# Single-Cell RNA-seq Analysis Report

Analysis

OmicsDesk

2026-05-14

## Table of contents

# 1 Executive Summary

This report presents a single-cell RNA sequencing (scRNA-seq) analysis of myocardial remodeling in a mouse transverse aortic constriction (TAC) model across four conditions: Sham control, and TAC at 2, 4, and 6 weeks post-surgery. After quality filtering, 27,146 high-quality cells were retained across all four samples. **Cell type landscape.** Marker-based annotation identified eight major cell populations. The cardiac microenvironment is dominated by fibroblasts (41.1%), consistent with the pronounced fibrotic remodeling expected in pressure-overload models. Macrophages represent the second largest population (20.3%), reflecting an active inflammatory response. B cells (14.2%) and neutrophils (9.0%) indicate sustained immune infiltration. Endothelial cells (7.3%), T cells (5.6%), cardiomyocytes (1.4%), and smooth muscle cells (1.1%) complete the cellular landscape. The low cardiomyocyte capture rate is expected in droplet-based scRNA-seq due to their large cell size. **Temporal dynamics.** Cell type composition shifts markedly over the TAC time course. At 4 and 6 weeks, B cell and macrophage proportions increase substantially compared to Sham, while fibroblast proportions decrease relatively — suggesting a progressive transition from a fibrosis-dominated to an inflammation-dominated microenvironment as heart failure advances. **Differential expression.** At 2 weeks post-TAC, the transcriptomic response is modest (12 upregulated, 8 downregulated genes; padj $< 0.05$, $|\log2FC| > 1$). The top upregulated gene is Nppa (atrial natriuretic peptide), a canonical marker of cardiac stress and heart failure. By 6 weeks, the response is dramatically amplified: 73 genes are upregulated and 697 are downregulated. Upregulated genes are enriched for ribosomal proteins (Rps28, Rpl36, Rpl37, Rpl38), indicating increased translational activity, and B cell markers (Igkc, Cd79a), confirming the immune cell infiltration observed in the composition analysis. Downregulated genes include hemoglobin subunits (Hbb-bs, Hba-a1, Hba-a2; log2FC = -3.7 to -4.6) and multiple mitochondrial genes (mt-Nd5, mt-Nd2, mt-Nd4, mt-Cytb, mt-Atp6), pointing to mitochondrial dysfunction — a hallmark of the failing heart. Additionally, Vim (vimentin) and Rock1 (Rho-kinase, involved in cardiac fibrosis signaling) are significantly downregulated at 6 weeks.

# 2 Methods

## 2.1 Quality Control & Preprocessing

Raw count matrices (10x Genomics Cell Ranger output) were loaded using **scanpy** (v1.9+). Per-cell quality metrics were computed: number of detected genes, total UMI counts, and percentage of mitochondrial reads (mt- prefix). Doublets were detected and removed using **Scrublet** (v0.2+) with default parameters. Cells were filtered to retain those with:

- Genes per cell: see QC table
- Maximum mitochondrial fraction: see QC table

## 2.2 Normalization & Dimensionality Reduction

Library-size normalization was applied (**scanpy** `normalize_total`, target sum = 10,000), followed by log1p transformation. Highly variable genes (HVGs) were selected using the Seurat v3 method (top 2,000 genes). Data were scaled (max value = 10) and principal component analysis (PCA) was performed on HVGs (50 components).

## 2.3 Batch Integration

When multiple samples were present, batch effects were corrected using **Harmony** (harmonypy) on PCA embeddings, with sample identity as the batch key.

## 2.4 Clustering

A k-nearest neighbors graph was constructed (k = 15, 30 PCs) and cells were embedded using **UMAP**. Community detection was performed using the **Leiden** algorithm (igraph implementation).

## 2.5 Cell Type Annotation

Automated cell type annotation was performed using **CellTypist** with majority voting to assign consensus labels per Leiden cluster. The model was selected based on organism (mouse).

## 2.6 Marker Genes & Differential Expression

Cluster marker genes were identified using the Wilcoxon rank-sum test (**scanpy** `rank_genes_groups`, groupby = leiden). For condition-based differential expression, the same test was applied per contrast (numerator vs. denominator conditions). Significance thresholds: adjusted p-value (Benjamini-Hochberg) $< 0.05$, $|\log2 \text{ fold-change}| > 1$.

## 2.7 Spatial Integration

Spatial transcriptomics data (10x Visium) were processed with **squidpy** (v1.2+). A spatial neighborhood graph was constructed (generic coordinates, k = 10 neighbors). Spatially variable genes were identified using **Moran's I** autocorrelation statistic on the top 100 HVGs. Cell type labels were transferred from the scRNA-seq reference by scoring each spatial spot for marker gene enrichment (`scanpy.tl.score_genes`) using the top 15 marker genes per cell type, and assigning the cell type with the highest score. Neighborhood enrichment analysis was performed using the transferred cell type labels to assess spatial co-localization patterns.

# 3 Quality Control

## 3.1 QC Summary

*How to read this table:* One row per sample. "Cells (post-filter)" is how many cells survived the quality filters and the doublet check — the count the rest of the report is built on. "Median genes/cell" is the typical transcriptional complexity per cell; a healthy 10x run sits in the low thousands. "Median UMIs/cell" tracks library depth. "Median % mito" is the mitochondrial-read fraction — high values flag stressed or dying cells. Big disparities between samples (e.g. one sample with half the genes/cell of the others) are worth a look before trusting downstream comparisons.

Table 1: Per-sample QC metrics after filtering.

| Sample | Cells (post-filter) | Median genes/cell | Median UMIs/cell | Median % mito |
|---|---|---|---|---|
| Sham | 8218 | 2634 | 8931 | 4.5 |
| TAC2w | 6577 | 2911 | 10379 | 4.6 |
| TAC4w | 6811 | 1390 | 4356 | 5.0 |
| TAC6w | 5540 | 1639 | 4673 | 2.6 |

## 3.2 QC Distributions (Pre-filter)

*How to read these plots:* Each violin/histogram shows the distribution of a QC metric across all cells before filtering. Look for long tails on the right (very-high-gene cells = likely doublets) and on the left (very-low-gene cells = empty droplets or dying cells). The mitochondrial-fraction panel shows the cut-off we apply to remove stressed cells.

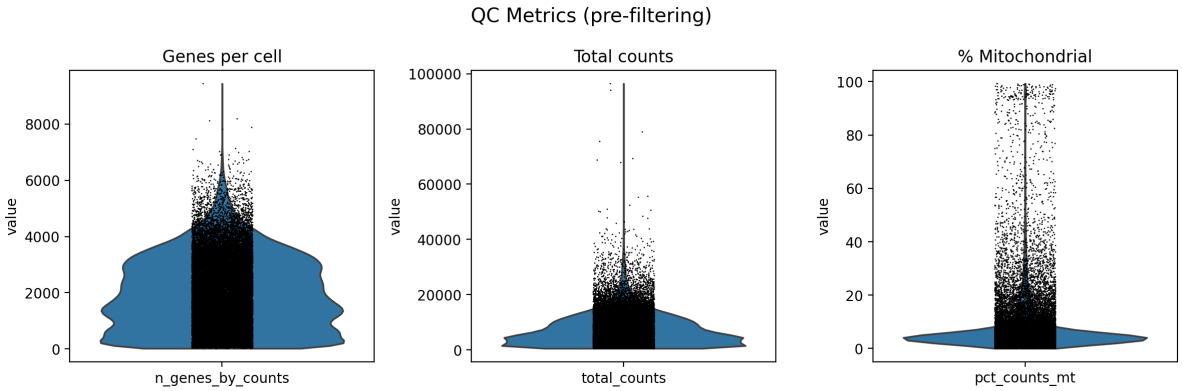


Figure 1: QC metrics before filtering: genes per cell, total UMI counts, and percentage mitochondrial reads.

## 3.3 QC Distributions (Post-filter)

*How to read these plots:* Same metrics as above, recomputed after low-quality cells and doublets are removed. The distributions should look tighter — long tails clipped, mitochondrial fraction capped — and roughly comparable across samples. This is the cell population every later analysis is run on.

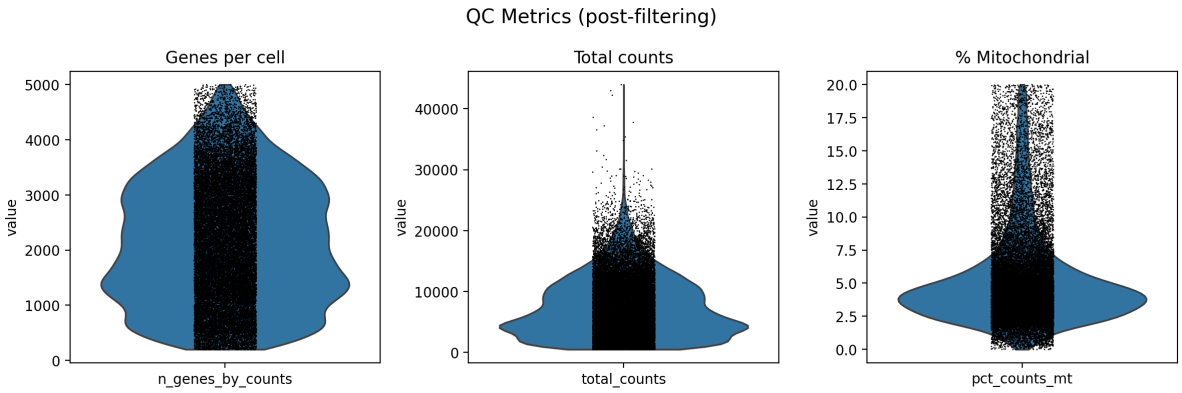


Figure 2: QC metrics after filtering and doublet removal.

## 3.4 QC Metrics per Sample

*How to read this figure:* One column per sample, three rows for the three QC metrics (genes/cell, UMIs/cell, % mito). Samples that look obviously different from the rest — much lower complexity, much higher mito — are flagged here *before* you trust them in the cell-type and DE panels.

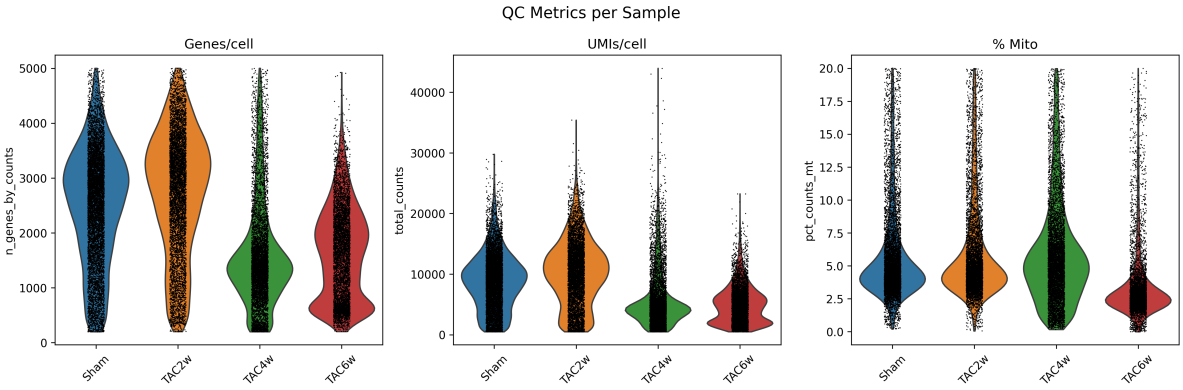


Figure 3: Distribution of genes per cell, UMIs per cell, and mitochondrial fraction across samples.

*What this means for your data:* After filtering and doublet removal, 27,146 high-quality cells remain across 4 sample(s) (median 2,136 genes/cell, median 4.5% mitochondrial reads).

## 3.5 PCA Variance

*How to read this figure:* Each bar is one principal component (PC) and its height is the share of total variance that PC explains. A few tall bars at the start mean a small number of components capture the bulk of the biological signal — that's typical for clean datasets. A long flat tail is fine; it just means the residual variance is spread thinly across many small dimensions.

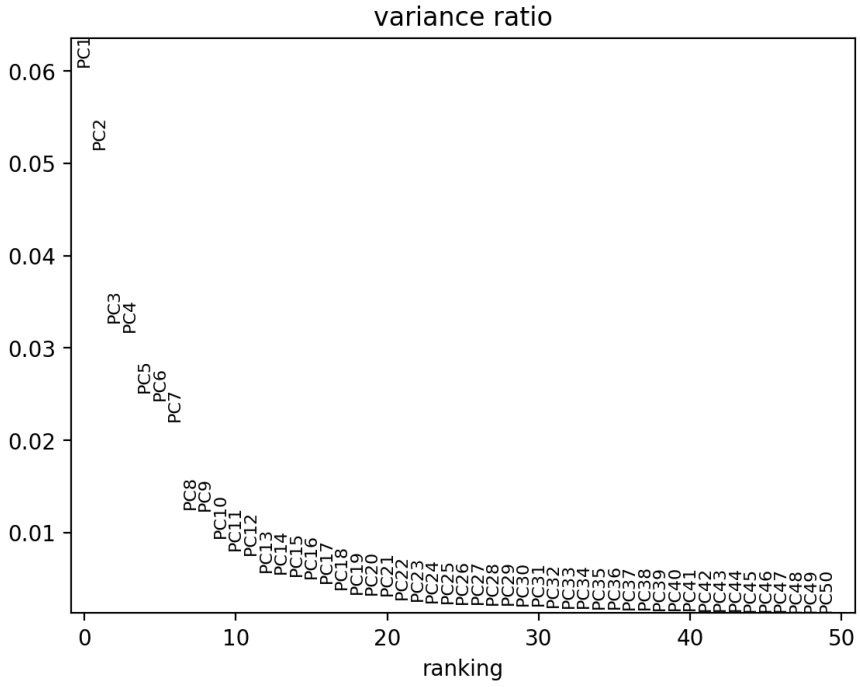


Figure 4: Proportion of variance explained by each principal component.

# 4 Clustering & Cell Types

## 4.1 Dataset Overview

*How to read this figure:* Each dot is one cell. UMAP places similar cells next to each other in 2D — the actual coordinates have no biological meaning, only the *grouping* does. The three panels colour the same cells three ways: by Leiden cluster (the unsupervised grouping), by cell-type label (what each cluster was annotated as), and by sample of origin (does each sample cover the whole map, or do samples form separate islands?). Islands per sample = a batch effect that warrants follow-up.

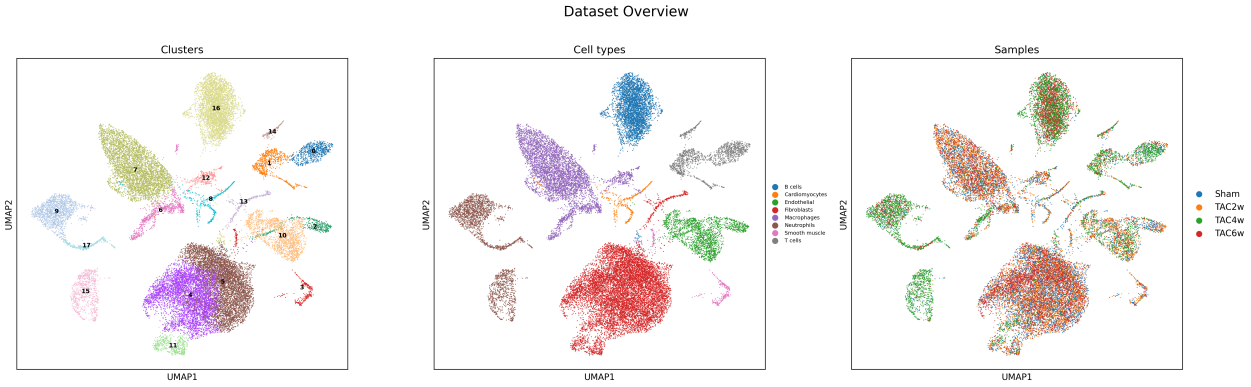


Figure 5: UMAP projections colored by Leiden cluster, cell type annotation, and sample of origin.

*What this means for your data:* UMAP shared across all panels: 27,146 cells in 18 Leiden cluster(s) / 8 cell type(s) from 4 sample(s). Sample-coloured panel checks for batch effects (mixing = good; per-sample islands = warrants Harmony review).

## 4.2 Cell Types

*How to read this figure:* Same UMAP, now coloured by the cell-type label that CellTypist assigned to each cluster via majority voting. Each coloured region is a population — e.g. fibroblasts, macrophages, T cells — defined by its marker-gene profile, not by where it sits on the plot. Compare side-by-side with the Leiden panel: well-annotated data shows a clean one-cluster-per-type mapping; messy annotations show types spread across many clusters or one cluster split between types.

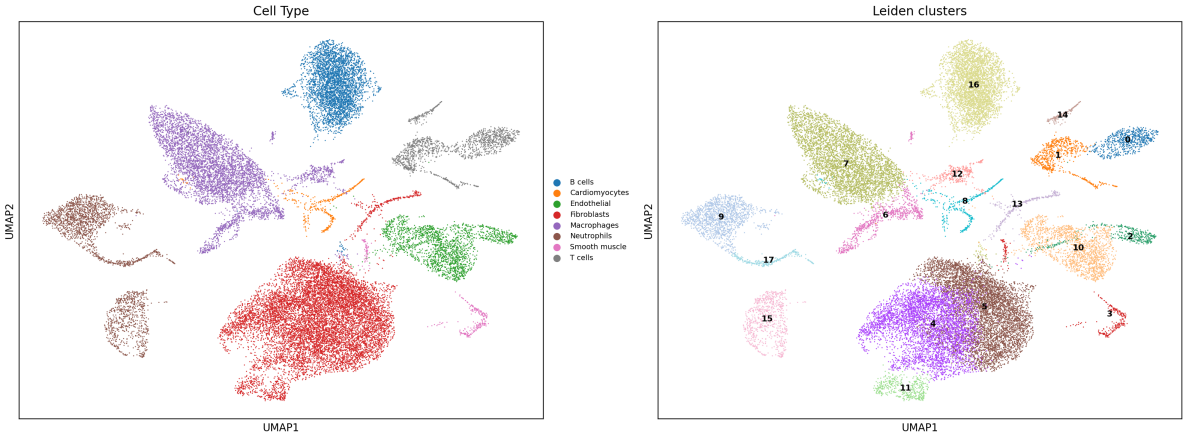


Figure 6: UMAP colored by CellTypist annotation (majority voting) alongside Leiden clusters.

*What this means for your data:* CellTypist annotated 27,146 cells into 8 cell type(s) across 18 Leiden cluster(s). Most abundant population: **Fibroblasts** (41.1%).

## 4.3 Cell Type Composition

*How to read this table:* One row per cell type. "Count" is the number of cells assigned to that type across the whole dataset; "Percentage" is its share of all cells. The top rows are the dominant populations in your tissue. Rare populations (typically $< 1$ %) show up at the bottom — useful to know exist, but under-powered for statistics on their own.

Table 2: Cell type counts and proportions.

| Cell Type | Count | Percentage |
|---|---|---|
| Fibroblasts | 11170 | 41.1 |
| Macrophages | 5515 | 20.3 |
| B cells | 3855 | 14.2 |
| Neutrophils | 2439 | 9.0 |
| Endothelial | 1983 | 7.3 |
| T cells | 1517 | 5.6 |
| Cardiomyocytes | 378 | 1.4 |
| Smooth muscle | 289 | 1.1 |

## 4.4 Composition per Cluster

*How to read this figure:* Each stacked bar is one Leiden cluster, segmented by the cell-type labels assigned to its cells. A clean cluster is one tall colour; a mixed cluster shows multiple colours, meaning either over-clustering (split one type in two) or biologically similar types that the algorithm couldn't separate.

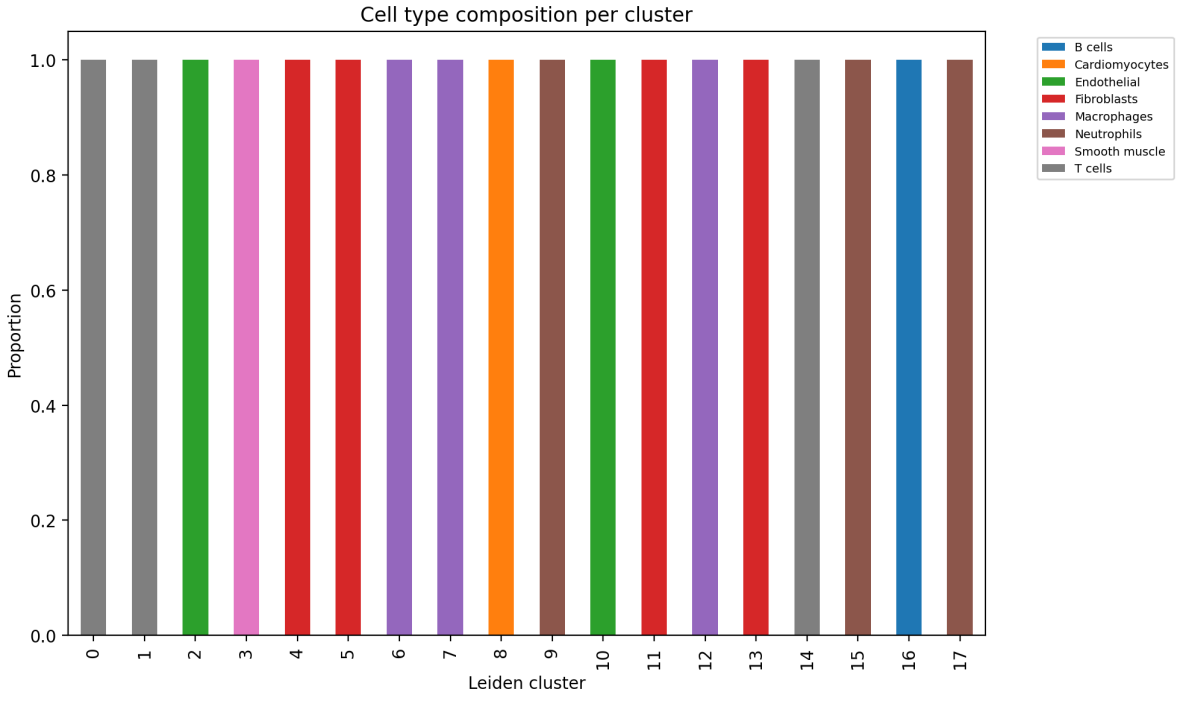


Figure 7: Cell type composition within each Leiden cluster.

## 4.5 Cell Type Proportions per Sample

*How to read this figure:* Each stacked bar is one sample (or condition); colours show the cell-type composition. A condition that loses or gains a coloured slab compared with the others is a candidate composition shift — formal significance is tested in §6 (Differential Abundance) further down.

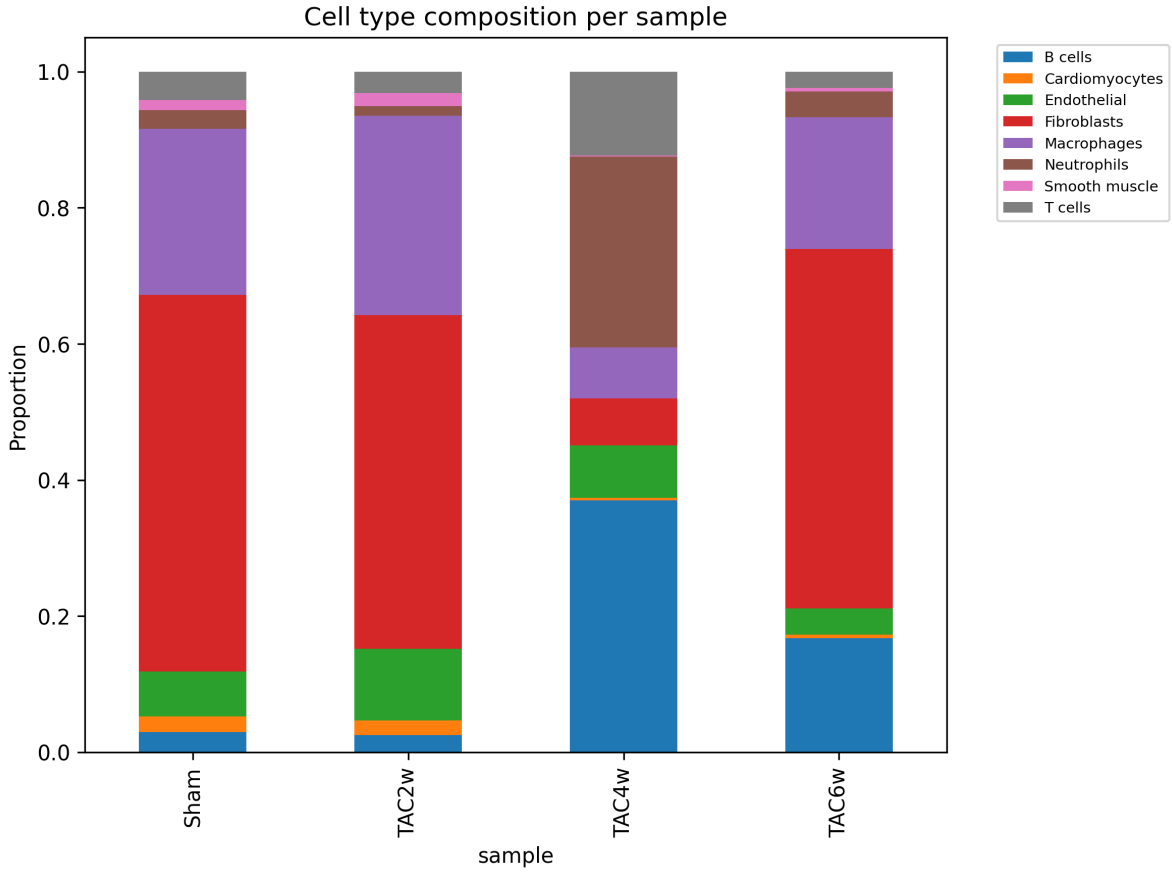


Figure 8: Stacked bar chart of cell type proportions across samples/conditions.

*What this means for your data:* Stacked bars: cell-type composition per condition. Largest shift across conditions is in **Fibroblasts** (48.4 percentage points).

# 5 Marker Genes

## 5.1 Dot Plot — Top Markers per Cluster

*How to read this figure:* Rows are clusters, columns are the top marker genes selected for them. Each dot at the intersection has two visual channels: dot **size** = the fraction of cells in that cluster expressing the gene (specificity), dot **colour** = the average expression level among the cells that do express it (intensity). A good marker is big *and* dark in its own cluster and small/pale everywhere else.

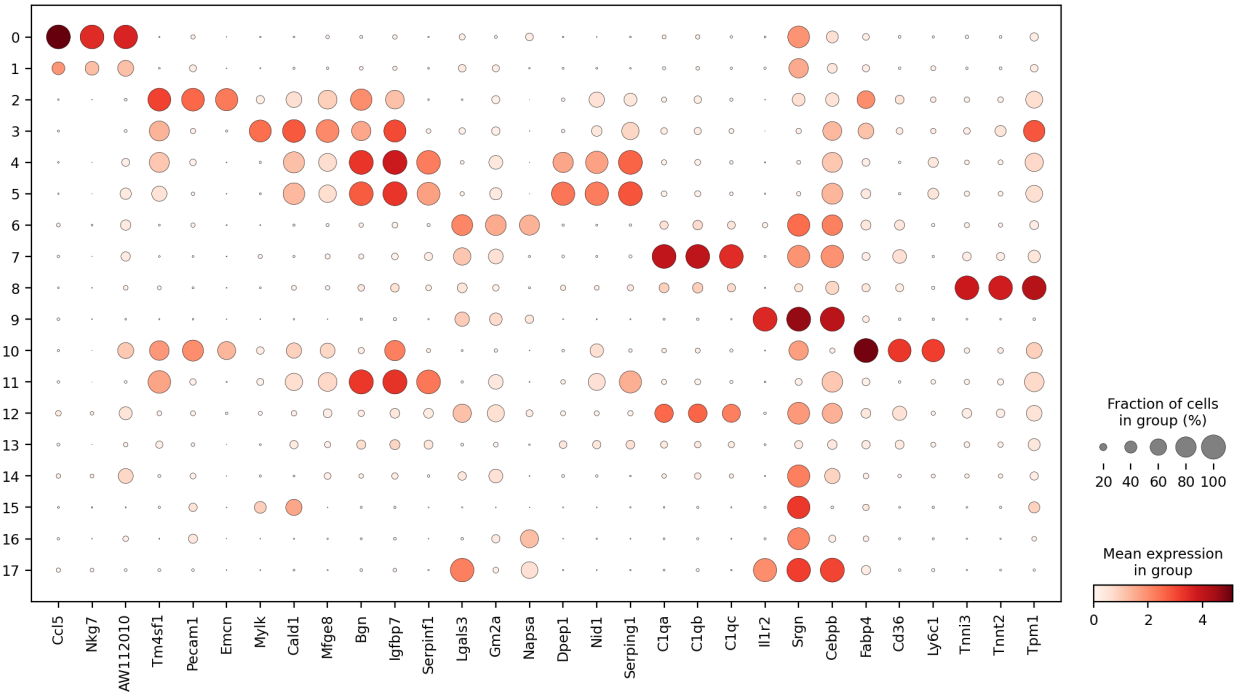


Figure 9: Dot plot of top marker genes per Leiden cluster. Dot size: fraction of cells expressing the gene. Color: mean expression level.

*What this means for your data:* Top marker genes are shown for each of the 18 Leiden cluster(s) (Wilcoxon rank-sum test). Dot size encodes the fraction of cells expressing the gene; color encodes mean expression in expressing cells.

## 5.2 Heatmap — Top Markers

*How to read this figure:* Rows are individual cells, grouped by cluster; columns are the top 5 marker genes per cluster. Cell colour encodes the expression level (darker = higher). Diagonal blocks of high expression mean each cluster has its own distinct marker signature — the cleaner the diagonal, the better separated the populations are.

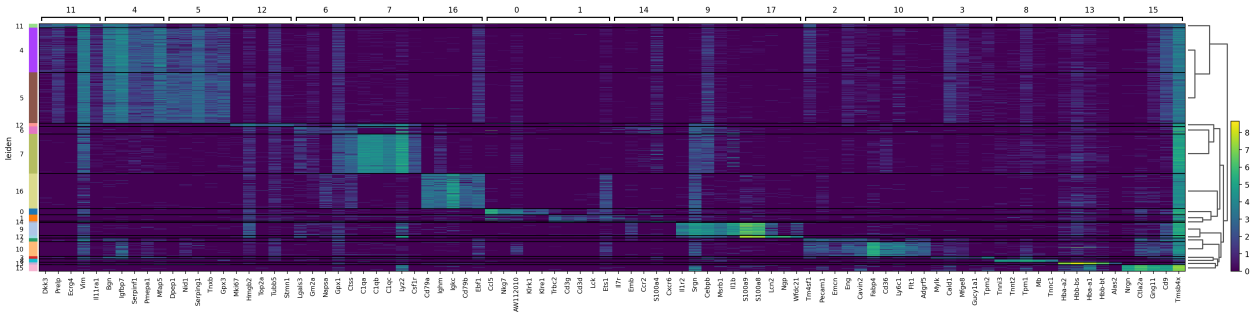

Figure 10: Heatmap of top 5 marker genes per cluster (Wilcoxon rank-sum test).

## 5.3 Marker Violin Plots

*How to read this figure:* One row per gene, one column per cluster. Each violin shows the distribution of that gene's expression in that cluster — width = how many cells sit at that expression level. A gene that is "on" only in one cluster (one tall violin, the rest flat) is a clean marker; a gene "on" in several clusters is a shared programme rather than a marker.

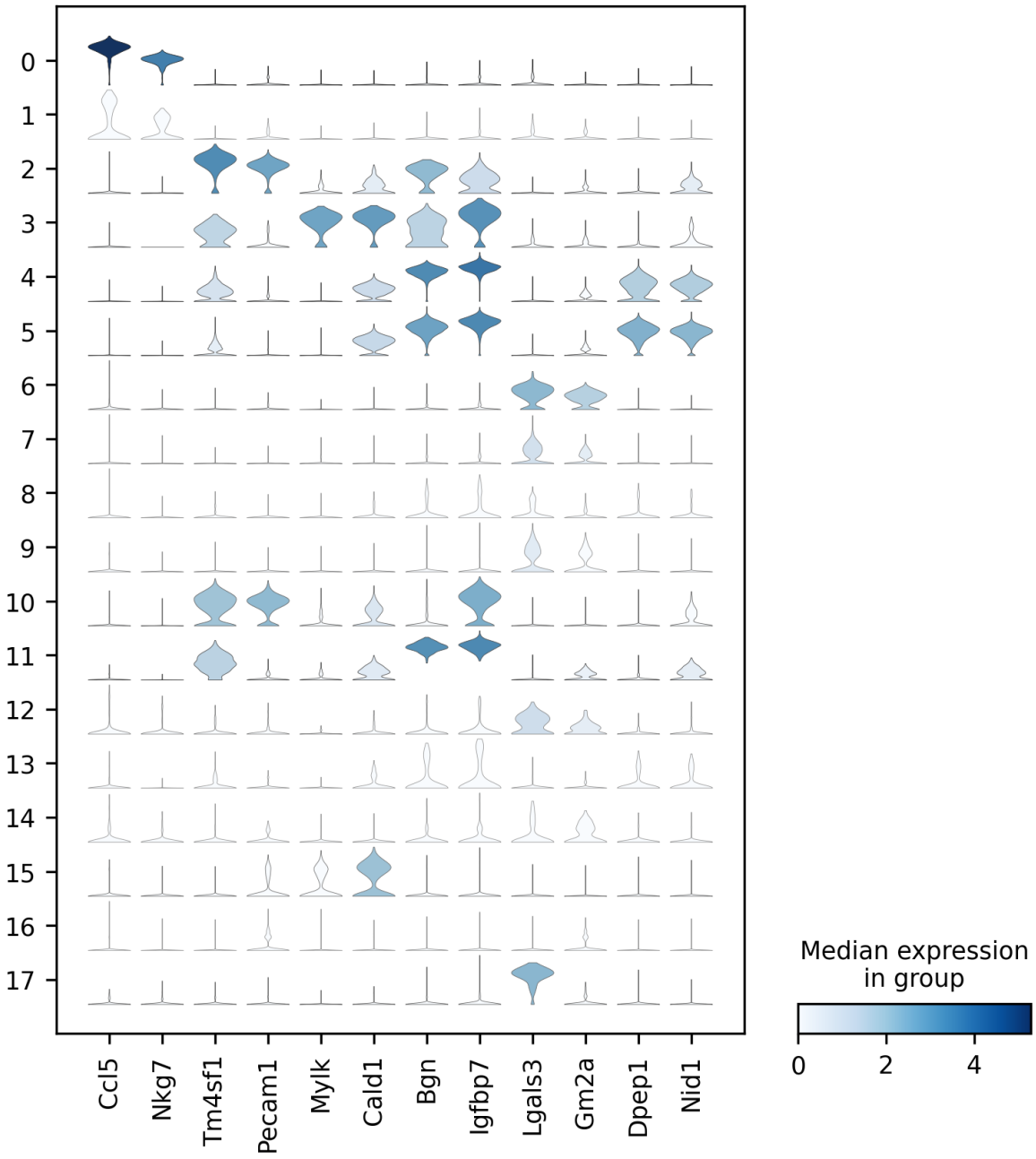


Figure 11: Stacked violin plot of top marker genes across Leiden clusters.

## 5.4 Top Marker Genes

*How to read this table:* For each cluster, the top 5 genes ranked by Wilcoxon rank-sum score. “Score” is the test statistic (higher = stronger separation against other clusters); “log2FC” is the average fold-change versus

the rest of the dataset; “padj” is the Benjamini-Hochberg-adjusted p-value (lower = stronger evidence).

Table 3: Top 5 marker genes per cluster (Wilcoxon rank-sum test, ordered by score).

| Cluster | Gene | Score | log2FC | padj |
|---|---|---|---|---|
| 0 | Ccl5 | 42.91 | 10.60 | 0 |
| 0 | Nkg7 | 42.77 | 9.39 | 0 |
| 0 | AW112010 | 41.97 | 6.63 | 0 |
| 0 | Klrk1 | 41.48 | 8.26 | 0 |
| 0 | Klre1 | 41.23 | 10.12 | 0 |
| 1 | Trbc2 | 40.59 | 8.01 | 0 |
| 1 | Cd3g | 38.28 | 8.28 | 0 |
| 1 | Cd3d | 37.75 | 8.14 | 0 |
| 1 | Lck | 35.17 | 6.27 | 0 |
| 1 | Ets1 | 34.58 | 3.45 | 0 |
| 2 | Tm4sf1 | 29.75 | 5.09 | 0 |
| 2 | Pecam1 | 29.75 | 5.37 | 0 |
| 2 | Emcn | 29.64 | 6.56 | 0 |
| 2 | Eng | 28.70 | 4.57 | 0 |
| 2 | Cavin2 | 28.35 | 4.83 | 0 |
| 3 | Mylk | 25.38 | 7.08 | 0 |
| 3 | Cald1 | 25.18 | 4.12 | 0 |
| 3 | Mfge8 | 24.68 | 3.94 | 0 |
| 3 | Gucy1a1 | 23.89 | 5.65 | 0 |
| 3 | Tpm2 | 23.71 | 7.11 | 0 |
| 4 | Bgn | 91.42 | 4.17 | 0 |
| 4 | Igfbp7 | 90.69 | 4.20 | 0 |
| 4 | Serpinf1 | 88.33 | 3.51 | 0 |
| 4 | Pmepa1 | 86.78 | 3.37 | 0 |
| 4 | Mfap5 | 86.65 | 4.10 | 0 |
| 5 | Dpep1 | 94.82 | 4.27 | 0 |
| 5 | Nid1 | 94.45 | 3.76 | 0 |
| 5 | Serping1 | 92.37 | 4.00 | 0 |
| 5 | Tnxb | 91.44 | 3.61 | 0 |
| 5 | Gpx3 | 90.71 | 3.82 | 0 |
| 6 | Lgals3 | 35.36 | 4.11 | 0 |
| 6 | Gm2a | 32.73 | 2.98 | 0 |
| 6 | Napsa | 32.36 | 3.61 | 0 |
| 6 | Gpx1 | 31.98 | 2.21 | 0 |
| 6 | Ctss | 30.94 | 3.09 | 0 |
| 7 | C1qa | 101.56 | 8.15 | 0 |
| 7 | C1qb | 101.50 | 7.99 | 0 |
| 7 | C1qc | 100.52 | 7.88 | 0 |
| 7 | Lyz2 | 96.32 | 6.41 | 0 |
| 7 | Csf1r | 95.34 | 5.95 | 0 |
| 8 | Tnni3 | 32.50 | 8.40 | 0 |
| 8 | Tnnt2 | 32.50 | 8.33 | 0 |
| 8 | Tpm1 | 31.94 | 6.52 | 0 |
| 8 | Mb | 31.71 | 8.24 | 0 |
| 8 | Tnnc1 | 30.55 | 8.38 | 0 |
| 9 | Il1r2 | 63.06 | 9.19 | 0 |
| 9 | Srgn | 62.52 | 5.57 | 0 |
| 9 | Cebpb | 62.41 | 4.97 | 0 |
| 9 | Msrb1 | 62.36 | 6.49 | 0 |
| 9 | Il1b | 62.27 | 8.60 | 0 |
| 10 | Fabp4 | 66.14 | 8.57 | 0 |
| 10 | Cd36 | 61.41 | 7.02 | 0 |
| 10 | Ly6c1 | 59.59 | 6.48 | 0 |
| 10 | Flt1 | 59.07 | 7.50 | 0 |

Table 3: Top 5 marker genes per cluster (Wilcoxon rank-sum test, ordered by score). *(continued)*

| Cluster | Gene | Score | log2FC | padj |
|---|---|---|---|---|
| 10 | Adgrf5 | 57.40 | 7.49 | 0 |
| 11 | Dkk3 | 34.43 | 5.73 | 0 |
| 11 | Prelp | 33.06 | 4.52 | 0 |
| 11 | Ecrg4 | 31.37 | 6.94 | 0 |
| 11 | Vim | 29.70 | 2.85 | 0 |
| 11 | Il11ra1 | 29.19 | 3.15 | 0 |
| 12 | Mki67 | 31.47 | 8.31 | 0 |
| 12 | Hmgb2 | 30.46 | 4.45 | 0 |
| 12 | Top2a | 30.09 | 8.32 | 0 |
| 12 | Tubb5 | 30.06 | 3.58 | 0 |
| 12 | Stmn1 | 30.01 | 6.72 | 0 |
| 13 | Hba-a2 | 27.76 | 10.91 | 0 |
| 13 | Hbb-bs | 27.76 | 10.74 | 0 |
| 13 | Hba-a1 | 27.75 | 10.79 | 0 |
| 13 | Hbb-bt | 27.59 | 10.65 | 0 |
| 13 | Alas2 | 26.89 | 10.09 | 0 |
| 14 | Il7r | 19.61 | 7.63 | 0 |
| 14 | Emb | 17.52 | 4.99 | 0 |
| 14 | Ccr2 | 16.28 | 5.18 | 0 |
| 14 | S100a4 | 16.22 | 4.49 | 0 |
| 14 | Cxcr6 | 15.79 | 7.90 | 0 |
| 15 | Nrgn | 45.88 | 10.94 | 0 |
| 15 | Ctla2a | 45.86 | 9.90 | 0 |
| 15 | Gng11 | 45.84 | 8.03 | 0 |
| 15 | Cd9 | 45.67 | 5.99 | 0 |
| 15 | Tmsb4x | 45.62 | 4.54 | 0 |
| 16 | Cd79a | 95.06 | 9.74 | 0 |
| 16 | Ighm | 95.02 | 6.23 | 0 |
| 16 | Igkc | 94.25 | 9.67 | 0 |
| 16 | Cd79b | 89.34 | 7.77 | 0 |
| 16 | Ebf1 | 87.47 | 4.02 | 0 |
| 17 | S100a9 | 28.99 | 11.32 | 0 |
| 17 | S100a8 | 28.98 | 11.19 | 0 |
| 17 | Lcn2 | 28.98 | 9.72 | 0 |
| 17 | Ngp | 28.91 | 12.07 | 0 |
| 17 | Wfdc21 | 28.89 | 8.78 | 0 |

# 6 Differential Expression

## 6.1 Results Summary

*How to read this table:* One row per contrast (one biological comparison). "Total tested" is the number of genes that survived expression-level pre-filtering. "Sig. UP" / "Sig. DOWN" count genes that cross both the adjusted-p-value threshold (0.05) and the effect-size threshold ($|\log2FC| > 1$). A near-balanced UP/DOWN split typically means a broad perturbation; a strongly skewed split usually points to a directional biology (e.g. an inflammatory programme being switched on or repressed).

Table 4: DEG counts (padj $< 0.05$, $|\log2FC| > 1$)

| contrast | Total tested | Sig. UP | Sig. DOWN |
|---|---|---|---|
| TAC2w_vs_Sham | 32285 | 12 | 8 |
| TAC6w_vs_Sham | 32285 | 73 | 697 |

*What this means for your data:* For **TAC6w_vs_Sham**: 73 up, 697 down at FDR<0.05 and |log2FC|>1.0. Top significant gene: **Rps28**.

## 6.2 Volcano Plots

*How to read these plots:* Each dot is one gene. The horizontal axis (log2 fold-change) shows how much a gene moves up or down between the two conditions (positive = higher in the first group, negative = lower). The vertical axis ($-\log$ padj) shows confidence — higher = stronger statistical evidence. Genes in the **top-right corner** are confidently up; **top-left** confidently down. Dashed lines mark the thresholds used in this report.

### 6.2.1 12_volcano_TAC2w_vs_Sham

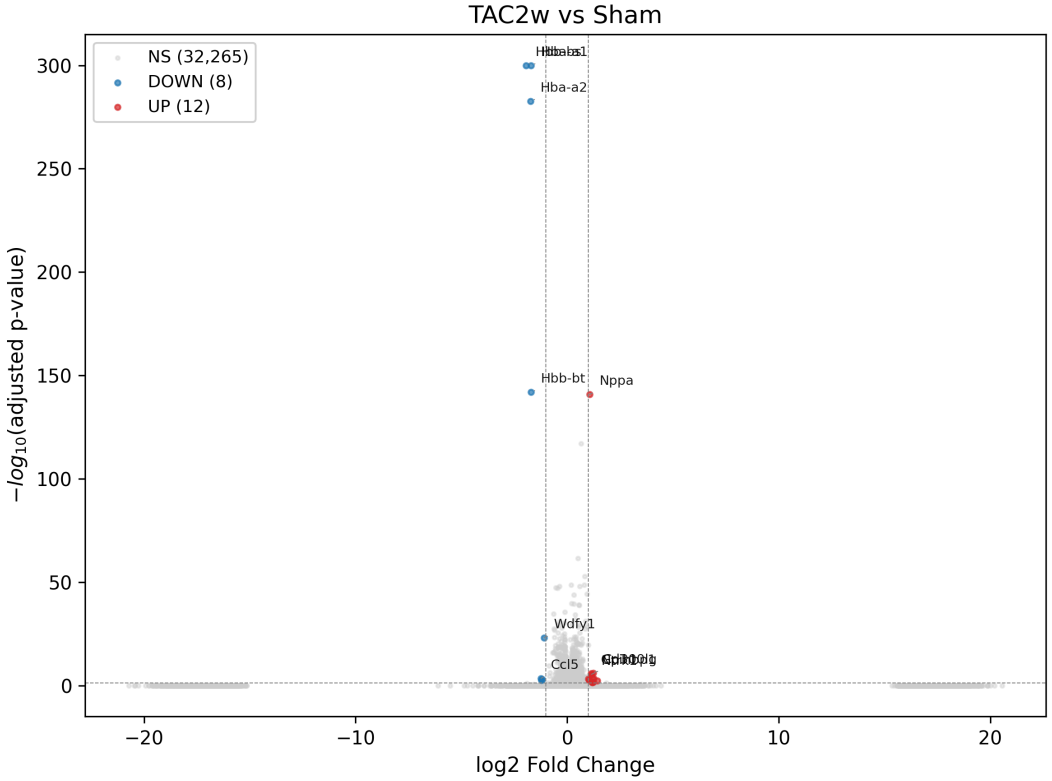


Figure 12: Volcano plot: 12_volcano_TAC2w_vs_Sham. Crimson: significant; grey: not significant.

### 6.2.2 12_volcano_TAC6w_vs_Sham

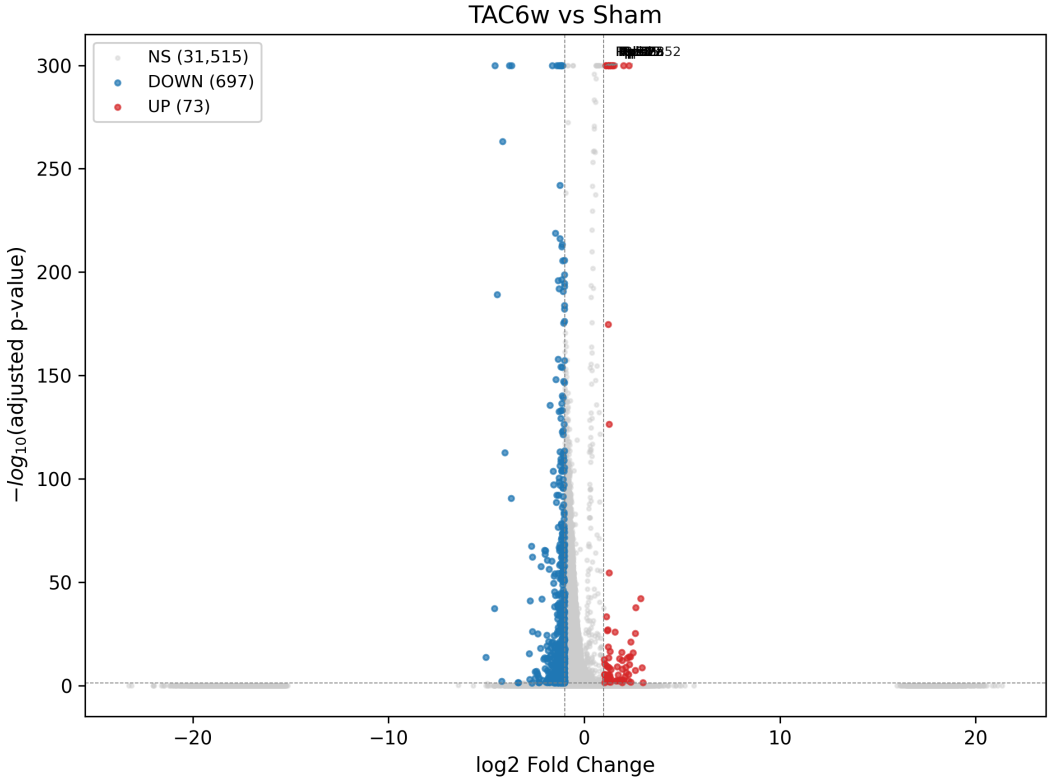


Figure 13: Volcano plot: 12_volcano_TAC6w_vs_Sham. Crimson: significant; grey: not significant.

## 6.3 Top Differentially Expressed Genes (Up-regulated)

*How to read this table:* Within each contrast, the top 15 up-regulated genes ordered by adjusted p-value (most confident first). "log2FC" is the effect-size: a value of 1 means the gene is roughly 2× higher in the numerator condition, 2 means 4× higher, 3 means 8× higher. "padj" is the Benjamini-Hochberg adjusted p-value — lower = stronger evidence. The full ranked list is in the Excel workbook (`results_workbook.xlsx`).

## 6.4 Top Differentially Expressed Genes (Down-regulated)

*How to read this table:* Same as the up-regulated table, but for genes significantly lower in the numerator condition. Negative "log2FC" values indicate the size of the decrease (−1 halved, −2 quartered).

Table 5: Top up-regulated genes per contrast (ordered by adjusted p-value).

| Contrast | Gene | log2FC | padj |
|---|---|---|---|
| TAC2w_vs_Sham | Nppa | 1.06 | 0.00e+00 |
| TAC2w_vs_Sham | Cd300lg | 1.22 | 1.00e-06 |
| TAC2w_vs_Sham | Gpihbp1 | 1.13 | 1.50e-06 |
| TAC2w_vs_Sham | Ntrk1 | 1.15 | 4.00e-06 |
| TAC2w_vs_Sham | Timp4 | 1.25 | 3.71e-04 |
| TAC2w_vs_Sham | Adgrl4 | 1.00 | 5.22e-04 |
| TAC2w_vs_Sham | Rbp7 | 1.20 | 5.99e-04 |
| TAC2w_vs_Sham | Gm42031 | 1.05 | 2.39e-03 |
| TAC2w_vs_Sham | Btnl9 | 1.43 | 6.38e-03 |
| TAC2w_vs_Sham | C1qtnf9 | 1.24 | 2.08e-02 |
| TAC2w_vs_Sham | Tcf15 | 1.16 | 2.32e-02 |
| TAC2w_vs_Sham | Slc15a2 | 1.20 | 4.25e-02 |
| TAC6w_vs_Sham | Rps28 | 1.53 | 0.00e+00 |
| TAC6w_vs_Sham | Uba52 | 2.30 | 0.00e+00 |
| TAC6w_vs_Sham | Rps29 | 1.33 | 0.00e+00 |
| TAC6w_vs_Sham | Rps27 | 1.44 | 0.00e+00 |
| TAC6w_vs_Sham | Rpl37a | 1.23 | 0.00e+00 |
| TAC6w_vs_Sham | Rpl35 | 1.43 | 0.00e+00 |
| TAC6w_vs_Sham | Rpl37 | 1.10 | 0.00e+00 |
| TAC6w_vs_Sham | Rps21 | 1.22 | 0.00e+00 |
| TAC6w_vs_Sham | Rpl38 | 1.28 | 0.00e+00 |
| TAC6w_vs_Sham | Rpl36 | 1.11 | 0.00e+00 |
| TAC6w_vs_Sham | Gm11808 | 2.02 | 0.00e+00 |
| TAC6w_vs_Sham | Tmsb10 | 1.34 | 0.00e+00 |
| TAC6w_vs_Sham | Btg1 | 1.23 | 0.00e+00 |
| TAC6w_vs_Sham | Bloc1s1 | 1.26 | 0.00e+00 |
| TAC6w_vs_Sham | Jup | 1.27 | 0.00e+00 |

Table 6: Top down-regulated genes per contrast (ordered by adjusted p-value).

| Contrast | Gene | log2FC | padj |
|---|---|---|---|
| TAC2w_vs_Sham | Hba-a1 | -1.71 | 0.000000 |
| TAC2w_vs_Sham | Hbb-bs | -1.95 | 0.000000 |
| TAC2w_vs_Sham | Hba-a2 | -1.73 | 0.000000 |
| TAC2w_vs_Sham | Hbb-bt | -1.71 | 0.000000 |
| TAC2w_vs_Sham | Wdfy1 | -1.10 | 0.000000 |
| TAC2w_vs_Sham | Ccl5 | -1.23 | 0.000337 |
| TAC2w_vs_Sham | Cadm3 | -1.18 | 0.001120 |
| TAC2w_vs_Sham | Stmn4 | -1.21 | 0.002690 |
| TAC6w_vs_Sham | mt-Nd5 | -1.24 | 0.000000 |
| TAC6w_vs_Sham | Hba-a2 | -3.82 | 0.000000 |
| TAC6w_vs_Sham | Hba-a1 | -3.72 | 0.000000 |
| TAC6w_vs_Sham | mt-Cytb | -1.16 | 0.000000 |
| TAC6w_vs_Sham | mt-Nd4 | -1.41 | 0.000000 |
| TAC6w_vs_Sham | mt-Atp6 | -1.18 | 0.000000 |
| TAC6w_vs_Sham | mt-Nd2 | -1.64 | 0.000000 |
| TAC6w_vs_Sham | Hbb-bs | -4.57 | 0.000000 |
| TAC6w_vs_Sham | Npm1 | -1.11 | 0.000000 |
| TAC6w_vs_Sham | Sept7 | -1.31 | 0.000000 |
| TAC6w_vs_Sham | Hbb-bt | -4.16 | 0.000000 |
| TAC6w_vs_Sham | Capza2 | -1.24 | 0.000000 |
| TAC6w_vs_Sham | Pid1 | -1.48 | 0.000000 |
| TAC6w_vs_Sham | Atrx | -1.23 | 0.000000 |
| TAC6w_vs_Sham | Sdcbp | -1.13 | 0.000000 |

## 6.5 Cell Type Distribution per Condition

*How to read this figure:* The same UMAP coordinates as the overview, split into one panel per condition. Cells coloured by cell type. Compare panels side-by-side: a cluster that's bright in one condition and empty in another is either appearing or disappearing biologically — that's the visual signal underlying the differential abundance test below.

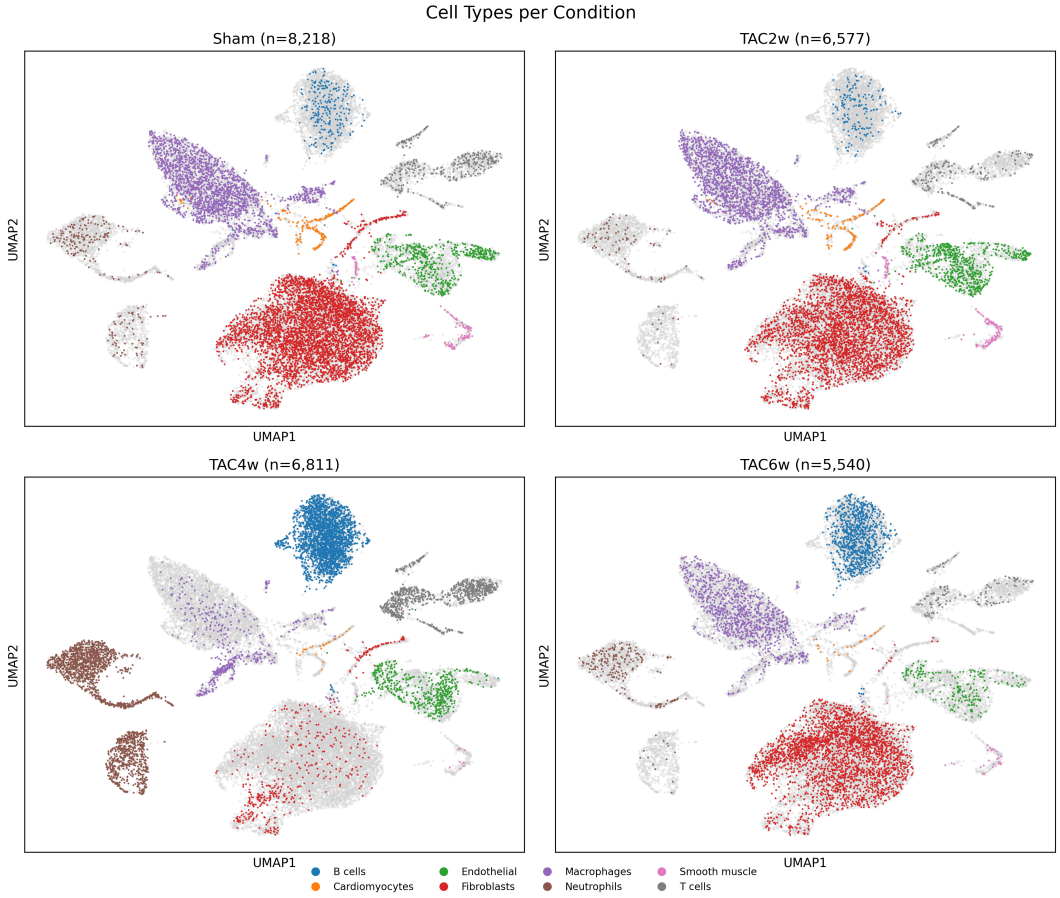


Figure 14: UMAP projections split by condition, colored by cell type. Each panel shows cells from one condition on the shared embedding.

*What this means for your data:* Stacked bars: cell-type composition per condition. Largest shift across conditions is in **Fibroblasts** (48.4 percentage points).

# 7 Differential Abundance

*How to read this table:* One row per cell type. The "%" columns are the share of cells of that type in each condition (rows must sum to ~100 % across all cell types within a condition, not within a row). "FDR" is the Benjamini-Hochberg-adjusted p-value for the test that the proportion *changes* across conditions; "Significant" is the FDR $< 0.05$ flag. Compositional caveat: because the proportions sum to 1, when one cell type goes up, others mechanically go down — interpret co-moving rows together rather than as independent events.

Table 7: Differential abundance testing of cell type proportions across conditions (scCODA).

| cell_type | Sham (%) | TAC2w (%) | TAC4w (%) | TAC6w (%) | FDR | Significant |
|---|---|---|---|---|---|---|
| B cells | 3.0 | 2.5 | 37.0 | 16.8 | 0 | Yes |
| Cardiomyocytes | 2.3 | 2.2 | 0.4 | 0.5 | 0 | Yes |
| Endothelial | 6.6 | 10.6 | 7.7 | 3.9 | 0 | Yes |
| Fibroblasts | 55.3 | 49.0 | 6.9 | 52.8 | 0 | Yes |
| Macrophages | 24.5 | 29.3 | 7.4 | 19.3 | 0 | Yes |
| Neutrophils | 2.7 | 1.4 | 28.1 | 3.8 | 0 | Yes |
| Smooth muscle | 1.5 | 1.9 | 0.2 | 0.5 | 0 | Yes |
| T cells | 4.1 | 3.1 | 12.3 | 2.4 | 0 | Yes |

*How to read this figure:* One panel or bar per cell type, showing how its proportion shifts across conditions. The y-axis is the share of all cells, not absolute counts. Asterisks (or a coloured annotation) flag cell types whose shift passes the FDR cut-off. Compare the visible trend with the FDR column in the table above — eye-catching shifts that aren't significant usually mean high sample-to-sample variability.

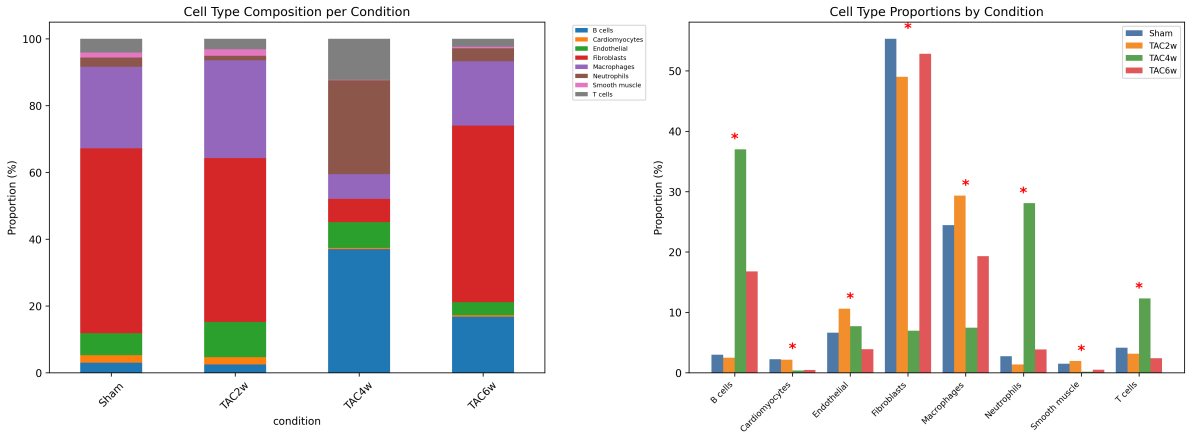


Figure 15: Cell type proportion changes across conditions with statistical significance.

*What this means for your data:* 8/8 cell type(s) show significant proportion shifts across conditions (FDR $<$ 0.05): B cells, Cardiomyocytes, Endothelial, Fibroblasts, Macrophages, Neutrophils, Smooth muscle, T cells. Most significant: **B cells** (FDR = 0.00e+00).

# 8 Trajectory Analysis

*How to read this figure:* PAGA summarises the cluster-level "wiring" of the dataset: each node is one cell-type/cluster and each edge represents how many cells lie at the transcriptional boundary between the two — thicker edges = stronger connectivity. Use it to see which populations are neighbours in expression space (often suggestive of lineage relationships) and which are isolated.

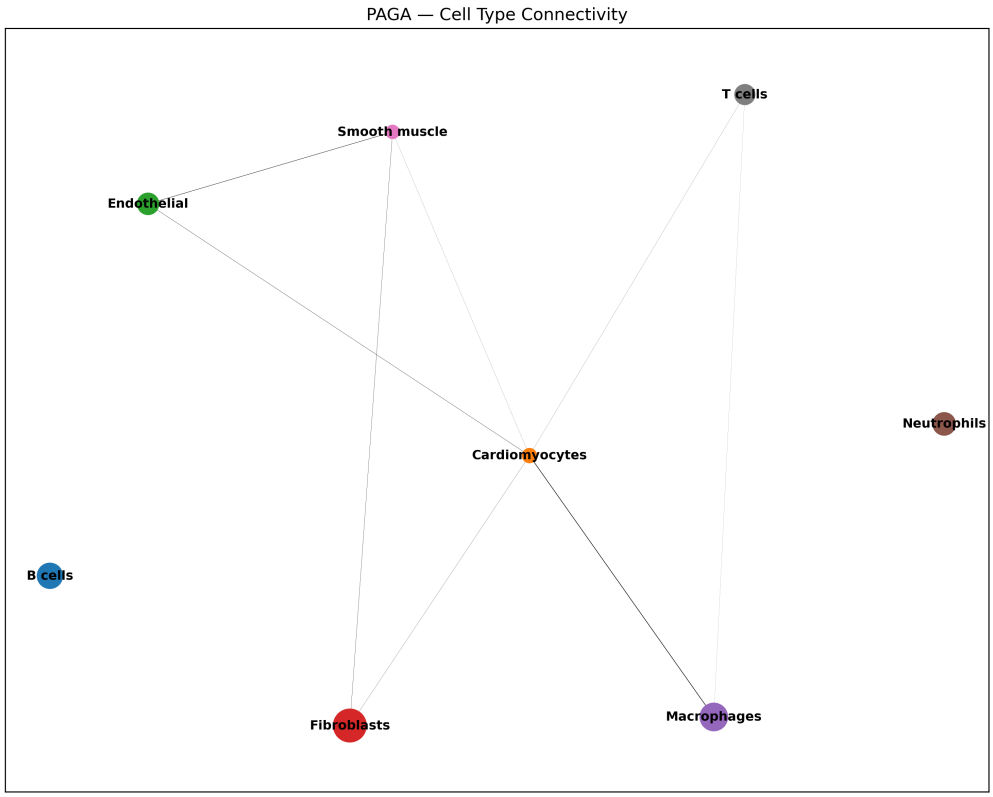


Figure 16: PAGA graph showing connectivity between cell type clusters. Edge thickness indicates connectivity strength.

*How to read this figure:* Same UMAP layout, now coloured by diffusion pseudotime. Pseudotime is a relative measure of how far each cell has moved from a chosen root along the inferred trajectory — it is not real-clock time. A smooth colour gradient along a UMAP arm = a candidate continuous transition; abrupt colour boundaries = discrete state changes.

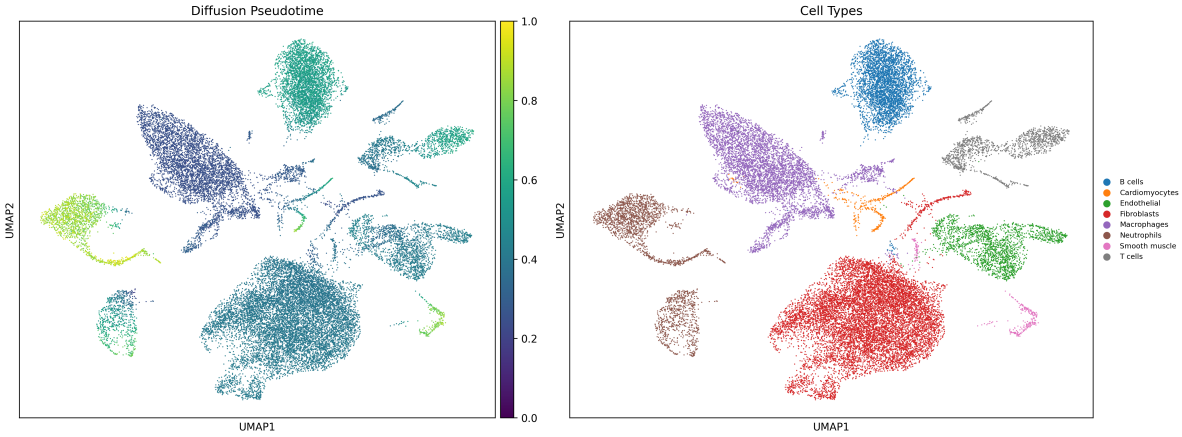


Figure 17: Diffusion pseudotime projected onto UMAP. Color indicates pseudotime progression from root cell.

*How to read this figure:* Each panel/row is one condition; the curve shows the distribution of pseudotime values across that condition's cells. A condition whose mass shifts to the right has more "advanced" cells along the inferred trajectory; one whose mass shifts left has more cells near the root. This pairs naturally with the differential abundance section — the same shifts can show up here as a pseudotime drift.

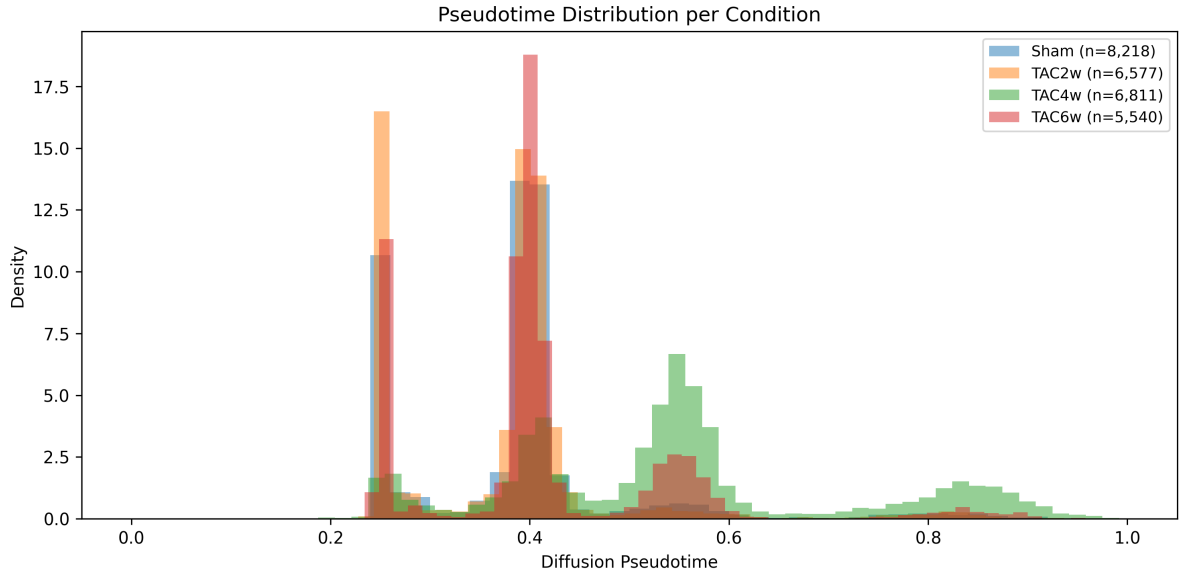


Figure 18: Pseudotime distribution per condition, showing cell state progression across timepoints.

# 9 Spatial Analysis

## 9.1 Spatially Variable Genes (Moran's I)

*How to read this table:* Moran's I measures spatial autocorrelation — how much a gene's expression at one tissue location predicts its expression at neighbouring locations. Values near 1 = strongly clustered (gene is on in specific tissue regions); values near 0 = scattered randomly across the slide. The genes at the top of this list are the most spatially organised features in the tissue.

Table 8: Top 20 spatially variable genes ranked by Moran's I statistic.

| ...1 | I | pval_norm | var_norm | pval_norm_fdr_bh |
|---|---|---|---|---|
| Ttn | 0.3175 | 0.0000 | 1e-04 | 0.0000 |
| Gsn | 0.2827 | 0.0000 | 1e-04 | 0.0000 |
| Ptgds | 0.2822 | 0.0000 | 1e-04 | 0.0000 |
| Ccl21a | 0.2201 | 0.0000 | 1e-04 | 0.0000 |
| Lcn2 | 0.1533 | 0.0000 | 1e-04 | 0.0000 |
| Xcl1 | 0.1421 | 0.0000 | 1e-04 | 0.0000 |
| Myl9 | 0.1347 | 0.0000 | 1e-04 | 0.0000 |
| Tpm2 | 0.1255 | 0.0000 | 1e-04 | 0.0000 |
| Col3a1 | 0.1060 | 0.0000 | 1e-04 | 0.0000 |
| Vim | 0.1056 | 0.0000 | 1e-04 | 0.0000 |
| Ankrd23 | 0.1014 | 0.0000 | 1e-04 | 0.0000 |
| Cfh | 0.0980 | 0.0000 | 1e-04 | 0.0000 |
| S100a8 | 0.0820 | 0.0000 | 1e-04 | 0.0000 |
| Tent5c | 0.0703 | 0.0000 | 1e-04 | 0.0000 |
| S100a9 | 0.0609 | 0.0000 | 1e-04 | 0.0000 |
| S100a4 | 0.0464 | 0.0000 | 1e-04 | 0.0000 |
| Abca4 | 0.0327 | 0.0010 | 1e-04 | 0.0062 |
| Hck | 0.0308 | 0.0019 | 1e-04 | 0.0104 |
| Stradb | 0.0306 | 0.0020 | 1e-04 | 0.0104 |
| Fubp1 | 0.0230 | 0.0147 | 1e-04 | 0.0733 |

*How to read these plots:* Each panel is the tissue section; each spot is one Visium voxel (~55 µm). Colour encodes the expression level of one of the top spatially variable genes — darker/redder spots indicate higher expression. Regions that light up together define putative tissue domains.

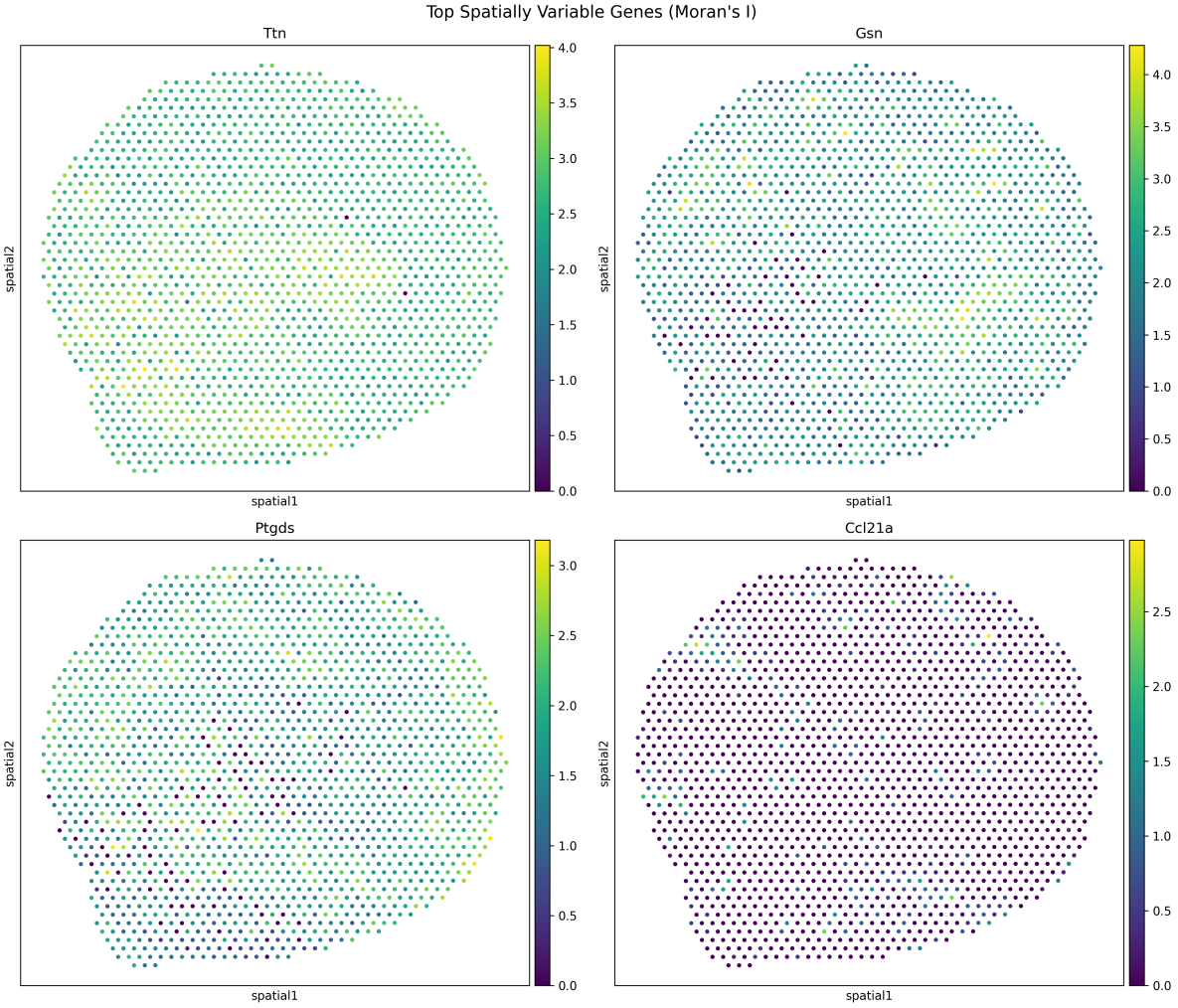


Figure 19: Top spatially variable genes colored by expression level.

## 9.2 Cell Type Spatial Mapping

Cell type labels were transferred from the scRNA-seq reference to spatial spots using marker gene scoring. Each spot was assigned the cell type with the highest enrichment score.

*How to read this figure:* The tissue section, with each Visium spot coloured by the cell type that scored highest from the scRNA-seq-derived marker panels. Bands of one colour = a tissue region dominated by that cell type.


Figure 20: Spatial tissue map colored by transferred cell type labels from scRNA-seq reference.

*How to read these plots:* One panel per cell type. Each spot is coloured by how strongly its expression matches that cell type's scRNA-seq marker signature — darker = higher score. Areas that score strongly in one panel and weakly in others identify spatially confined populations.


Figure 21: Per-cell-type enrichment scores across tissue. Darker red indicates higher marker gene set enrichment.

## 9.3 Differential Expression — Spatial Overlay

Top differentially expressed genes from the scRNA-seq condition comparisons are shown on the spatial tissue coordinates.

*How to read these plots:* Each panel shows the spatial expression of one top-ranked DE gene from the scRNA-seq comparisons. Look for whether the genes that came out of the dissociated-cell analysis are localised to specific tissue regions in situ — concordance strengthens the call.

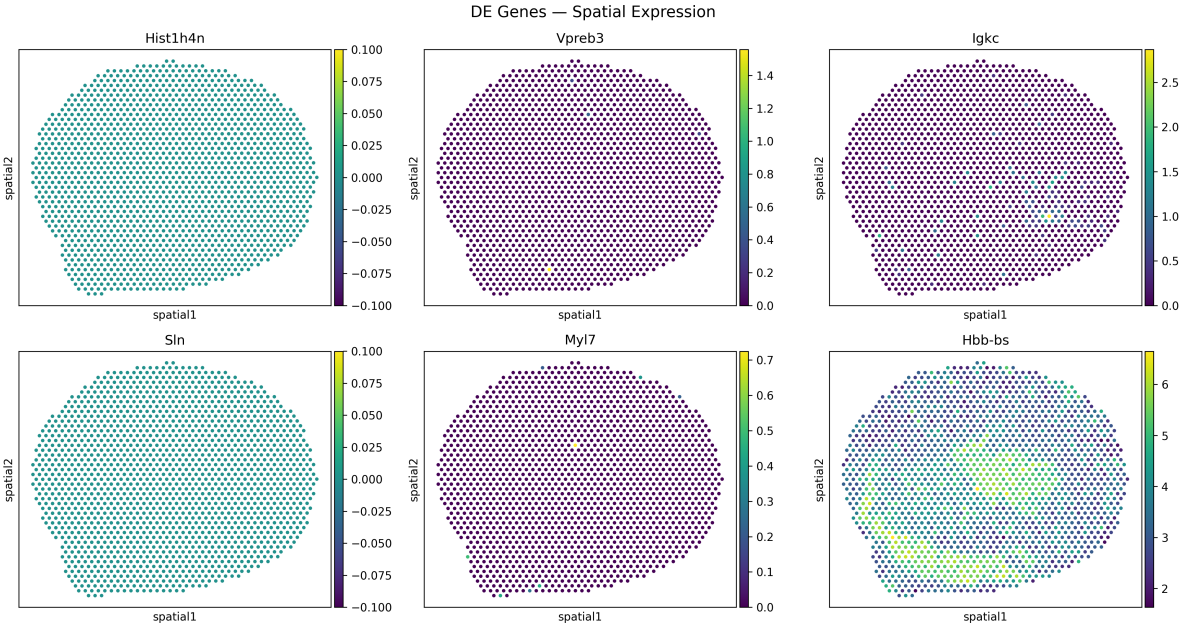


Figure 22: Spatial expression of top DE genes (scRNA-seq TAC vs Sham).

## 9.4 Cell Type Co-localization

*How to read this heatmap:* Rows and columns are both cell types; each cell shows whether those two types are physically next to each other on the tissue more often than expected by chance. Red/positive values = co-localised; blue/negative = avoid each other. The diagonal is each type with itself (regions where the type is internally clustered).

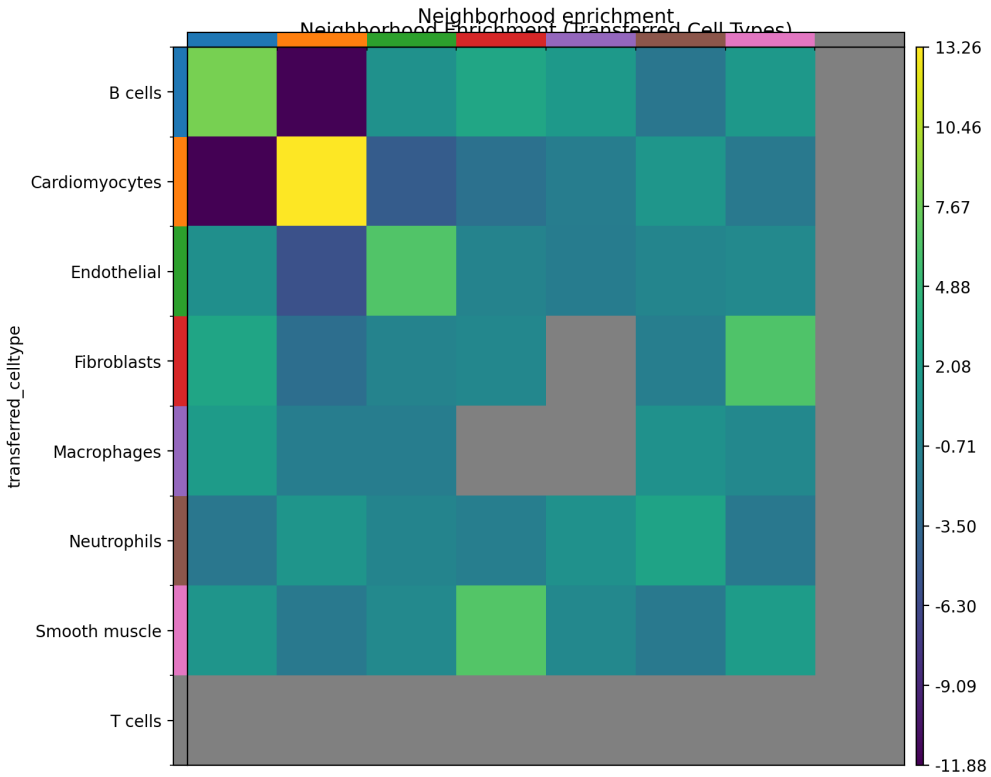


Figure 23: Neighborhood enrichment heatmap using transferred cell type labels. Positive values indicate spatial co-localization.

# 10 Conclusions

This single-cell analysis of the mouse TAC model reveals a clear spatiotemporal progression of myocardial remodeling at cellular resolution: 1. **Early response (2 weeks):** The transcriptomic signature is dominated by Nppa upregulation — the classical natriuretic peptide stress response — with minimal broader transcriptional changes. The cellular composition at this stage still resembles Sham, suggesting compensated hypertrophy. 2. **Late response (6 weeks):** A dramatic shift occurs, with 770 differentially expressed genes. Three key biological themes emerge: - **Immune infiltration:** B cell markers (Igkc, Cd79a) are among the most upregulated genes, and B cell proportions increase markedly in the composition data. This suggests an adaptive immune component to chronic heart failure that is often underappreciated in the TAC model literature. - **Mitochondrial dysfunction:** Five mitochondrial electron transport chain genes are significantly downregulated, consistent with the metabolic switch from fatty acid oxidation to glycolysis that characterizes the failing heart. - **Translational reprogramming:** Widespread upregulation of ribosomal proteins (Rps/Rpl family) suggests increased protein synthesis demands, possibly driven by the hypertrophic and fibrotic remodeling program. 3. **Clinical relevance:** The fibroblast-dominated microenvironment (41% of all cells) and its progressive shift toward immune-dominated composition mirrors findings in human heart failure biopsies, supporting the translational relevance of this model. These findings provide a cellular-resolution roadmap of heart failure progression and identify B cell infiltration and mitochondrial gene downregulation as potential therapeutic targets or biomarkers for disease staging.

---

*Report generated by Client project — 2026-06-05 16:27:18.420575 Pipeline: scanpy + Scrublet + Harmony + Leiden + CellTypist + Wilcoxon + scCODA + PAGA Deliverables: PDF report + Excel workbook (`results_workbook.xlsx`) Full DEG tables available in the Excel workbook. For questions contact: hello@omicsdesk.com*