Integrative Analysis of GWAS Risk Genes in the Human Protein Atlas

Zhuoran Cai

2026-02-13

This report is generated automatically from the GWAS and Human Protein Atlas (HPA) analysis pipeline.

It is meant as: - a **reference** for what the code can do, and
- a **summary** of the main results for a given disease or trait.

The report assumes that:

- risk genes were retrieved from the **GWAS Catalog**, expression and cluster information were retrieved from **HPA** (can be done through the searching script), and
- enrichment and specificity analyses were run using the provided scripts.

Full datasets can be found in the exported file for further investigation.

0. Input and Search Terms

This section lists the key inputs used to generate this report: EFO terms used for GWAS risk gene retrieval; Date of the GWAS search; Human Protein Atlas (HPA) version used for expression data.

```
## # A tibble: 7 x 3
##   efo_id        trait                                     uri
##   <chr>         <chr>                                     <chr>
## 1 EFO_0006514   Alzheimer's disease biomarker measurement  http://www.ebi.ac.uk~
## 2 EFO_0006801   Alzheimer's disease neuropathologic change http://www.ebi.ac.uk~
## 3 MONDO_0004975 Alzheimer disease                          http://purl.obolibra~
## 4 EFO_1001870   late-onset Alzheimer's disease             http://www.ebi.ac.uk~
## 5 EFO_0009268   family history of Alzheimer's disease      http://www.ebi.ac.uk~
## 6 OBA_2001000   age of onset of Alzheimer disease          http://purl.obolibra~
## 7 EFO_0022957   early-onset Alzheimers disease             http://www.ebi.ac.uk~
```

```
## [1] "2026-02-13 18:38:39 CET"
```

```
## [1] "https://v24.proteinatlas.org/api/search_download.php"
```

1. Summary of statistic

This section gives a quick overview of how many risk genes were identified, how many were found in HPA, and how many could be mapped onto the brain expression UMAP.

```
## Total genes in gene list: 530
##  Risk genes found in HPA: 476
##  Risk genes that can be mapped in brain UMAP: 453
```

2. What clusters do risk gene enrich?

2a. Perform Over-representation analysis (ORA) using fisher's exact test to see enriched clusters.

-An over-representation analysis is used to determine which clusters are significantly enriched. Then, fold enrichment score for each cluster is calculated. **Fold Enrichment** measures how much the observed overlap exceeds random expectation. It is calculated as:

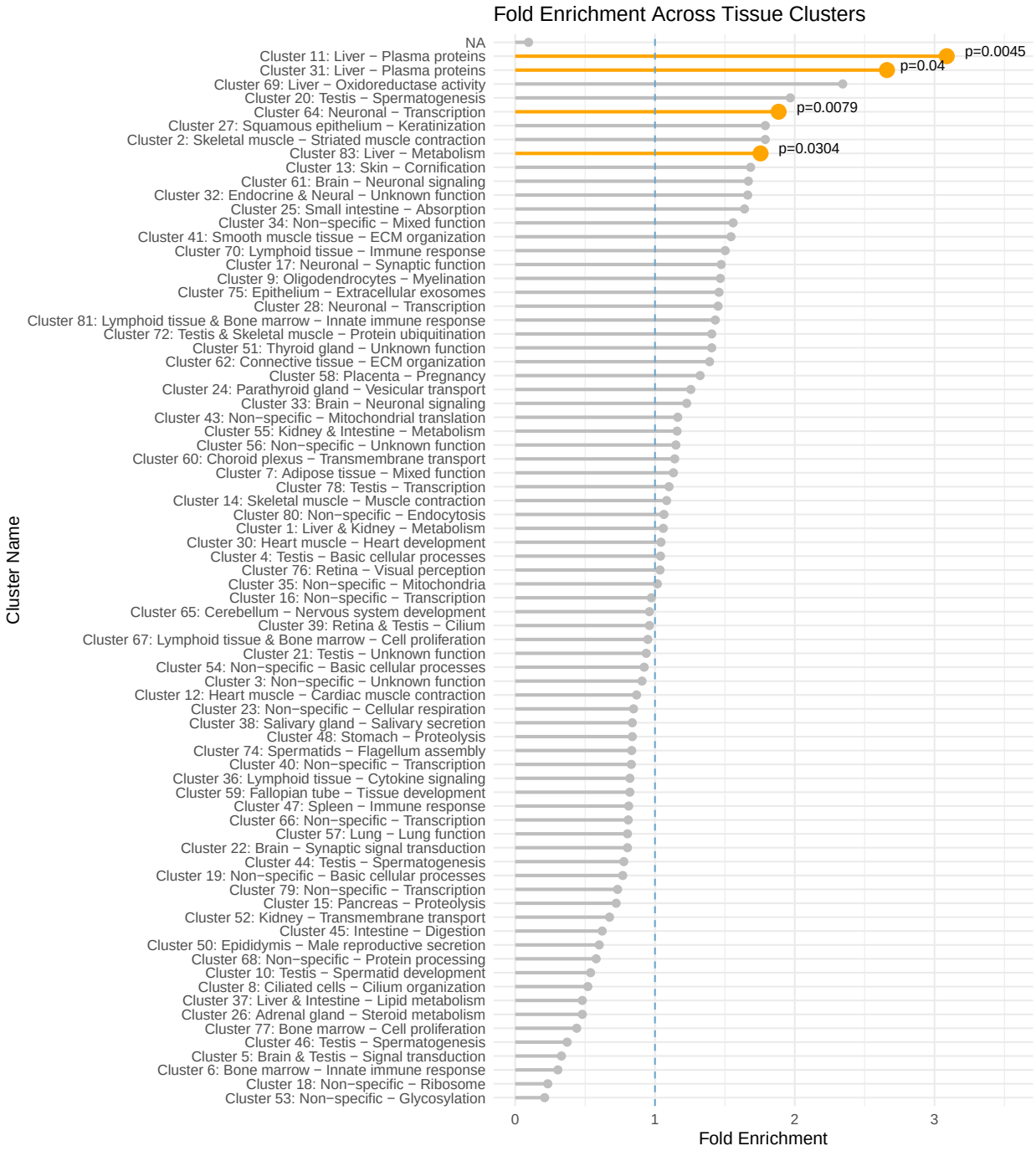
$$\text{Fold Enrichment} = \frac{\text{Observed Overlap}}{\text{Expected Overlap}}$$

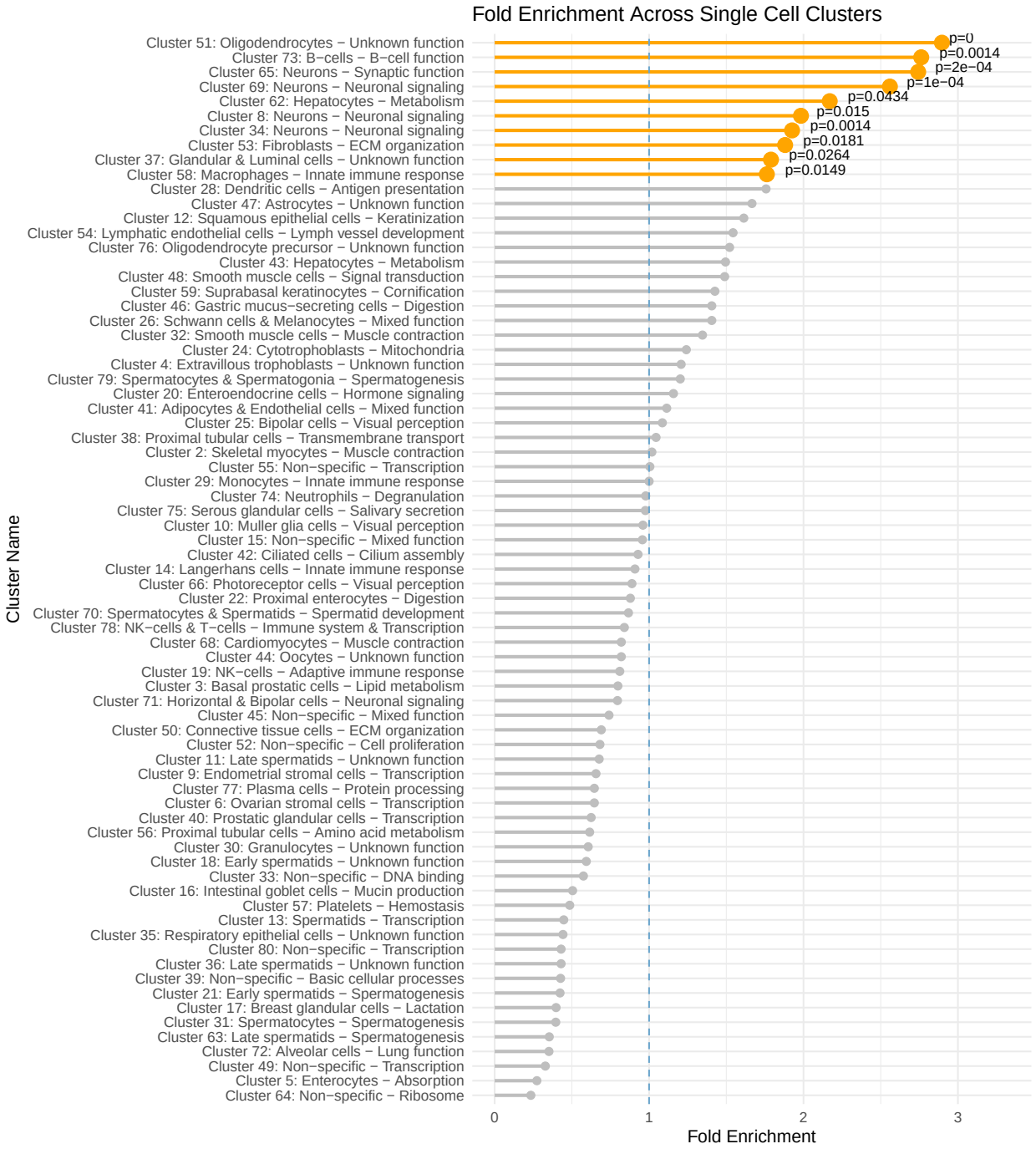
Where:

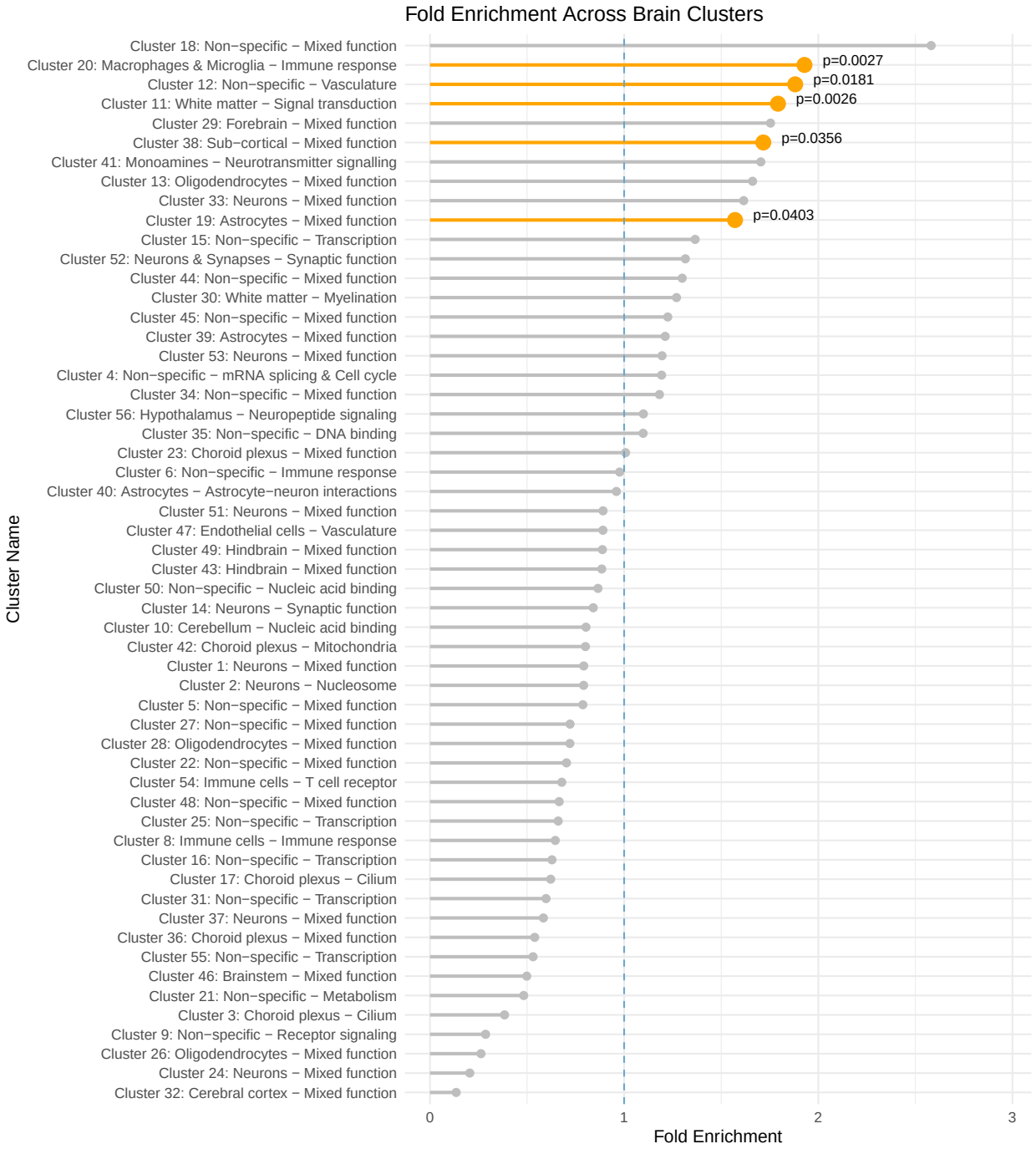
$$\text{Expected Overlap} = \frac{\text{Risk Genes} \times \text{Cluster Size}}{\text{Total Genes}}$$

- **Fold Enrichment > 1**: Indicates enrichment.
- **Fold Enrichment $= 1$**: Indicates no enrichment.
- **Fold Enrichment < 1**: Indicates depletion.

  In the lollipop plots below, enriched ($p<0.05$) clusters are highlighted in orange. x-axis represents fold enrichment.
- In the lollipop plots, each dot is a cluster. The x-axis shows fold enrichment and the y-axis shows the cluster label. Orange dots indicate significantly enriched clusters ($p < 0.05$). Note that since this pipeline is initially designed for HPA brain section/brain related diseases, genes not found in the HPA brain section database are excluded from further analysis.


Fold Enrichment Across Tissue Clusters
Cluster Name
NA
Cluster 11: Liver − Plasma proteins
Cluster 31: Liver − Plasma proteins
Cluster 69: Liver − Oxidoreductase activity
Cluster 20: Testis − Spermatogenesis
Cluster 64: Neuronal − Transcription
Cluster 27: Squamous epithelium − Keratinization
Cluster 2: Skeletal muscle − Striated muscle contraction
Cluster 83: Liver − Metabolism
Cluster 13: Skin − Cornification
Cluster 61: Brain − Neuronal signaling
Cluster 32: Endocrine & Neural − Unknown function
Cluster 25: Small intestine − Absorption
Cluster 34: Non−specific − Mixed function
Cluster 41: Smooth muscle tissue − ECM organization
Cluster 70: Lymphoid tissue − Immune response
Cluster 17: Neuronal − Synaptic function
Cluster 9: Oligodendrocytes − Myelination
Cluster 75: Epithelium − Extracellular exosomes
Cluster 28: Neuronal − Transcription
Cluster 81: Lymphoid tissue & Bone marrow − Innate immune response
Cluster 72: Testis & Skeletal muscle − Protein ubiquitination
Cluster 51: Thyroid gland − Unknown function
Cluster 62: Connective tissue − ECM organization
Cluster 58: Placenta − Pregnancy
Cluster 24: Parathyroid gland − Vesicular transport
Cluster 33: Brain − Neuronal signaling
Cluster 43: Non−specific − Mitochondrial translation
Cluster 55: Kidney & Intestine − Metabolism
Cluster 56: Non−specific − Unknown function
Cluster 60: Choroid plexus − Transmembrane transport
Cluster 7: Adipose tissue − Mixed function
Cluster 78: Testis − Transcription
Cluster 14: Skeletal muscle − Muscle contraction
Cluster 80: Non−specific − Endocytosis
Cluster 1: Liver & Kidney − Metabolism
Cluster 30: Heart muscle − Heart development
Cluster 4: Testis − Basic cellular processes
Cluster 76: Retina − Visual perception
Cluster 35: Non−specific − Mitochondria
Cluster 16: Non−specific − Transcription
Cluster 65: Cerebellum − Nervous system development
Cluster 39: Retina & Testis − Cilium
Cluster 67: Lymphoid tissue & Bone marrow − Cell proliferation
Cluster 21: Testis − Unknown function
Cluster 54: Non−specific − Basic cellular processes
Cluster 3: Non−specific − Unknown function
Cluster 12: Heart muscle − Cardiac muscle contraction
Cluster 23: Non−specific − Cellular respiration
Cluster 38: Salivary gland − Salivary secretion
Cluster 48: Stomach − Proteolysis
Cluster 74: Spermatids − Flagellum assembly
Cluster 40: Non−specific − Transcription
Cluster 36: Lymphoid tissue − Cytokine signaling
Cluster 59: Fallopian tube − Tissue development
Cluster 47: Spleen − Immune response
Cluster 66: Non−specific − Transcription
Cluster 57: Lung − Lung function
Cluster 22: Brain − Synaptic signal transduction
Cluster 44: Testis − Spermatogenesis
Cluster 19: Non−specific − Basic cellular processes
Cluster 79: Non−specific − Transcription
Cluster 15: Pancreas − Proteolysis
Cluster 52: Kidney − Transmembrane transport
Cluster 45: Intestine − Digestion
Cluster 50: Epididymis − Male reproductive secretion
Cluster 68: Non−specific − Protein processing
Cluster 10: Testis − Spermatid development
Cluster 8: Ciliated cells − Cilium organization
Cluster 37: Liver & Intestine − Lipid metabolism
Cluster 26: Adrenal gland − Steroid metabolism
Cluster 77: Bone marrow − Cell proliferation
Cluster 46: Testis − Spermatogenesis
Cluster 5: Brain & Testis − Signal transduction
Cluster 6: Bone marrow − Innate immune response
Cluster 18: Non−specific − Ribosome
Cluster 53: Non−specific − Glycosylation
p=0.0045
p=0.04
p=0.0079
p=0.0304
0
1
2
3
Fold Enrichment


Fold Enrichment Across Single Cell Clusters
Cluster 51: Oligodendrocytes − Unknown function
Cluster 73: B−cells − B−cell function
Cluster 65: Neurons − Synaptic function
Cluster 69: Neurons − Neuronal signaling
Cluster 62: Hepatocytes − Metabolism
Cluster 8: Neurons − Neuronal signaling
Cluster 34: Neurons − Neuronal signaling
Cluster 53: Fibroblasts − ECM organization
Cluster 37: Glandular & Luminal cells − Unknown function
Cluster 58: Macrophages − Innate immune response
Cluster 28: Dendritic cells − Antigen presentation
Cluster 47: Astrocytes − Unknown function
Cluster 12: Squamous epithelial cells − Keratinization
Cluster 54: Lymphatic endothelial cells − Lymph vessel development
Cluster 76: Oligodendrocyte precursor − Unknown function
Cluster 43: Hepatocytes − Metabolism
Cluster 48: Smooth muscle cells − Signal transduction
Cluster 59: Suprabasal keratinocytes − Cornification
Cluster 46: Gastric mucus−secreting cells − Digestion
Cluster 26: Schwann cells & Melanocytes − Mixed function
Cluster 32: Smooth muscle cells − Muscle contraction
Cluster 24: Cytotrophoblasts − Mitochondria
Cluster 4: Extravillous trophoblasts − Unknown function
Cluster 79: Spermatocytes & Spermatogonia − Spermatogenesis
Cluster 20: Enteroendocrine cells − Hormone signaling
Cluster 41: Adipocytes & Endothelial cells − Mixed function
Cluster 25: Bipolar cells − Visual perception
Cluster 38: Proximal tubular cells − Transmembrane transport
Cluster 2: Skeletal myocytes − Muscle contraction
Cluster 55: Non−specific − Transcription
Cluster 29: Monocytes − Innate immune response
Cluster 74: Neutrophils − Degranulation
Cluster 75: Serous glandular cells − Salivary secretion
Cluster 10: Muller glia cells − Visual perception
Cluster 15: Non−specific − Mixed function
Cluster 42: Ciliated cells − Cilium assembly
Cluster 14: Langerhans cells − Innate immune response
Cluster 66: Photoreceptor cells − Visual perception
Cluster 22: Proximal enterocytes − Digestion
Cluster 70: Spermatocytes & Spermatids − Spermatid development
Cluster 78: NK−cells & T−cells − Immune system & Transcription
Cluster 68: Cardiomyocytes − Muscle contraction
Cluster 44: Oocytes − Unknown function
Cluster 19: NK−cells − Adaptive immune response
Cluster 3: Basal prostatic cells − Lipid metabolism
Cluster 71: Horizontal & Bipolar cells − Neuronal signaling
Cluster 45: Non−specific − Mixed function
Cluster 50: Connective tissue cells − ECM organization
Cluster 52: Non−specific − Cell proliferation
Cluster 11: Late spermatids − Unknown function
Cluster 9: Endometrial stromal cells − Transcription
Cluster 77: Plasma cells − Protein processing
Cluster 6: Ovarian stromal cells − Transcription
Cluster 40: Prostatic glandular cells − Transcription
Cluster 56: Proximal tubular cells − Amino acid metabolism
Cluster 30: Granulocytes − Unknown function
Cluster 18: Early spermatids − Unknown function
Cluster 33: Non−specific − DNA binding
Cluster 16: Intestinal goblet cells − Mucin production
Cluster 57: Platelets − Hemostasis
Cluster 13: Spermatids − Transcription
Cluster 35: Respiratory epithelial cells − Unknown function
Cluster 80: Non−specific − Transcription
Cluster 36: Late spermatids − Unknown function
Cluster 39: Non−specific − Basic cellular processes
Cluster 21: Early spermatids − Spermatogenesis
Cluster 17: Breast glandular cells − Lactation
Cluster 31: Spermatocytes − Spermatogenesis
Cluster 63: Late spermatids − Spermatogenesis
Cluster 72: Alveolar cells − Lung function
Cluster 49: Non−specific − Transcription
Cluster 5: Enterocytes − Absorption
Cluster 64: Non−specific − Ribosome
p=0
p=0.0014
p=2e−04
p=1e−04
p=0.0434
p=0.015
p=0.0014
p=0.0181
p=0.0264
p=0.0149
Cluster Name
0
1
2
3
Fold Enrichment



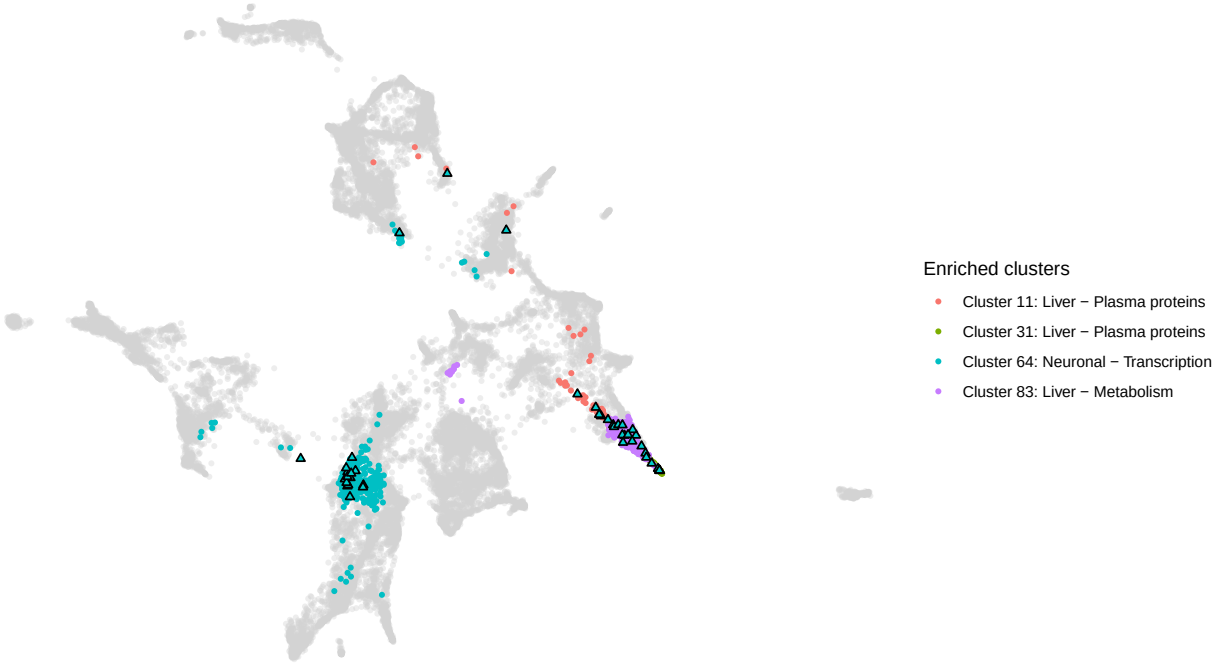
2b. Visualization of enriched clusters

The UMAP plots below show:

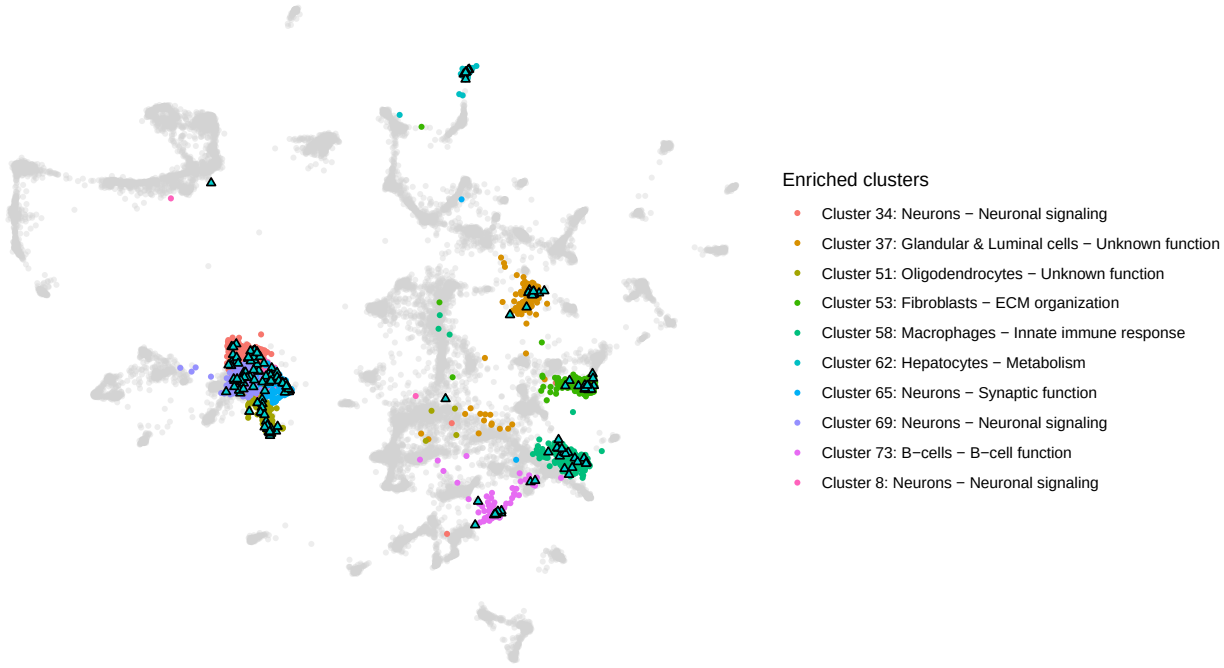
- All genes (grey)
- Genes in enriched clusters (colored)
- Risk genes within enriched clusters (cyan triangles)

These plots highlight where risk genes concentrate in transcriptomic space.

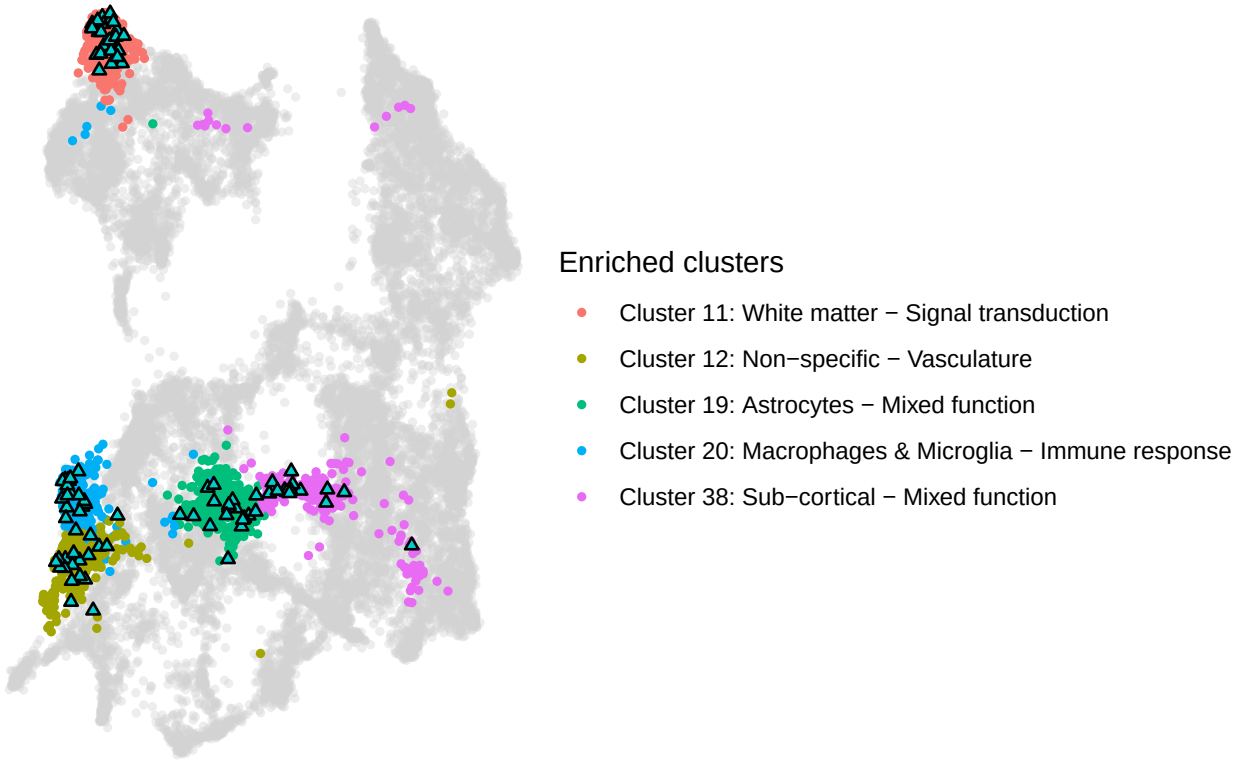
UMAP Plot (Tissue v24.1): Enriched clusters with risk genes



UMAP Plot (Single−cell v24.1): Enriched clusters with risk genes



UMAP Plot (Brain v24.1): Enriched clusters with risk genes



3. Specificity of Risk Genes

HPA assigns genes to specificity categories that describe how selectively they are expressed across tissues or cell types:

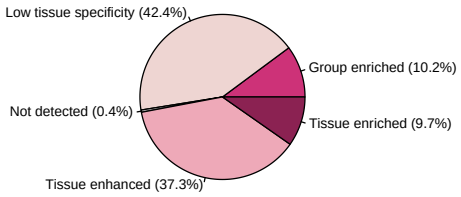
- **Tissue/Cell-type Enriched** genes shows markedly higher expression in one tissue or cell type compared to all others
- **Group enriched genes** displayed elevated expression in a small group of related tissues or cell types
- **Tissue/Cell-typeEnhanced** genes showed increased expression compared to the average across all tissues or cell types
- Genes that did not meet these thresholds were considered to have low specificity or broadexpression.

  We examine how risk genes are distributed across these categories in tissues, single-cell types, and brain datasets.

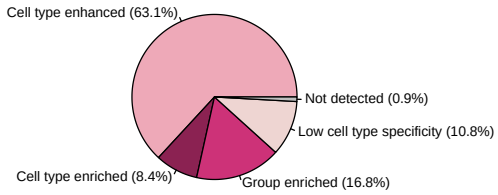
3a. Specificity component of risk genes

These plots show how many risk genes fall into each specificity class in different expression contexts.

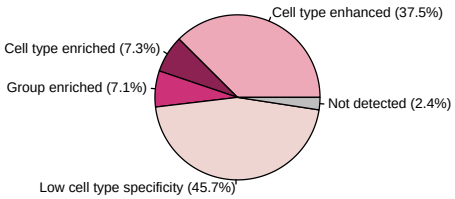
RNA.tissue.specificity



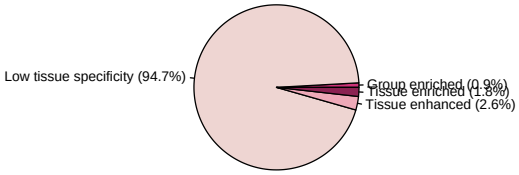
RNA.single.cell.type.specificity



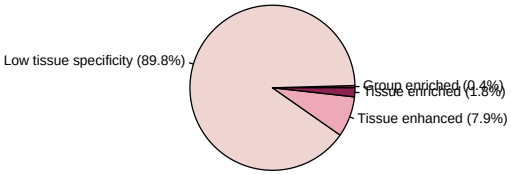
RNA.single.nuclei.brain.specificity



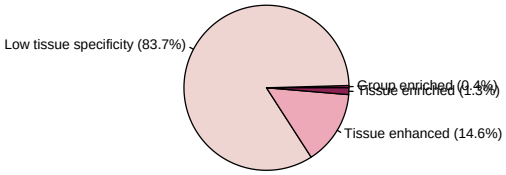
Human.brain.regional.RNA.Specificity



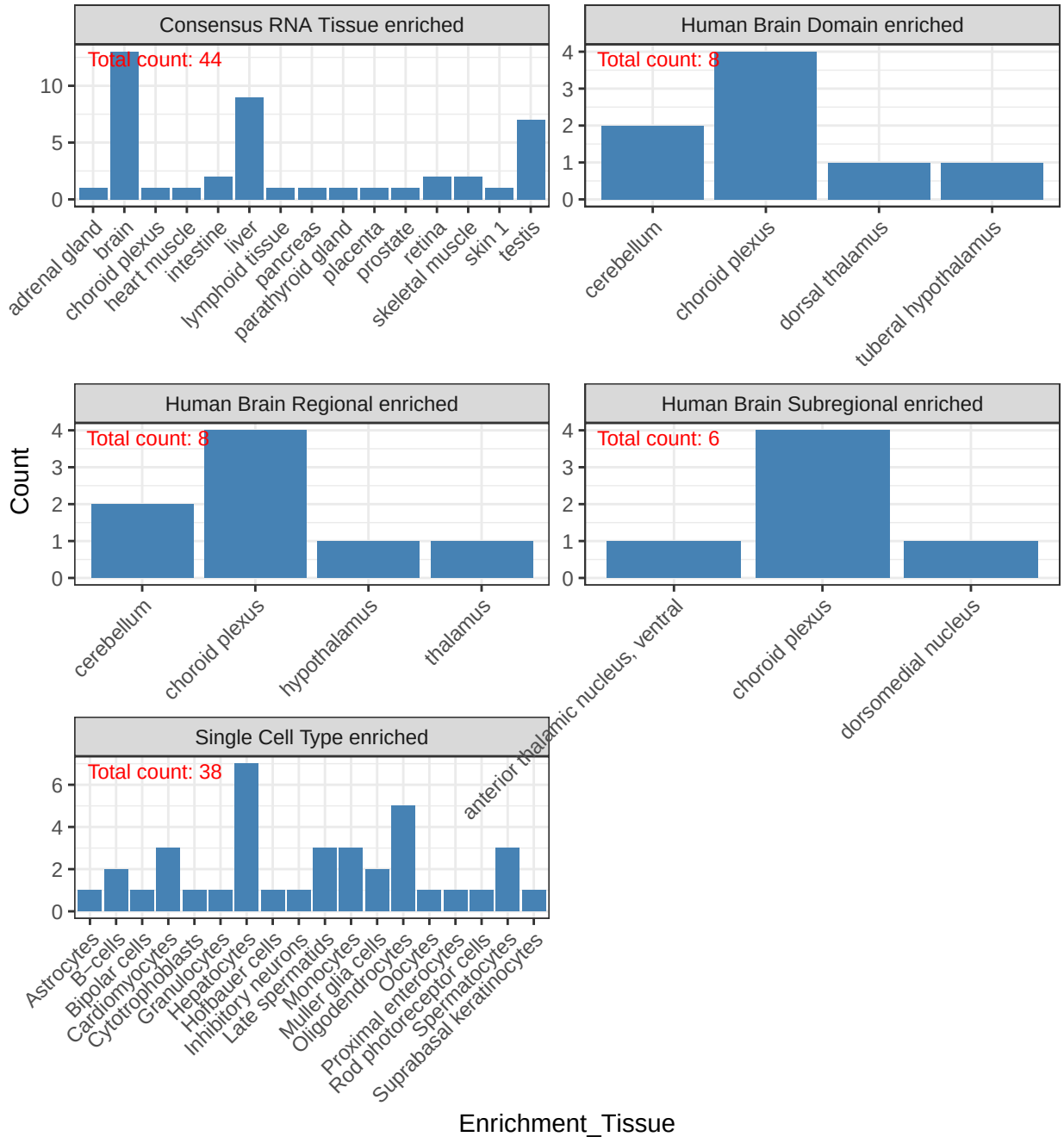
Human.brain.domain.RNA.Specificity

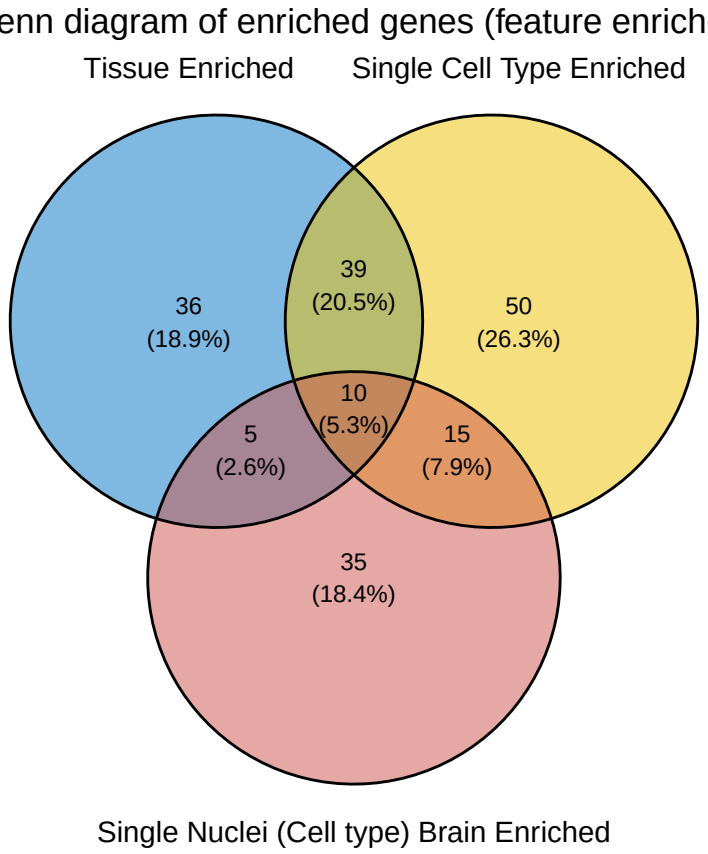


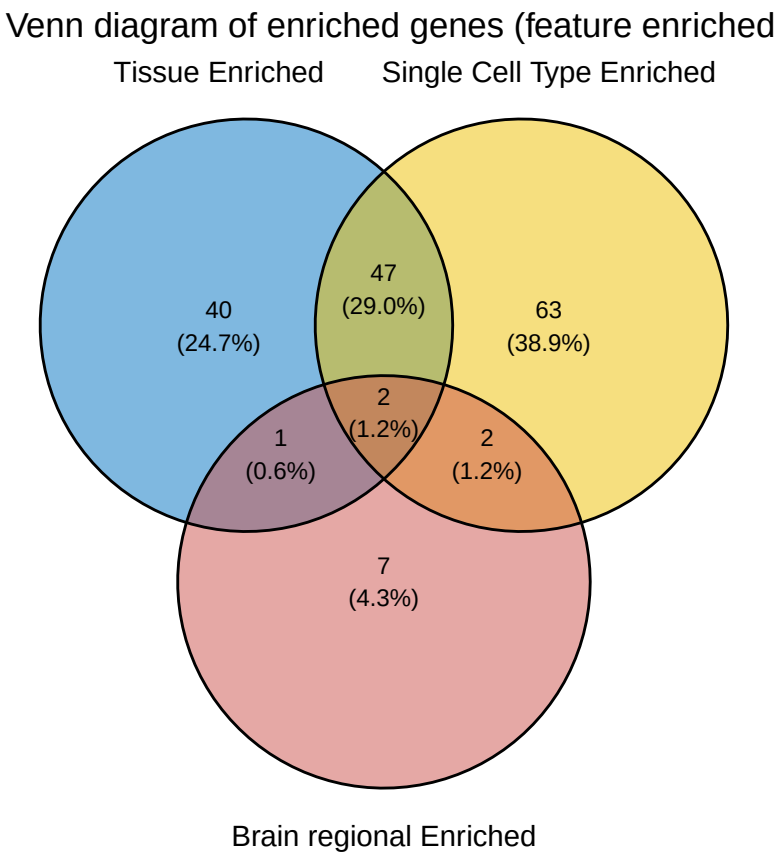
Human.brain.subregional.RNA.Specificity



3b. Detailed component of feature enriched genes (group enriched excluded)




Venn diagram of enriched genes (feature enriched+group enriche
Tissue Enriched
Single Cell Type Enriched
36
(18.9%)
39
(20.5%)
50
(26.3%)
10
(5.3%)
5
(2.6%)
15
(7.9%)
35
(18.4%)
Single Nuclei (Cell type) Brain Enriched


Venn diagram of enriched genes (feature enriched+group enriche
Tissue Enriched
Single Cell Type Enriched
40
(24.7%)
47
(29.0%)
63
(38.9%)
2
(1.2%)
1
(0.6%)
2
(1.2%)
7
(4.3%)
Brain regional Enriched