HSE University
Faculty of Computer science

Data analysis of single-cell sequencing

Study of brain tumor

Created by: Orlova Anastasia
FCS DSBA 212
Supervisor: Poptsova Maria Sergeevna
Head of the International Laboratory for Bioinformatics

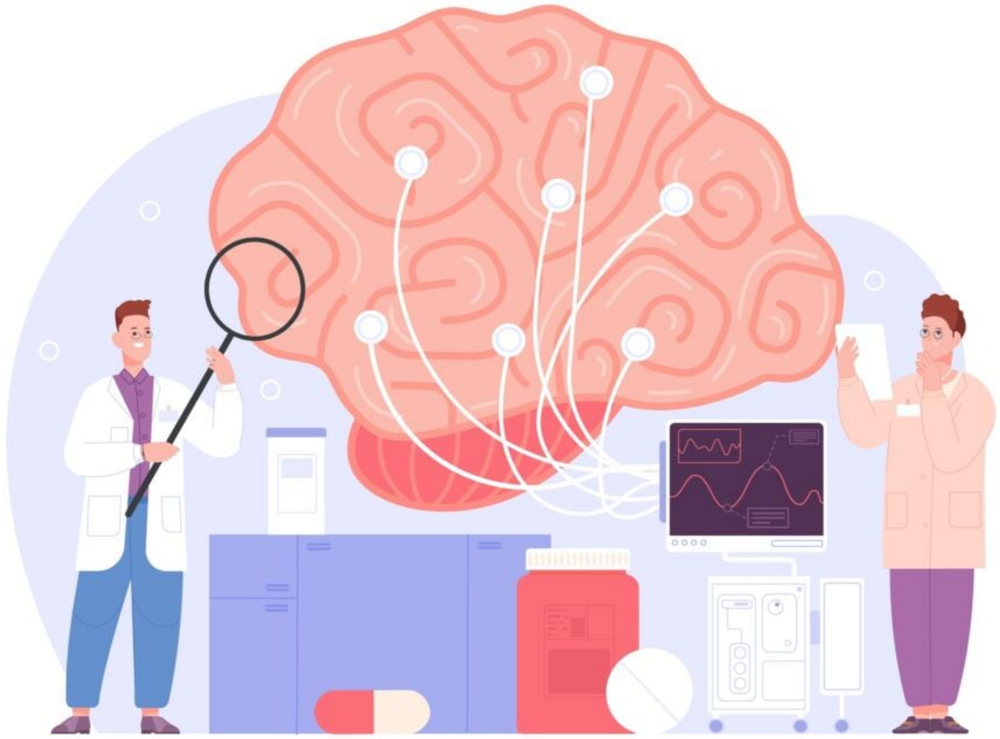
Data description and preprocessing

- One sample of high-grade glioma
- Gene Expression Omnibus (GEO) - source of data
- Conducting quality control (QC) to ensure the reliability of the analysis:

  - Removal of low-quality cells and artifacts

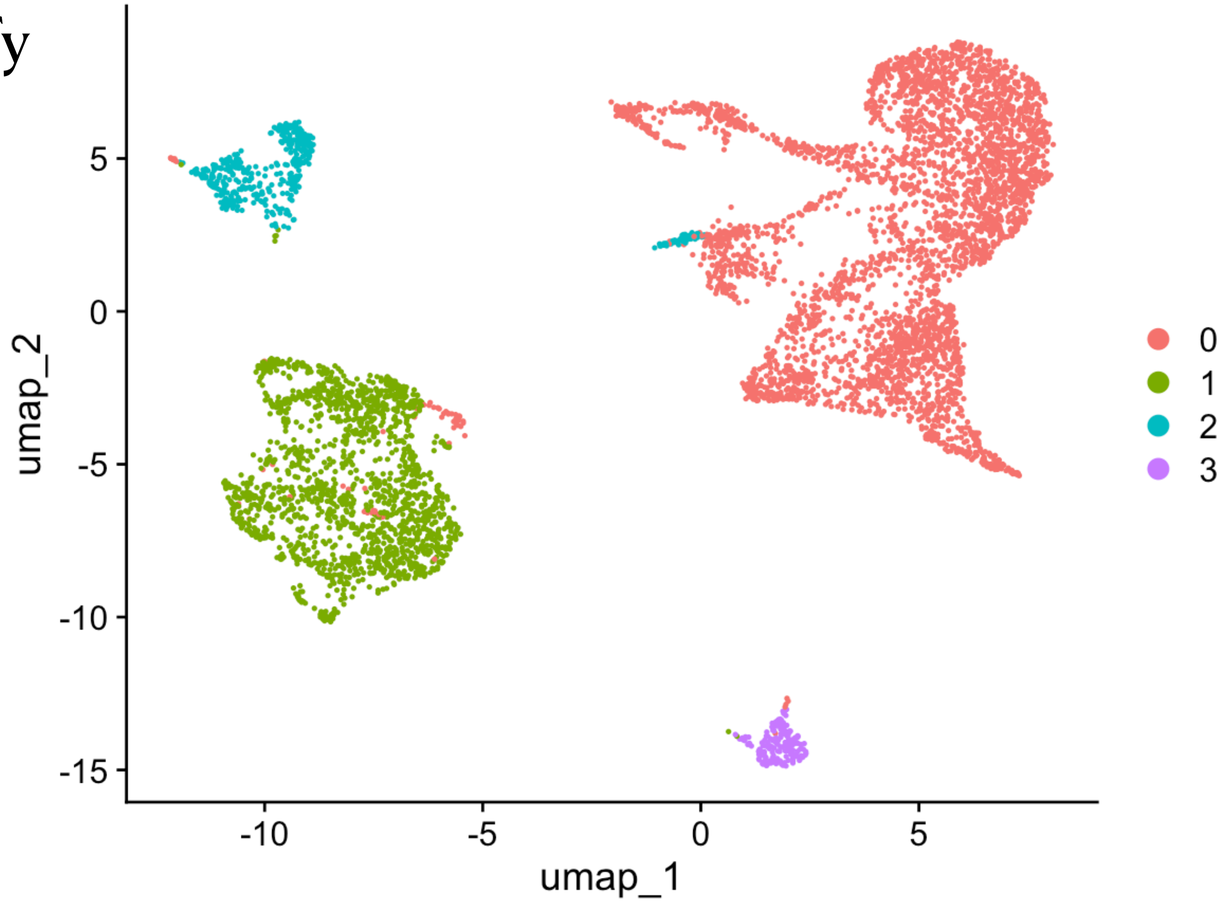
  - Key QC parameters: feature_RNA, nCount_RNA, percent.mt

  - Identification and removal of highly variable genes, including noise



Clusterization

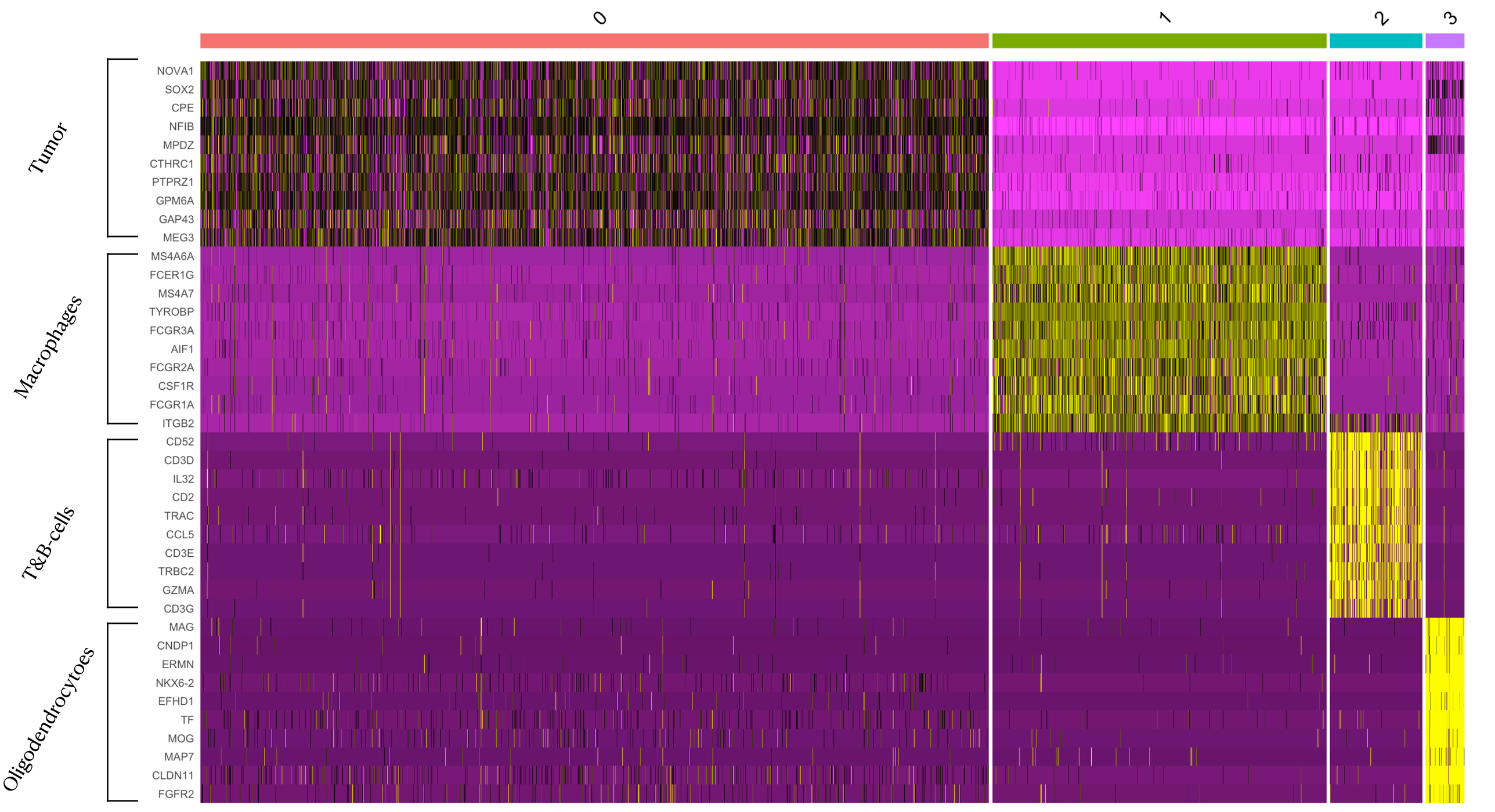
- Selection of highly variable genes to identify biologically significant signals
- Scaling the data to prepare for PCA
- PCA to reduce the dimension and identify the main sources of variation
- Clustering using the KNN and the Louvain algorithms
- Visualization of clusters in 2D using UMAP
- Getting four clusters for subsequent annotation



Defining cell types

Differential expression

- Selection of highly variable genes to improve accuracy
- Setting thresholds for the logarithmic change in expression (0.25) and the percentage of expressing cells (25%)
- Selection of the **top 15 marker genes** with a logarithmic change in expression >1 for each cluster
- Visualization of the results using **a heat map** to display the expression of marker genes in different clusters


0
1
2
3
Tumor
NOVA1
SOX2
CPE
NFIB
MPDZ
CTHRC1
PTPRZ1
GPM6A
GAP43
MEG3
Macrophages
MS4A6A
FCER1G
MS4A7
TYROBP
FCGR3A
AIF1
FCGR2A
CSF1R
FCGR1A
ITGB2
T&B-cells
CD52
CD3D
IL32
CD2
TRAC
CCL5
CD3E
TRBC2
GZMA
CD3G
Oligodendrocytoes
MAG
CNDP1
ERMN
NKX6-2
EFHD1
TF
MOG
MAP7
CLDN11
FGFR2

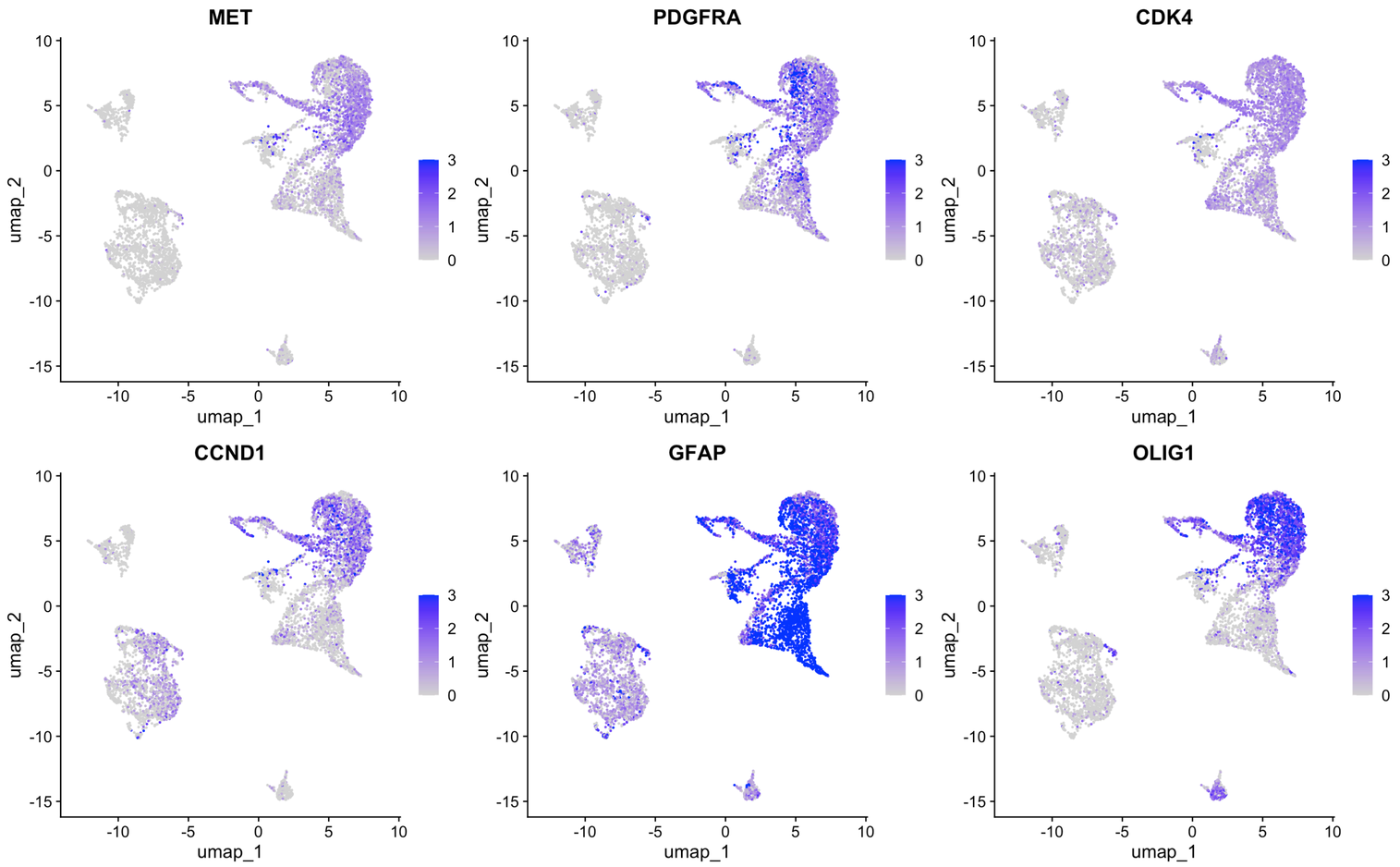
Intended annotations after DE

1. Cluster zero: A variety of genes indicating different cell types, possibly **cancer cells**
2. The first cluster: FCGR3A, AIF 1 (IBA 1) - markers of **macrophages**
3. The second cluster: CD3D, CD3E, TRAC, TRBC2 are markers of **immune cells**
4. The third cluster: MAG, MAG, CLDN11 are markers of **oligodendrocytes**

Tinting markers

We found well-known cancer markers using Network of Cancer Genes(NCG) and Human Gene Database websites and highlighted them on the cluster map

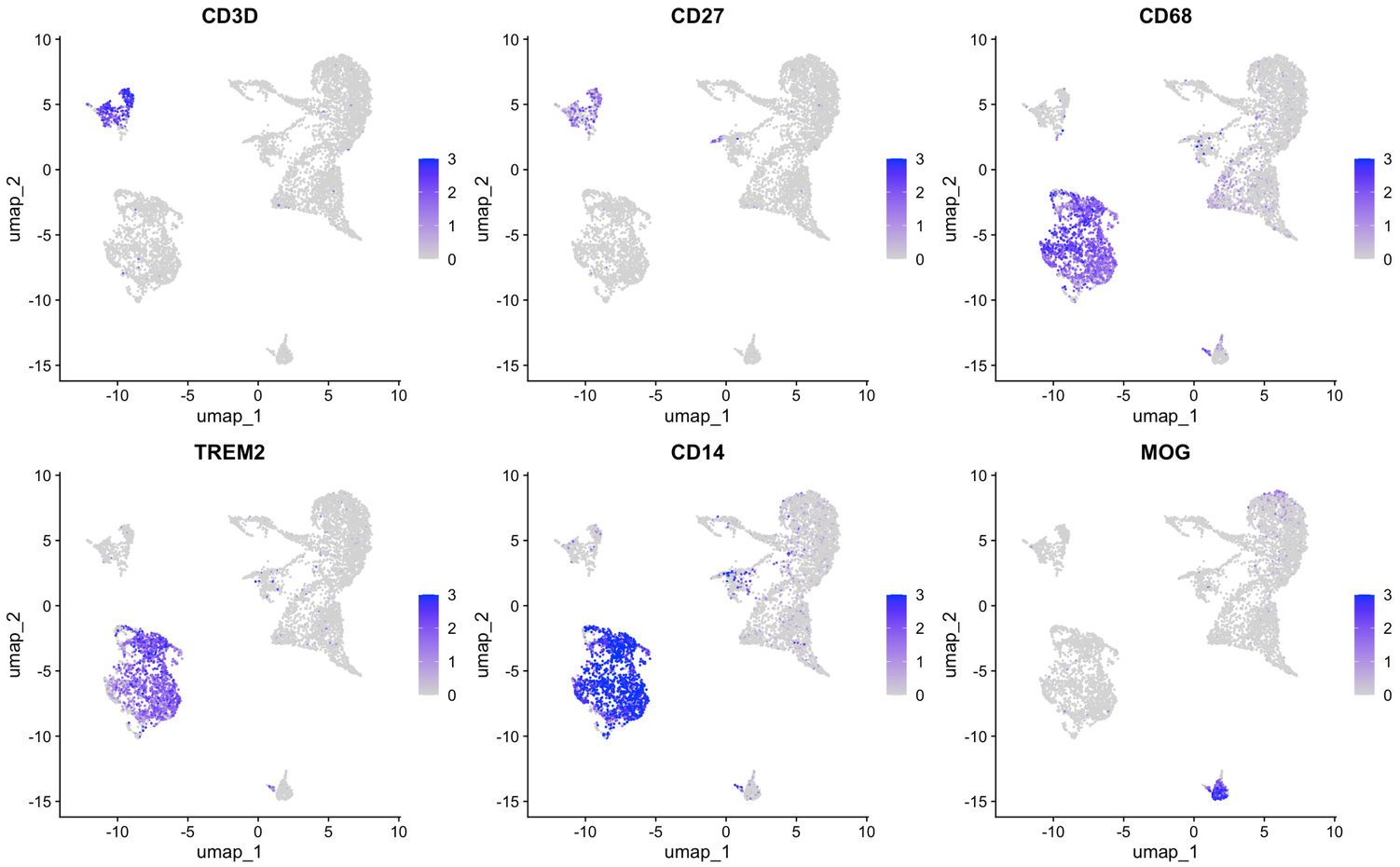
- High-grade glioma confirmed by overexpression of cancer markers: **MET, PDGFRA, CDK4, NF1, KRAS**.
- Expression of astrocyte markers (**GFAP**) and oligodendrocyte markers (**OLIG1-2**) suggests the tumor could be astrocytoma or oligodendroglioma.
- Co-expression of markers in cancer and immune cells: **CCND1-2-3 ->** Potential impact on tumor behavior.



Tinting markers

We found well-known cancer markers using Network of Cancer Genes(NCG) and Human Gene Database websites and highlighted them on the cluster map

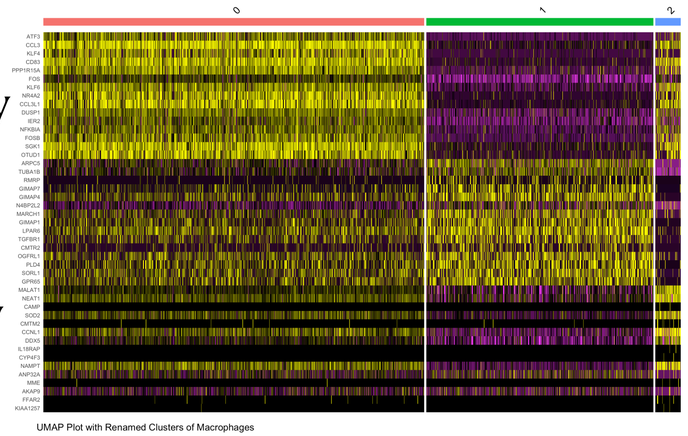
- Checked most well-known markers of other cell-types
- CD3D, CD27, etc - **T-cells & B-cells**
- CD68, CD14, TREM2, etc - **Macrophages**
- MOG - **oligodendrocytoes**



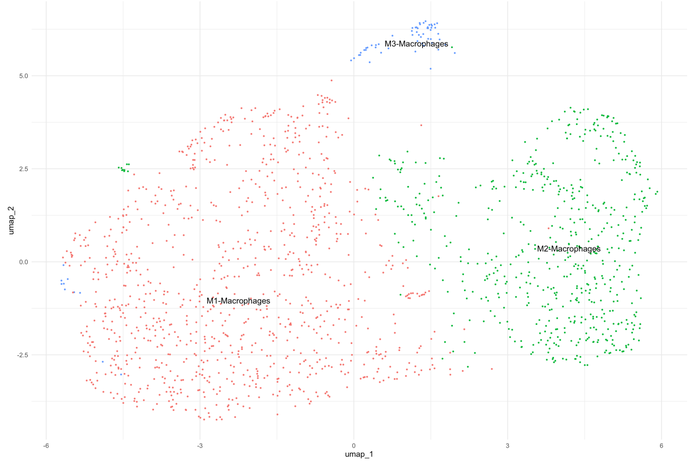
Go into each cluster individually

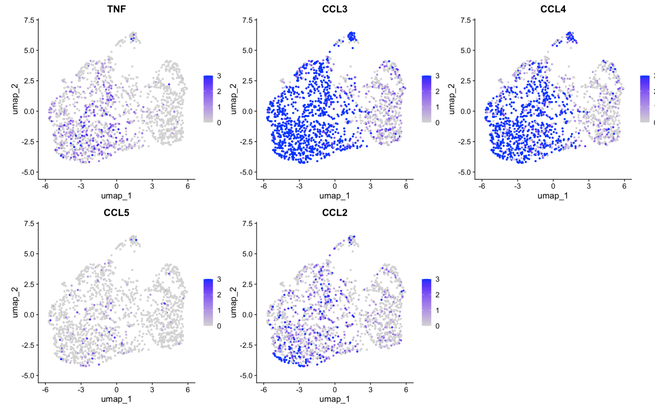
Macrophages cluster

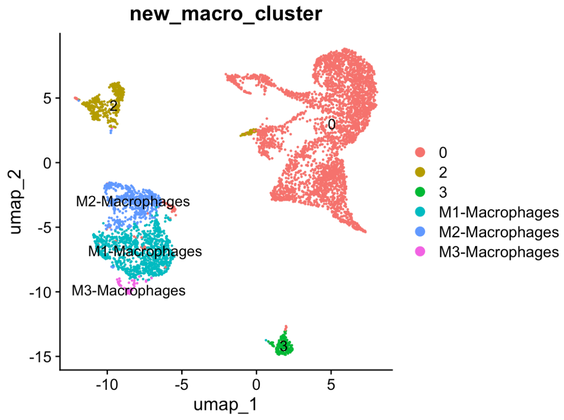
- Important to identify M1-M1 macrophages:
    1. **M1 Macrophages:** Anti-tumor, activated by pro-inflammatory signals, destroy cancer cells, associated with good prognosis.
    2. **M2 Macrophages:** Pro-tumor, activated by anti-inflammatory signals, promote blood vessel growth, suppress antitumor response, associated with poor prognosis.
- **Three subclusters identified**: M1, M2, M3 macrophages
- **MIF and PTEN**: Highly expressed in both macrophages and cancer clusters we suggest potential impact on cancer which confirmed by previous studies



UMAP Plot with Renamed Clusters of Macrophages

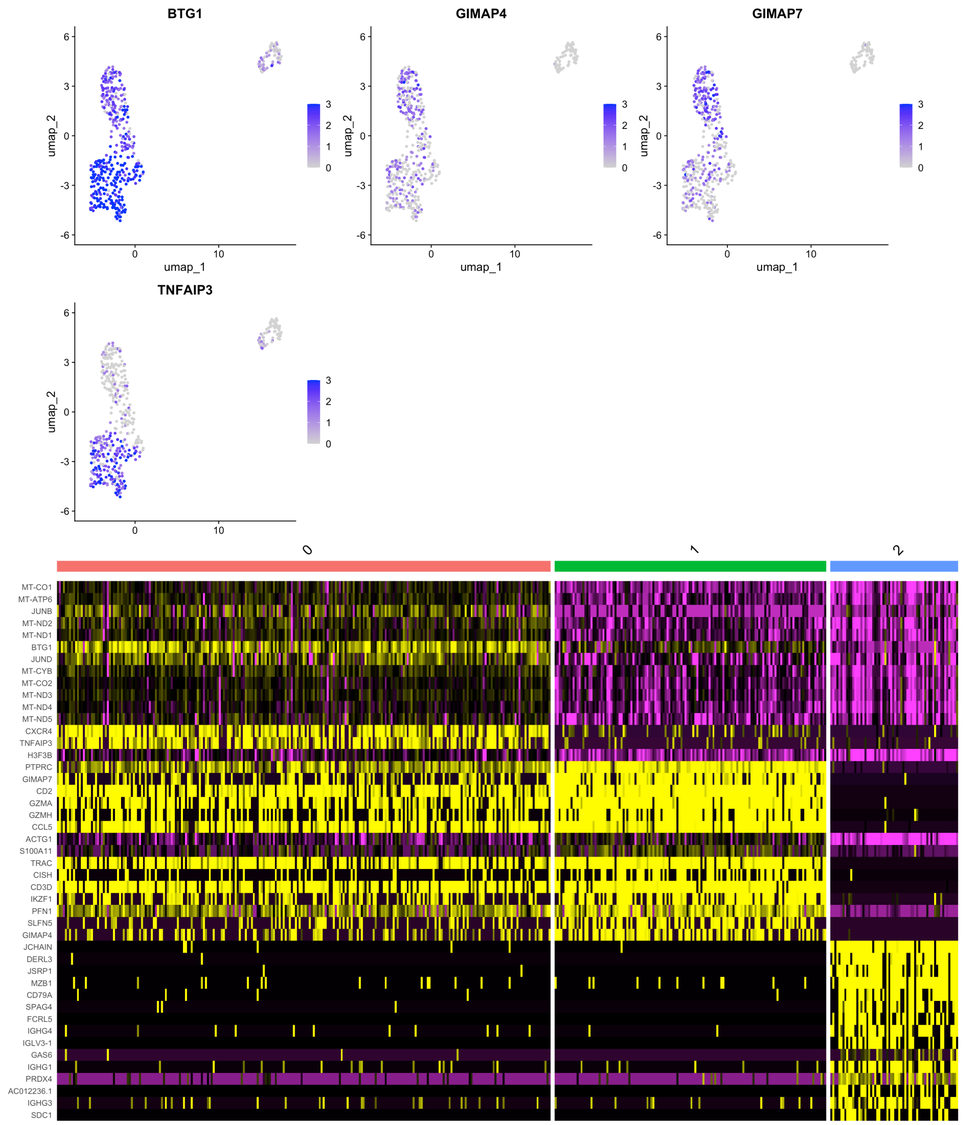






Immune cells cluster

- Important to identify T-cells and B-cells
    1. **T Cells:** Recognize and destroy cancer cells.
    2. **B Cells:** Produce antibodies against tumor antigens.
- **Three subclusters identified**: T-cells, B-cells, Plasma cells
- **BTG1 and TNFAIP3**: Highly expressed in both immune and cancer clusters we suggest potential impact on cancer which confirmed by previous studies



Cancer cells cluster

Cluster 0: **OPC** - Oligodendrocyte Precursors

- High expression of genes: OLIG1, OLIG2, SOX4, SOX8, GRIA2, BCAN.
- Functions: Intercellular communication and axon support.

Cluster 1: **Glial Support**

- High expression of genes: GFAP, VIM, IGFBP7, CLU, ID3, S100A10.
- Functions: Cell protection, survival, and differentiation regulation.

Cluster 2: **Active Proliferation**

- High expression of genes: TOP2A, TK1, BIRC5, CENPK, CENPM, CA12, OASL.
- Functions: Cell proliferation, division, and metabolic processes.

Cluster 3: **High Survival and Inflammation**

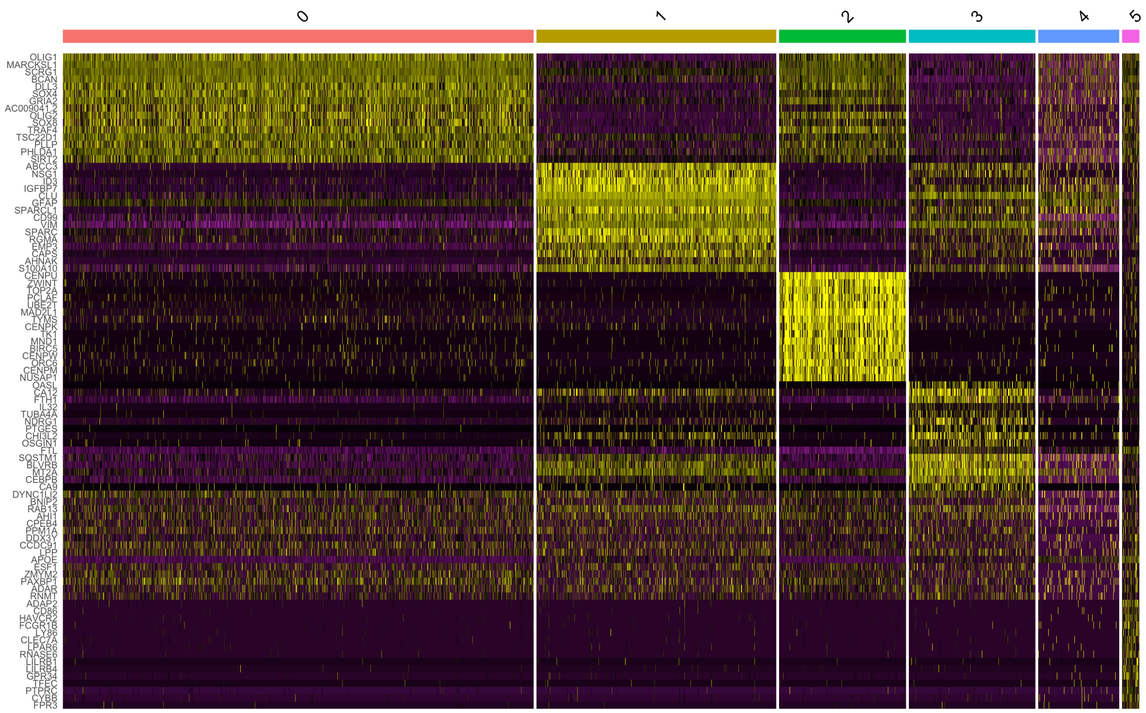
- High expression of genes: FTH1, FTL, IL32, NDRG1, SQSTM1, OSGIN1.
- Functions: Adaptation to stress and inflammatory responses.

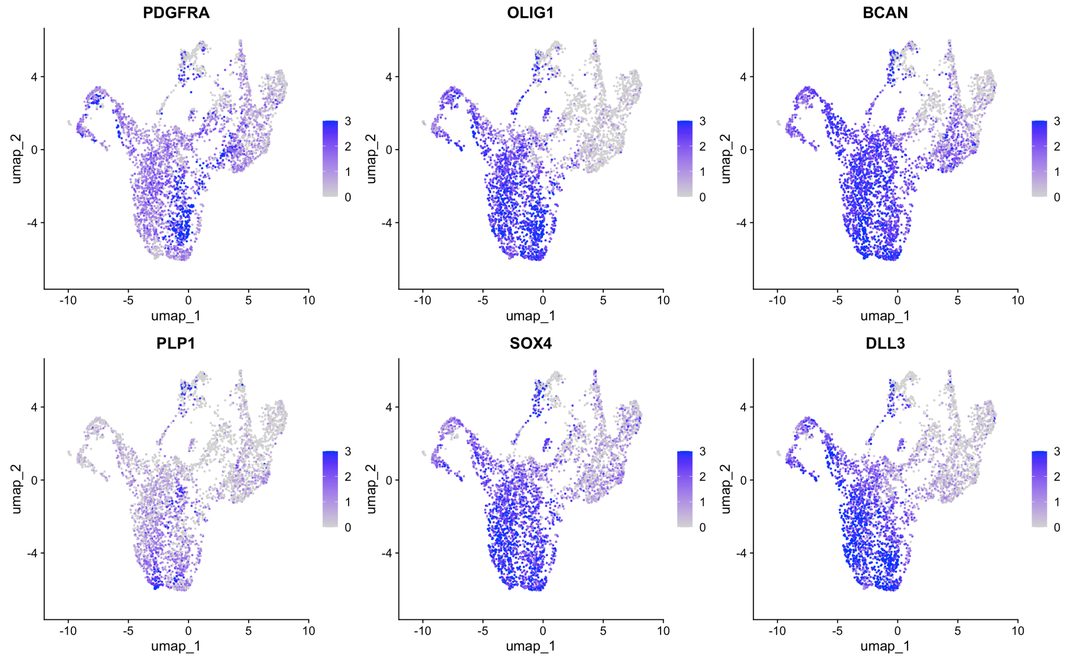
Cluster 4: **Migration and Adhesion**

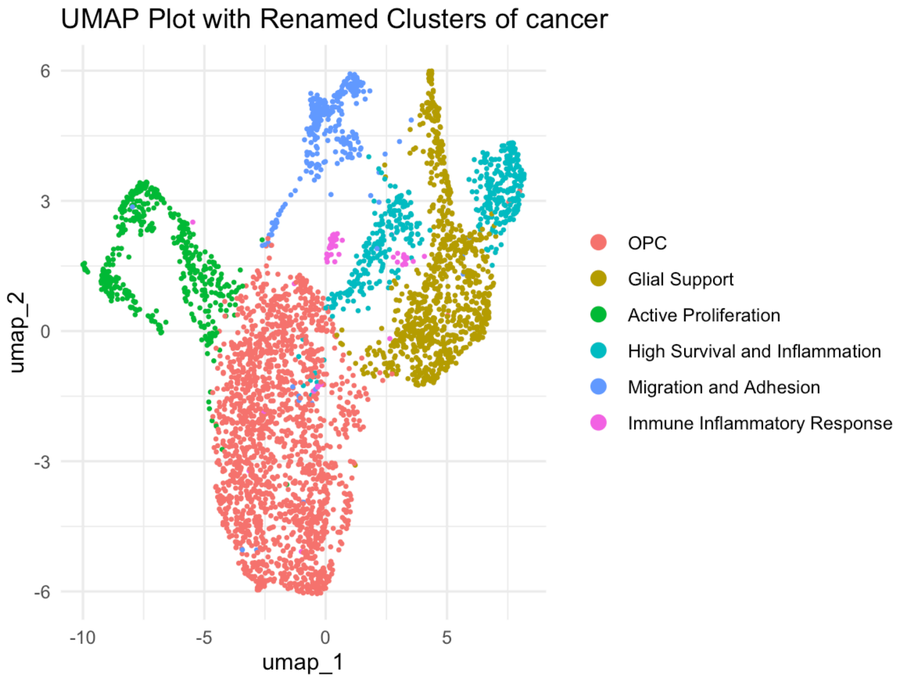
- High expression of genes: RAB13, LPP, APOE, CD86, ADAR.
- Functions: Cell migration, adhesion, and intercellular interactions.

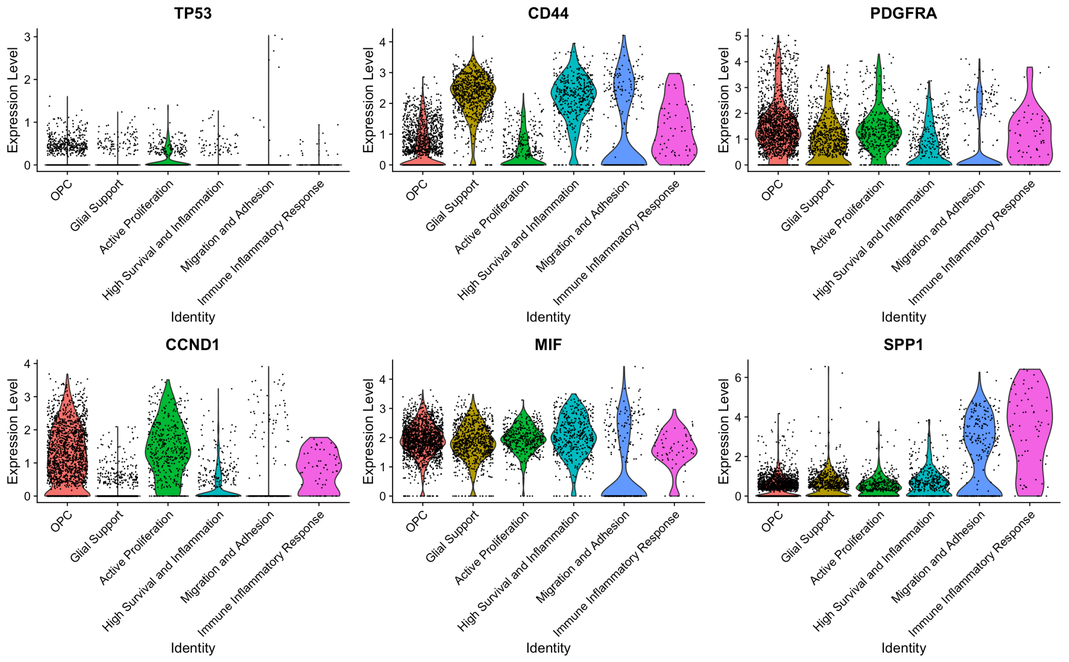
Cluster 5: **Immune Inflammatory Response**

- High expression of genes: HAVCR2, CLEC7A, FCGR1B, LILRB4, CYBB, RNASE6.
- Functions: Phagocytosis, antigen presentation, cellular stress, and metabolism regulation.

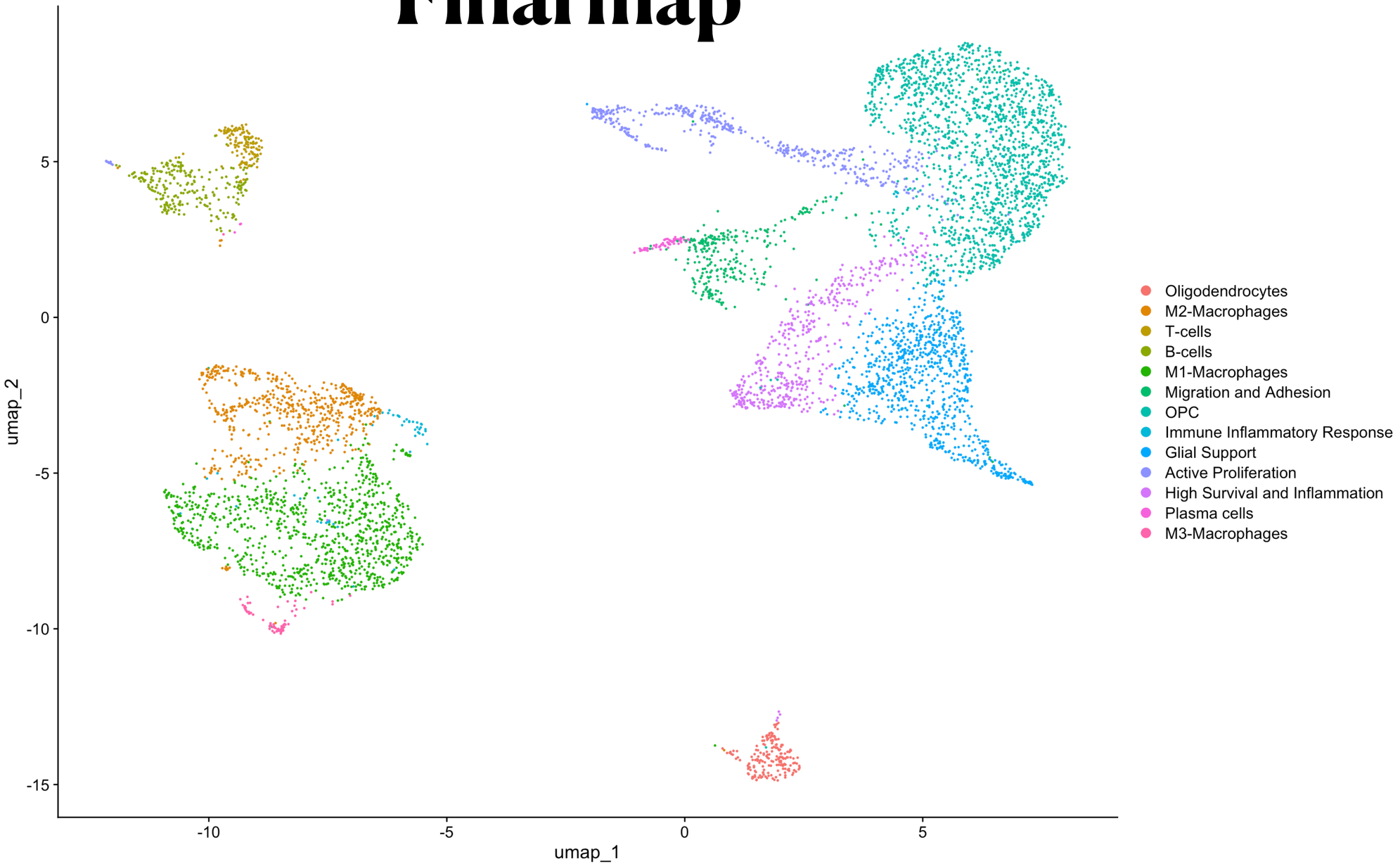

0
1
2
3
4
5


PDGFRA
OLIG1
BCAN
PLP1
SOX4
DLL3
umap_1
umap_2


UMAP Plot with Renamed Clusters of cancer
umap_1
umap_2
OPC
Glial Support
Active Proliferation
High Survival and Inflammation
Migration and Adhesion
Immune Inflammatory Response


TP53
CD44
PDGFRA
CCND1
MIF
SPP1
Expression Level
Identity
OPC
Glial Support
Active Proliferation
High Survival and Inflammation
Migration and Adhesion
Immune Inflammatory Response

Final map

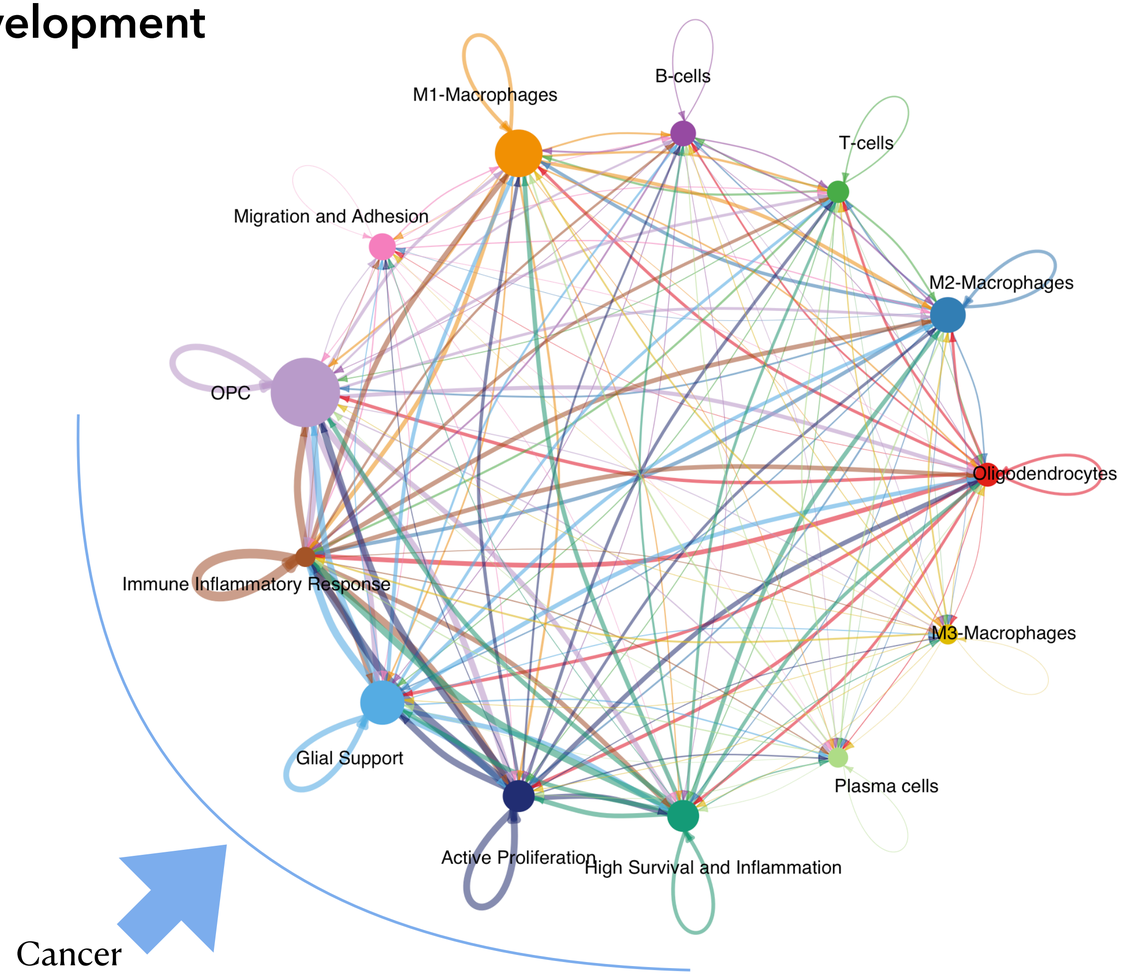

umap_2
umap_1
5
0
-5
-10
-15
-10
-5
0
5
Oligodendrocytes
M2-Macrophages
T-cells
B-cells
M1-Macrophages
Migration and Adhesion
OPC
Immune Inflammatory Response
Glial Support
Active Proliferation
High Survival and Inflammation
Plasma cells
M3-Macrophages

Cell-cell communication analysis

Network of intercellular communications

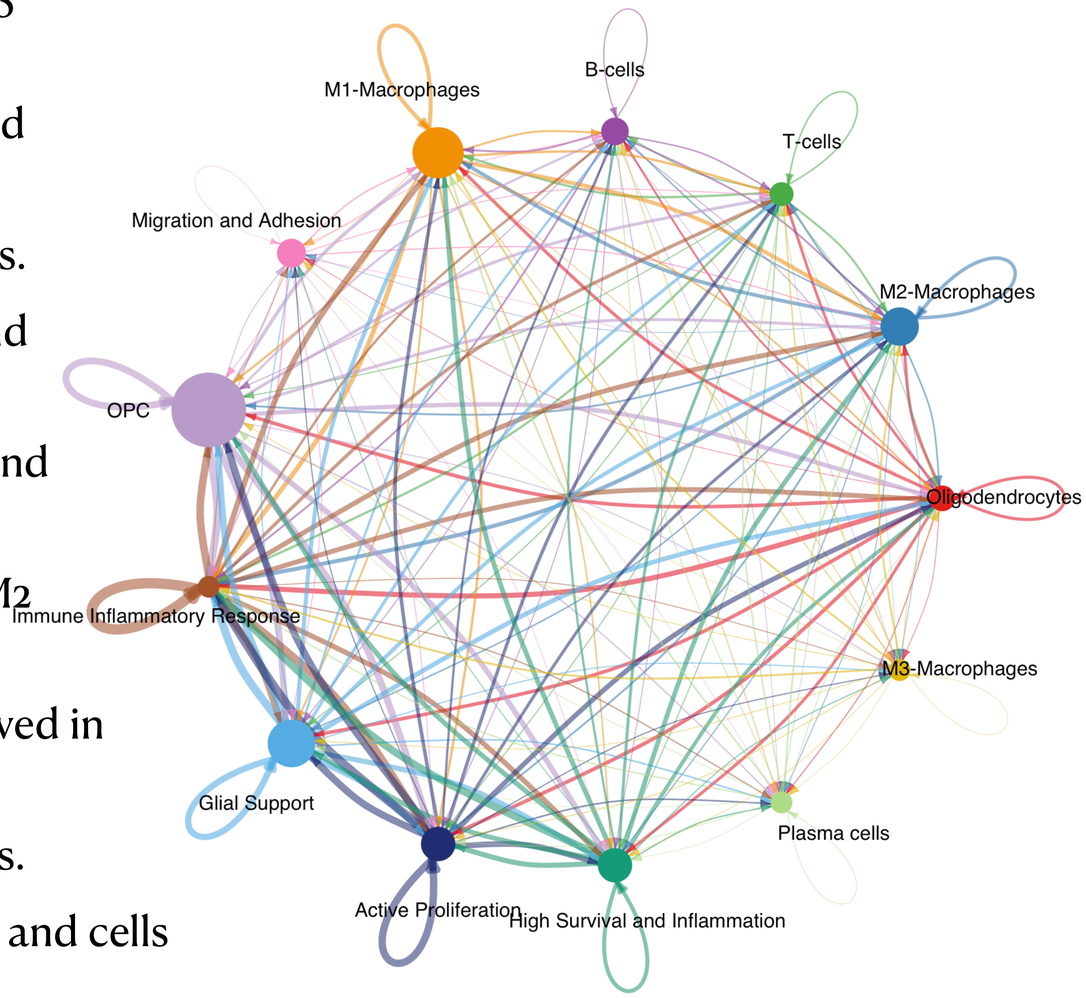
Investigate intercellular communication to understand interactions between different cell types: study mechanisms of diseases, immune responses, and cellular development

1. Created an object to analyze cell type-based communications.
2. Connected a human ligand-receptor database.
3. Isolated relevant interactions.
4. Identified overexpressed genes and interactions.
5. Reduced data dimensionality with PCA.
6. Calculated and filtered interaction probabilities.
7. Visualized networks of interactions between different cell types.



Network of intercellular communications

- **M1 Macrophages**: Interact with T cells, B cells, M2 macrophages, M3 macrophages, dendritic cells, and oligodendrocytes.
- **T Cells**: Interact with M1 macrophages, B cells, M2 macrophages, and oligodendrocytes.
- **B Cells**: Interact with T cells, M1 macrophages, and M2 macrophages.
- **M2 Macrophages**: Interact with M1 macrophages, T cells, B cells, and oligodendrocytes.
- **Dendritic Cells**: Interact with M1 macrophages, M2 macrophages, and oligodendrocytes.
- **Oligodendrocytes**: Interact with M1 macrophages, T cells, B cells, M2 macrophages, and dendritic cells.
- **OPCs (Oligodendrocyte Precursors)**: Interact with glial cells involved in support and inflammatory response.
- **Plasma Cells**: Interact with M1 macrophages and other immune cells.
- **Glial Support Cells**: Interact with OPCs, actively proliferating cells, and cells involved in survival and inflammation.

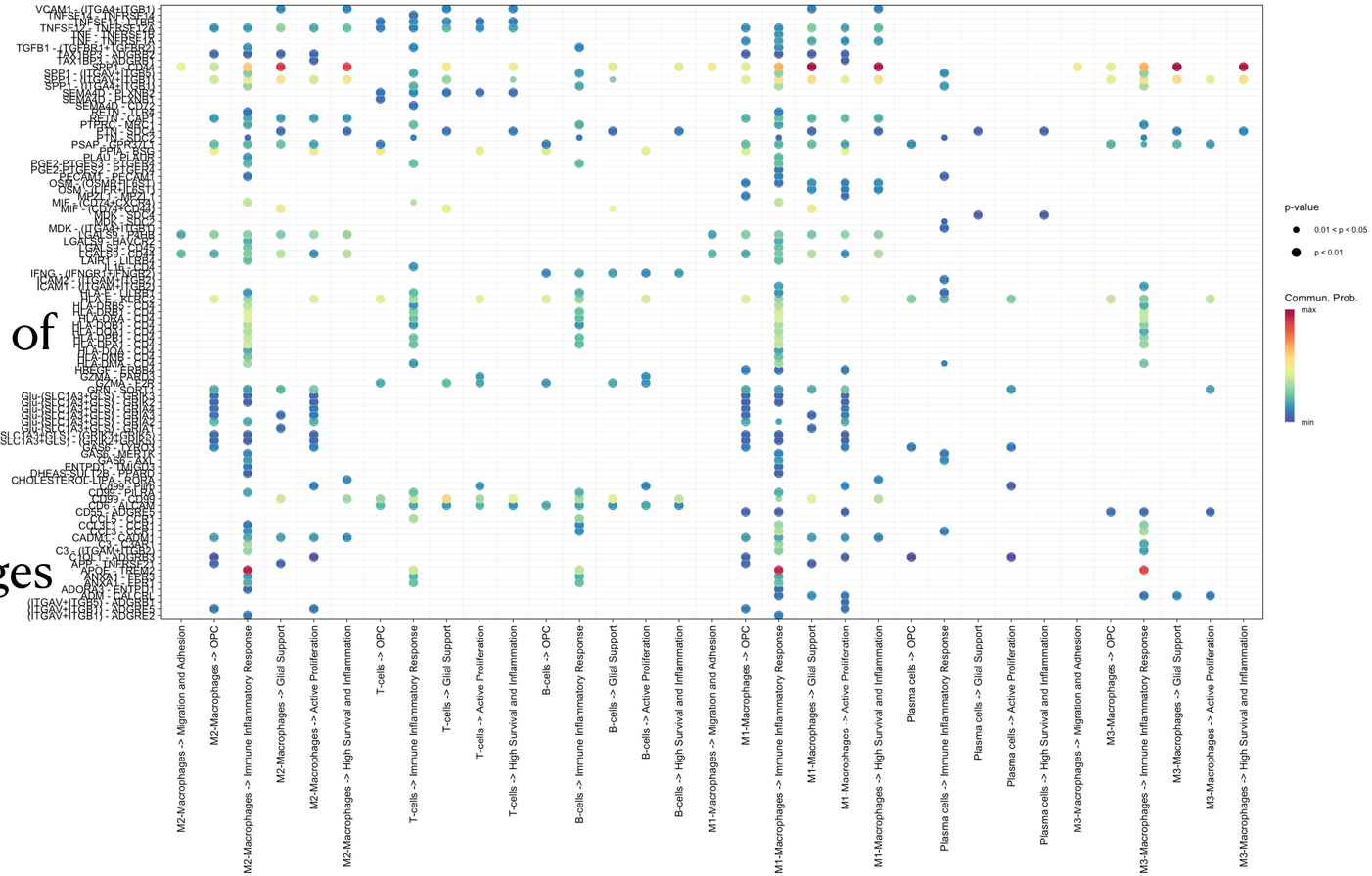


Signaling networks

Illustrate interactions between signal sources (stromal cells) and target cells (tumor cells)

Key Findings:

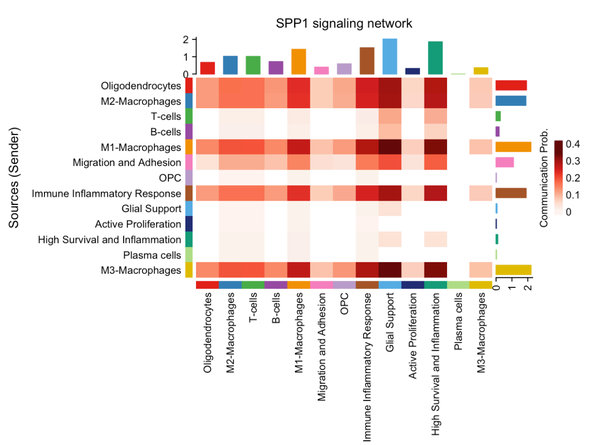
- Most Active Ligand-Receptor Pairs:
  - SPP1-CD44
  - APOE-TREM2
  - These pairs have the highest probability of communication.
- High Communication:
  - Notable interaction between macrophages and cancer cells.
- Significant Effect:
  - **MIF - (CD74+CD44)** pair impacts cell communication substantially.

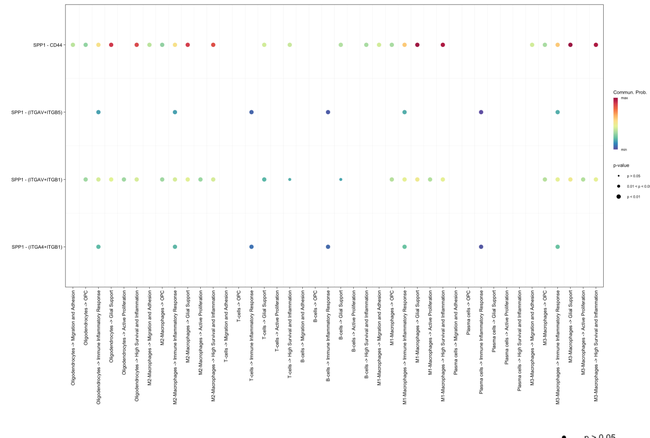


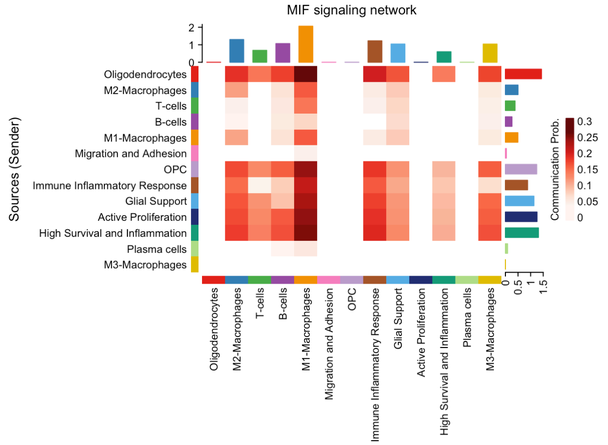
Signaling networks

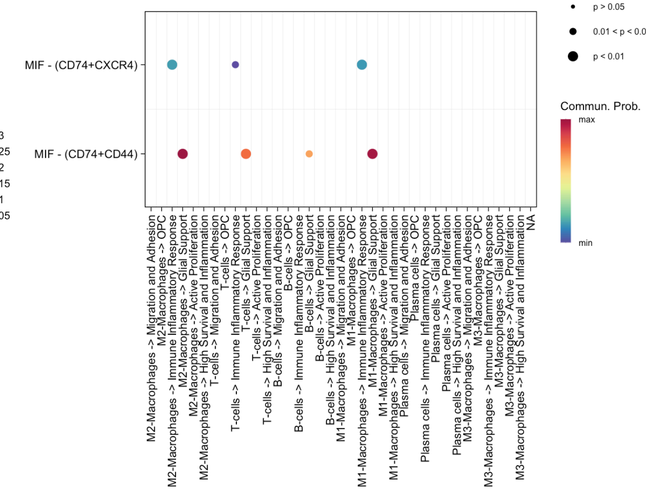
Detailed Examination of Key Ligands: SPP, MIF, APOE

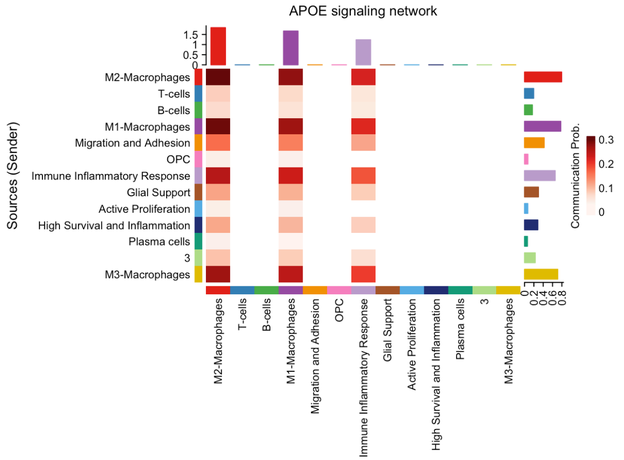
- **Bubble diagrams** and **signaling network graphs** to assess the impact of these ligands on cellular communication
- Key Findings:
    1. **MIF - (CD74+CD44):**
        - Pronounced interaction between M1 and M2 macrophages.
        - Significant effect on "Glial support" cells.
    2. **SPP1-CD44:**
        - Affects stromal cells' interaction with "Glial support" cells and other cancer subtypes.
    3. **APOE:**
        - Communication network reveals interactions only between macrophages and the "Immune Inflammatory Response".

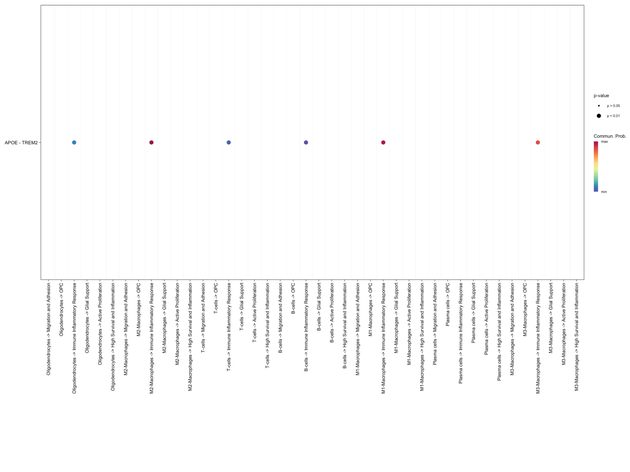








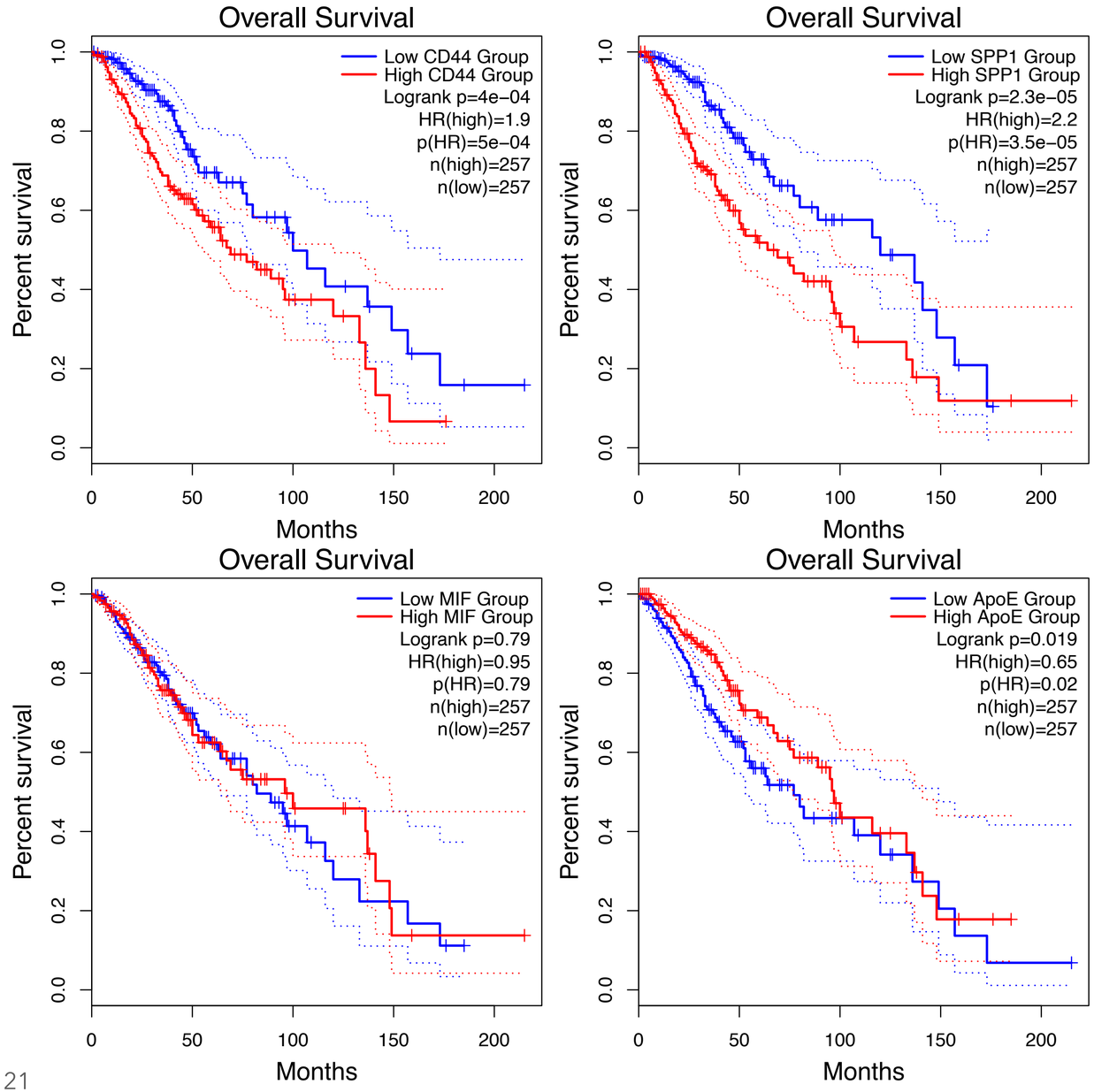




Survival analysis

Kaplan-Meier Curves

- High **SPP1, CD44** expression correlate with a **shorter** overall survival time of glioma patients
- High **MIF, APOE** expression correlate with a **longer** overall survival time of glioma patients
- **TNFAIP3 and BTG1** graphs showed that there was **no significant difference in survival** between the group where these genes were highly expressed and low



Main results

1. We confirmed previous research and demonstrated the impact of certain genes on the tumor: CDND1-2-3, MIF, PTEN, BTG1 and TNFAIP3.
2. We found that cancer cells in this sample are divided into 5 subpopulations, distinct in functions.
3. We discovered a significant influence of the ligand-receptor pair SPP1-CD44 on cell-cell interaction.
4. SPP1-CD44 negatively affects the lifespan of patients with glioma.
5. Ligands MIF, APOE also significantly affect cell-cell interaction, while increasing lifespan in diagnosed glioma.
6. Suggested that these macrophages more likely play pro-tumor roles in glioma.

List of references

https://satijalab.org/seurat/articles/get_started_v5_new

https://github.com/sqjin/CellChat

http://network-cancer-genes.org/

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185231

https://pubmed.ncbi.nlm.nih.gov/34805184/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7751482/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6472943/

https://link.springer.com/article/10.1007/s00262-020-02616-6

http://gepia2.cancer-pku.cn/#survival

https://www.nature.com/articles/s43018-022-00475-x

https://www.genecards.org/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8602110/

https://pubmed.ncbi.nlm.nih.gov/21088135/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8858002/

https://www.genecards.org/cgi-bin/carddisp.pl?gene=BTG1

https://cancerci.biomedcentral.com/articles/10.1186/s12935-021-02089-2

https://www.nature.com/articles/s41417-022-00582-y

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10774283/

https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/34789849/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10236634/

https://www.nature.com/articles/s41467-022-28372-y

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9249016/

https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2020.581209/full

Thank you