Neuro-genomics - Final Project

Avital Markus & Noa Cohen

---

https://github.com/TaliVas/neurogenomics-final-project.git

PART 1 - analysis of bulk sequencing

Objective & Background

In this part of the project, we performed a bulk RNA-seq analysis to study gene expression differences between control mice and knockout mice in cortex tissue. The goal of the analysis was to identify which genes are differentially expressed between the two conditions. Knockout mouse models are commonly used in neuroscience research to study the role of specific genes in brain function and disease. By comparing gene expression in knockout mice versus healthy controls, we can identify the biological pathways affected by the loss of a gene - and gain insights into the molecular basis of neurological conditions.

Input data

In this experiment, we had 6 mouse cortex samples: 3 control mice (C1, C2, C3) and 3 knockout mice (KO1, KO2, KO3). The raw data was provided as FASTQ files.

Software & Tools

The analysis was performed in Python. Transcript quantification was done using Kallisto, with the mouse reference transcriptome from Ensembl. Differential expression analysis was performed using pyDESeq2, a Python implementation of the DESeq2 method. Functional analysis was performed using g:Profiler for gene set enrichment and GeneCards for individual gene annotation.

Workflow

1. Indexing with Kallisto - before we could quantify gene expression, we needed to build a reference index. We used Kallisto, a pseudoalignment tool, to process the FASTQ files. Kallisto works by comparing each sequencing read to a reference transcriptome which is a database of all known mRNA sequences in the mouse genome. We downloaded the mouse reference transcriptome from Ensembl (GRCm39, release 110). We then built a Kallisto index from this file using the following command:

   kallisto index -i mouse_transcriptome.idx Mus_musculus.GRCm39.cdna.all.fa.gz

   This index file allows Kallisto to rapidly match reads from our samples to known transcripts.

2. Quantification with Kallisto - for each of the 6 samples, we ran Kallisto to estimate how many reads mapped to each transcript. Since our data was single-end sequencing, we specified the single flag along with the estimated fragment length and standard deviation

   kallisto quant -i mouse_transcriptome.idx -o C1_results --single -l 300 -s 50 C1.fastq

   Kallisto produced an abundance.tsv file for each sample, containing estimated counts and tpm values for every transcript.

3. Transcript-to-Gene Aggregation - Kallisto produces counts at the transcript level, but genes can produce multiple transcripts through alternative splicing. For differential expression analysis, we need counts at the gene level.
   We used the Ensembl GTF annotation file to create a mapping between each transcript ID and its corresponding gene name. We then summed all transcript counts belonging to the same gene, producing a gene-level count matrix with 6 samples (columns) and all detected genes (rows).

4. Differential Expression Analysis with DESeq2 - after aggregating counts to the gene level, we performed differential expression analysis using pyDESeq2 - a Python implementation of the DESeq2 method.
   We provided DESeq2 with the sample metadata (which samples are control and which are knockout) and defined the experimental design: comparing the control condition versus the knockout condition.
   DESeq2 performs the following steps:
   a. Normalization - corrects for differences in sequencing depth between samples.

b. Statistical modeling - models the count data using a negative binomial distribution.
c. Fold change calculation - computes the log2 fold change for each gene, representing how much more or less a gene is expressed in control vs. knockout.
d. Multiple testing correction - because we test thousands of genes simultaneously, DESeq2 applies the Benjamini-Hochberg correction to control the false discovery rate, producing an adjusted p-value (padj) for each gene.

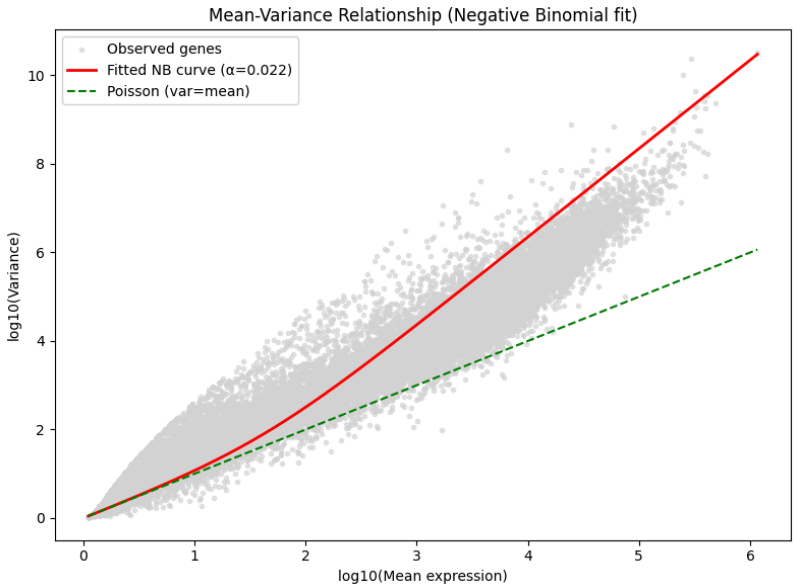
We defined a gene as significantly differentially expressed if its adjusted p-value was below 0.05 (**padj < 0.05**). Since the analysis was set up as control vs. knockout, a positive log2 fold change means the gene is more active in the control mice, while a negative log2 fold change means the gene is more active in the knockout mice.

5. DESeq2 Assumption Verification - before interpreting the results, we verified that the data meets DESeq2's core assumption - that gene expression counts follow a negative binomial distribution. This distribution assumes that variance grows faster than the mean (overdispersion).

6. Functional Analysis - after identifying the differentially expressed genes, we performed functional analysis to understand the biological meaning of our results. This was done using online databases and tools:
    a. GeneCards - we looked up each of the top differentially expressed genes individually on GeneCards, a comprehensive database that provides information about gene function, associated diseases, expression patterns, and known biological roles.
    b. g:Profiler - we used g:Profiler to perform gene set enrichment analysis. This tool takes a list of genes and asks: do these genes share a common biological function? We ran this analysis separately for each group of differentially expressed genes.

Results

DESeq2 Verification



We plotted the mean expression vs. variance for all genes and fitted a negative binomial curve using least squares: variance = mean + α × mean², where α is the dispersion parameter. The fitted curve (α = 0.022) closely follows the observed data, and lies clearly above the Poisson line (variance = mean), confirming that the data is overdispersed. This validates the use of DESeq2 for our differential expression analysis.

Differential Expression Results

49 genes were found to be significantly differentially expressed between control and knockout mice (padj < 0.05).
These were divided into two groups:

Group A (26 genes) - higher expression in control mice, meaning these genes are **downregulated** in the knockout.
The 10 most significant are shown in the table below.

Gene	log2FC	padj
Gtf2i	0.70	1.85e-27
Neurod6	0.81	5.15e-16
Cnp	0.43	1.14e-07
Mag	0.45	4.05e-07
Mog	0.43	6.15e-06
Myl4	0.76	7.09e-06
Fibcd1	0.64	1.34e-05
Mobp	0.57	8.69e-05

Cldn11	0.39	1.22e-04
Plp1	0.36	3.53e-04

Group B (23 genes) - higher expression in knockout mice, meaning these genes are **upregulated** in the knockout.
The 10 most significant are shown in the table below.

Gene	log2FC	padj
Lrg1	-3.85	7.28e-293
Amz1	-0.87	4.94e-25
Dmgdh	-2.22	1.61e-22
Tfr2	-0.87	2.93e-09
Myo18b	-1.17	4.79e-09
Ppp1ccb	-1.05	1.58e-07
Creb3l3	-2.06	2.60e-07
Ccdc42	-1.28	5.03e-06
Drc1	-0.76	1.27e-05
Hapln4	-0.50	3.51e-05

Gene Analysis

Group A

1. GTF2I (General Transcription Factor IIi) is a transcription factor that regulates gene expression by binding to promoter regions and controlling transcription initiation. It is associated with Williams-Beuren Syndrome, a neurodevelopmental disorder caused by deletion of this gene.
2. NEUROD6 is a transcription factor involved in the development and differentiation of neurons.
3. CNP (2',3'-Cyclic Nucleotide Phosphodiesterase) is one of the most abundant proteins in myelin. It is associated with multiple sclerosis and hypomyelinating leukodystrophy, both diseases involving damage to myelin.
4. MAG (Myelin Associated Glycoprotein) is a protein that mediates communication between myelin-producing cells and neurons, and plays a role in maintaining the myelin sheath around axons.
5. MOG (Myelin Oligodendrocyte Glycoprotein) is a protein located on the outer surface of the myelin sheath and is a key target in immune-mediated demyelinating diseases.

6. MYL4 (Myosin Light Chain 4) is a motor protein involved in muscle contraction and cytoskeleton organization, primarily expressed in embryonic muscle and heart tissue.
7. FIBCD1 is a transmembrane receptor that binds acetylated molecules and helps cells take them up through endocytosis.
8. MOBP (Myelin Associated Oligodendrocyte Basic Protein) is a structural component of the myelin sheath that helps compact and stabilize myelin in the CNS.
9. CLDN11 (Claudin 11) is a tight junction protein that is a major component of CNS myelin and plays an important role in regulating oligodendrocyte function.
10. PLP1 (Proteolipid Protein 1) is the most abundant protein in CNS myelin and plays a critical role in the formation, compaction, and maintenance of the myelin sheath.

The majority of genes downregulated in the knockout mice are directly involved in **myelination** - the process by which oligodendrocytes wrap axons with a protective myelin sheath in the CNS. The coordinated downregulation of those genes suggests that the knockout disrupts oligodendrocyte function and myelin production in the cortex. GTF2I and NEUROD6 are transcription factors involved in neural development, indicating that the knockout also affects broader gene regulation in the brain.

Group B

1. LRG1 (Leucine Rich Alpha-2-Glycoprotein 1) is a protein involved in cell signaling and the innate immune response, and is associated with inflammatory processes.
2. AMZ1 (Archaelysin Family Metallopeptidase 1) is a metallopeptidase enzyme involved in protein degradation.
3. DMGDH (Dimethylglycine Dehydrogenase) is a mitochondrial enzyme involved in choline metabolism, and mutations in this gene cause dimethylglycine dehydrogenase deficiency.
4. TFR2 (Transferrin Receptor 2) is a membrane protein that mediates cellular uptake of iron, and is associated with hereditary hemochromatosis, a disease of iron overload.
5. MYO18B (Myosin XVIIIB) is a motor protein involved in intracellular trafficking and regulation of muscle-specific genes, and has been linked to tumor suppression.
6. Ppp1ccb (Protein Phosphatase 1 Catalytic Subunit Gamma B) is a serine/threonine phosphatase involved in regulating circadian rhythm and is part of a larger phosphatase complex in the cell.

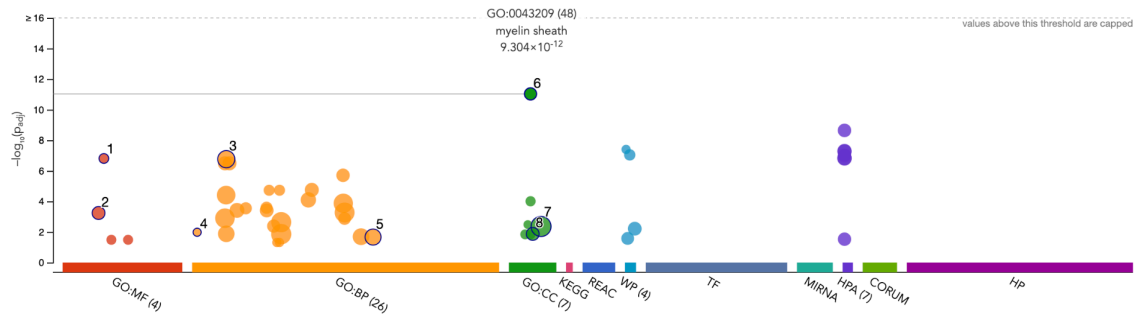
7. CREB3L3 is a transcription factor activated during cellular stress and inflammation, and plays a role in regulating triglyceride metabolism and acute phase response genes.
8. CCDC42 is a protein essential for sperm development and cilium assembly, primarily expressed in reproductive tissue.
9. DRC1 (Dynein Regulatory Complex Subunit 1) is a component of the ciliary dynein complex that regulates ciliary movement, and mutations in this gene cause primary ciliary dyskinesia.
10. HAPLN4 (Hyaluronan And Proteoglycan Link Protein 4) is an extracellular matrix protein involved in the formation of perineuronal nets - specialized structures that surround neurons and regulate synaptic transmission.

The genes upregulated in the knockout mice form a more diverse group compared to Group A. The most striking finding is LRG1, which showed the strongest signal in the entire dataset and is involved in innate immune signaling, suggesting an **inflammatory response** in the knockout cortex.
HAPLN4 is directly related to brain function, playing a role in perineuronal net formation and GABAergic synaptic transmission.
CREB3L3 and PPP1CCB are involved in cellular stress responses and signal regulation. The remaining genes are involved in diverse processes including iron uptake, metabolism, cytoskeletal organization, and ciliary function, and their upregulation may reflect broader cellular stress or compensatory responses to the loss of the knocked-out gene.

g:Profiler Enrichment Analysis

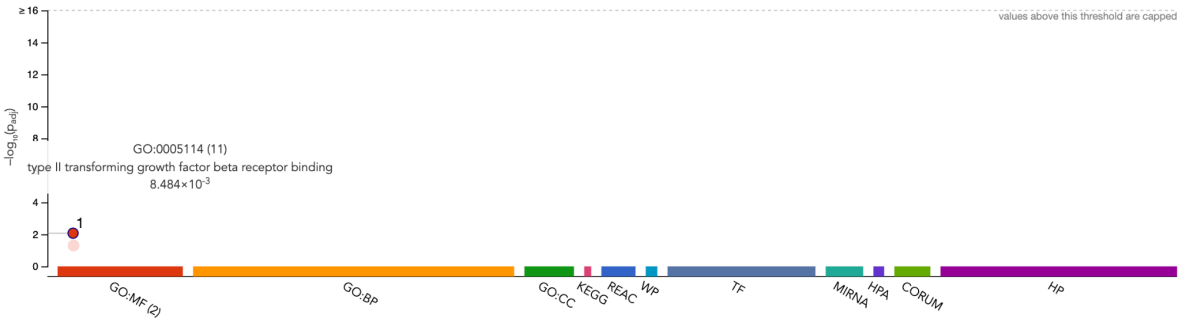
We performed gene set enrichment analysis using g:Profiler separately for Group A and Group B genes, with the organism set to Mus muscu§lus.

Group A showed highly significant enrichment for terms directly related to myelination and oligodendrocyte biology.These results confirm that the genes downregulated in the knockout mice are functionally connected to myelin formation and CNS development.



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0019911	structural constituent of myelin sheath	5.734×10^{-8}
2	GO:BP	GO:0007272	ensheathment of neurons	8.078×10^{-9}
3	GO:BP	GO:0007417	central nervous system development	1.831×10^{-5}
4	GO:BP	GO:0061564	axon development	7.822×10^{-4}
5	GO:BP	GO:0070447	positive regulation of oligodendrocyte progenit...	1.985×10^{-2}
6	GO:CC	GO:0043209	myelin sheath	3.126×10^{-13}
7	GO:CC	GO:0044304	main axon	1.267×10^{-2}
8	GO:CC	GO:0071944	cell periphery	1.649×10^{-2}

Group B showed one significant result: TGF-beta receptor binding. This is relevant because LRG1 - the most strongly upregulated gene in our dataset is a known regulator of TGF-beta signaling, which plays a role in inflammation and tissue remodeling. This suggests that the inflammatory response in the knockout cortex may involve the TGF-beta pathway.



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005114	type II transforming growth factor beta receptor binding	8.484×10^{-3}

Discussion

The differential expression analysis of mouse cortex tissue revealed 49 significantly differentially expressed genes between control and knockout mice, pointing to a clear biological phenotype.

The most prominent finding is the coordinated downregulation of myelin-related genes in Group A. These genes encode structural proteins of the myelin sheath - the insulating layer that surrounds axons in the central nervous system. The fact that multiple myelin genes are simultaneously downregulated strongly suggests that the knocked-out gene plays a critical role in oligodendrocyte function and myelin production. This pattern is consistent with demyelinating diseases such as Multiple

Sclerosis and hypomyelinating leukodystrophies, where loss of myelin leads to impaired nerve conduction and neurological dysfunction.
The g:Profiler enrichment analysis confirmed this interpretation - the top enriched terms for Group A were myelin sheath, ensheathment of neurons, and oligodendrocyte differentiation, all with very high statistical significance.
In Group B, the most striking result was the extreme upregulation of LRG1, a protein involved in innate immune signaling and TGF-beta pathway activation. This suggests a neuroinflammatory response secondary to the demyelination - a pattern commonly observed in demyelinating diseases, where myelin damage triggers immune activation in the brain.
Together, the two groups tell the biological story: the knockout leads to loss of myelin gene expression, followed by inflammatory activation in the cortex.

Possible Treatment Direction

Potential treatment strategies could focus on:

1. Promoting remyelination - drugs that stimulate oligodendrocyte cells to differentiate and produce new myelin, similar to approaches being developed for Multiple Sclerosis.
2. Anti-inflammatory treatment - targeting the LRG1/TGF-beta signaling pathway to reduce the neuroinflammatory response observed in the knockout mice
3. Gene therapy - restoring expression of the knocked-out gene specifically in oligodendrocytes to rescue myelin production.

PART 2 - analysis of single cell sequencing in situ

Objective & Background

In this part of the project, we analyzed an in situ single-cell gene expression dataset from a breast cancer patient biopsy in order to predict whether the patient is likely to benefit from PD-L1 checkpoint inhibitor immunotherapy. Because checkpoint inhibitors rely on immune cells being present and functionally engaged within the tumor microenvironment, we evaluated three criteria:

(1) whether immune cells comprise at least 10% of the biopsy.

(2) whether immune cells are spatially mixed with tumor cells (MBC) rather than separated into distinct regions.

(3) whether at least 10% of cells express the PD-L1 gene (official gene symbol CD274) with a raw count of at least 1.

Checkpoint inhibitor immunotherapy can improve anti-tumor immune activity by reducing inhibitory signaling that otherwise prevents immune cells - especially T

cells, from attacking tumor cells. However, in breast cancer, response is variable across patients. Clinically approved checkpoint inhibitors in this context target PD-L1; therefore, the presence of immune cells within or near tumor regions and measurable PD-L1 (CD274) expression in the biopsy are key indicators that the therapy may be effective. For this project, the patient is considered a potential responder only if the answers to all three criteria above are “Yes”.

Input data

The analysis was performed using three raw CSV files:

1. expression_matrix.csv - a raw count matrix for 291 genes measured in 8627 single cells from one patient biopsy (each cell has a unique identifier).
2. locations_of_cells.csv - spatial coordinates (x,y in pixels) for each sequenced cell, using the same cell identifiers. Each coordinate represents the approximate center of the cell in the tissue section.
3. marker_genes.csv - a reference table of marker genes per cell type, used to help annotate cell identities in the biopsy.

Software & tools

The analysis was performed in Python using Scanpy and AnnData, which are widely used and well-established tools for standard single-cell workflows. Scanpy supports the full pipeline required in this project: QC metrics and filtering, library-size normalization, log-transformation, PCA, kNN graph construction, Leiden clustering, and UMAP visualization, while keeping all results (clusters, embeddings, annotations, spatial coordinates) organized in a single reproducible object (AnnData).

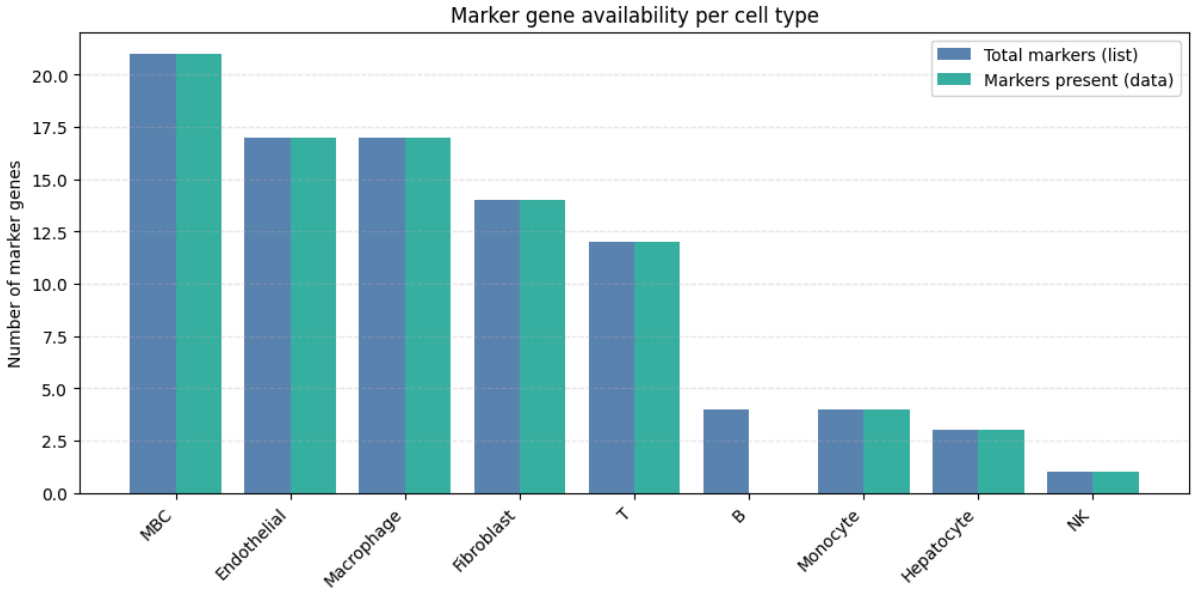
Seurat provides a very similar workflow in R using a Seurat object. In practice, Scanpy (Python) and Seurat (R) implement the same core analysis steps; we used Scanpy because it integrates naturally with pandas/numpy for cleaning and aligning the CSV inputs and with matplotlib for customized QC and spatial plots.

Workflow

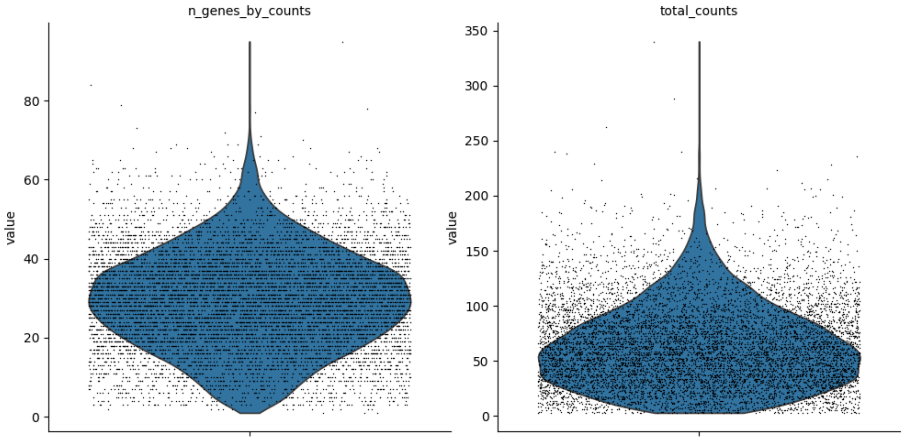
1. Load raw files and align cells - Loaded the three CSV files and performed basic cleanup: removed fully-empty columns in the expression table, set the first column as the cell-ID index, coerced values to numeric, and aligned expression and coordinates by intersecting cell IDs. After alignment, the expression matrix contained 8627 cells and 291 genes, and the locations table contained 8627 cells and 2 coordinate columns.

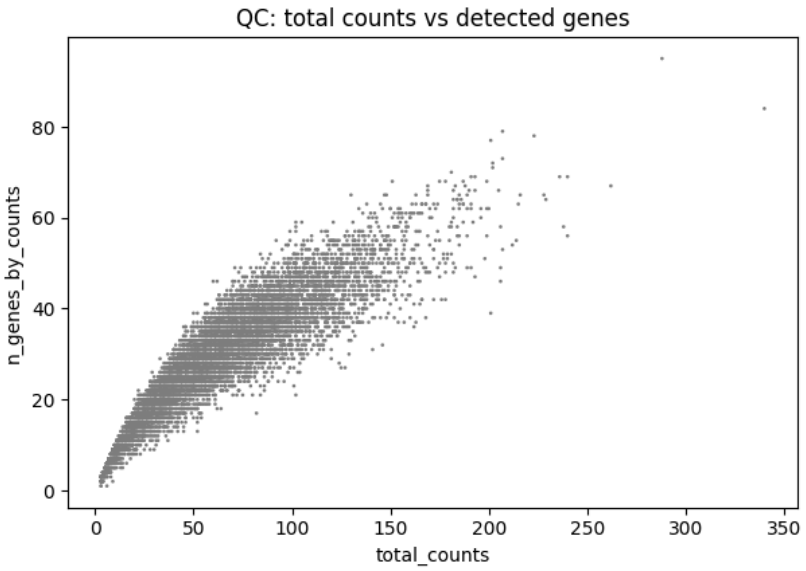
2. Marker gene availability check - Checked how many marker genes from the marker table are present in the dataset gene panel. This documents limitations of annotation when a cell type has few or no markers available in the measured genes. Although B cells were included in the provided marker list, we were not able to identify a B-cell population in this dataset because none of the B-cell marker genes appeared in the 291-gene expression panel. As a result, during the marker-based scoring step, B cells could not be scored/assigned and were effectively excluded from the cell-type classification..

CellType	total_markers_in_list	markers_present_in_data
MBC	21	21
Endothelial	17	17
Macrophage	17	17
Fibroblast	14	14
T	12	12
B	4	0
Monocyte	4	4
Hepatocyte	3	3
NK	1	1



3. Build AnnData and compute QC metrics -Marker-Gene Check: Before cell annotation, a verification step was performed to ensure the marker genes listed in marker_genes.csv were actually present in the sequenced gene panel. This step is crucial for the reliability of subsequent cell-type identification.
   Cell QC Metrics: Computed standard QC metrics using the scanpy library, including total counts per cell and the number of genes detected. Cells were visualized using violin and scatter plots to identify and potentially filter low-quality outliers.

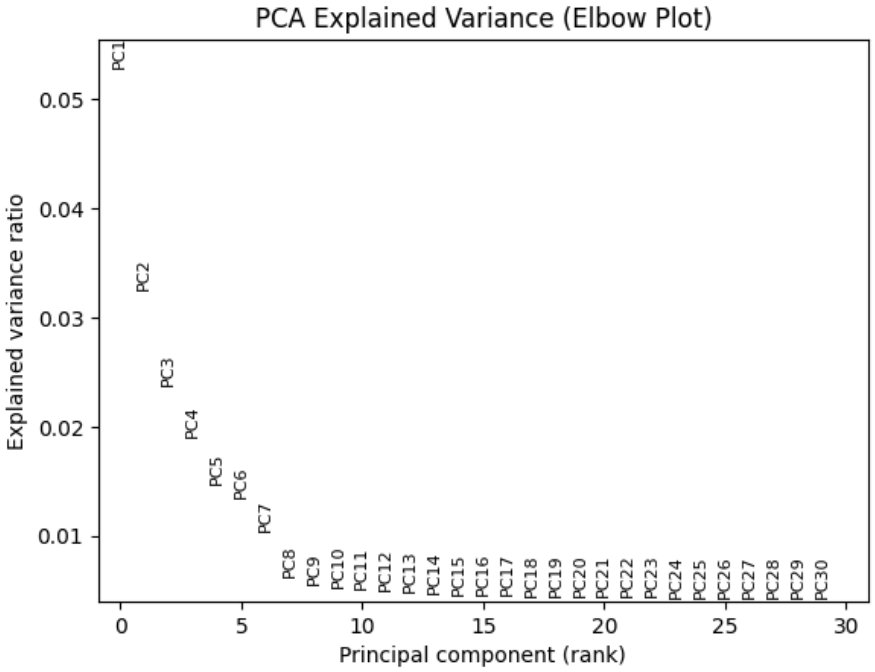




The plots summarize basic quality-control metrics across all cells. n_genes_by_counts shows how many genes were detected per cell, while total_counts reflects the total number of counts per cell (a proxy for library size).

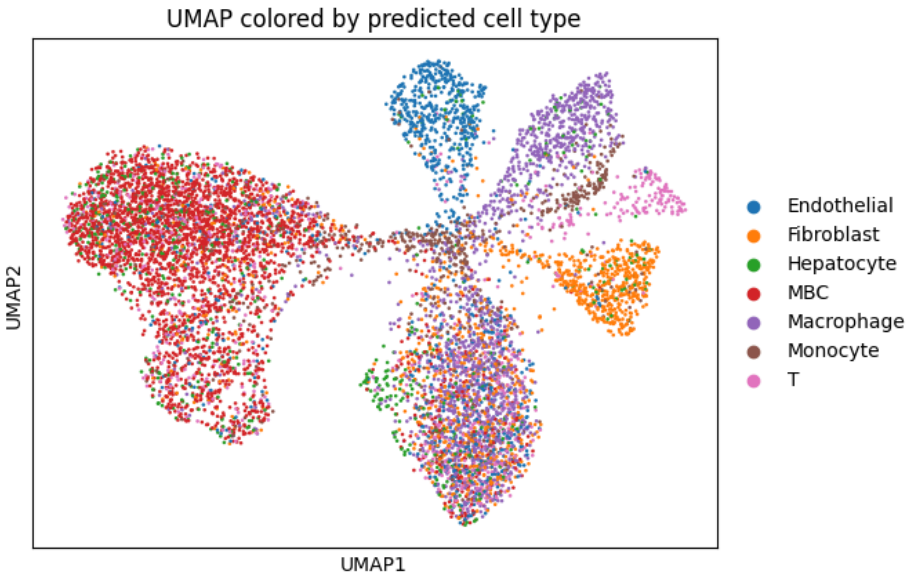
The scatter plot of total_counts vs. n_genes_by_counts shows the expected positive relationship: cells with higher total counts typically have more detected genes. Together, these plots help identify potential low-quality cells (very low counts/genes) and outliers with unusually high counts/genes that may represent technical artifacts.

4. Normalization, log-transformation, and PCA - saved a snapshot of the matrix for raw-count queries, which preserves the original count-based information needed for rules that explicitly rely on raw counts (e.g., considering expression as raw count ≥ 1). We then performed library-size normalization (`target_sum=1e4`) to make cells comparable despite different sequencing depths, followed by a `log1p` transformation to stabilize variance and reduce the impact of highly expressed genes.
   After scaling the features, we computed PCA to reduce the dimensionality of the dataset by projecting each cell onto a smaller set of components that capture the main sources of variation in gene expression. This denoised, low-dimensional representation is more suitable for building a nearest-neighbor graph, clustering, and UMAP visualization than the original high-dimensional gene space. Finally, an elbow/variance-ratio plot was used to select an appropriate number of PCs that retain most of the biological signal while minimizing noise for downstream analysis

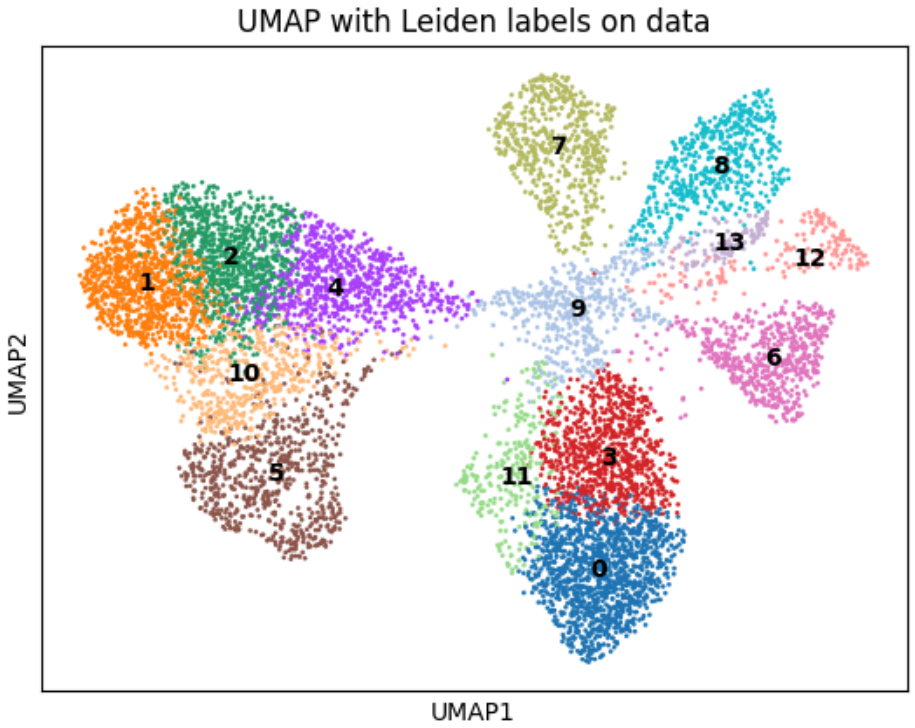


We selected 15 principal components based on the PCA elbow. After approximately the first ~15 PCs, the explained variance per additional component becomes small (the curve starts to plateau), suggesting that later PCs mainly capture noise or minor variation. Therefore, using 15 PCs provides a good balance for downstream neighbor graph construction, Leiden clustering, and UMAP visualization.

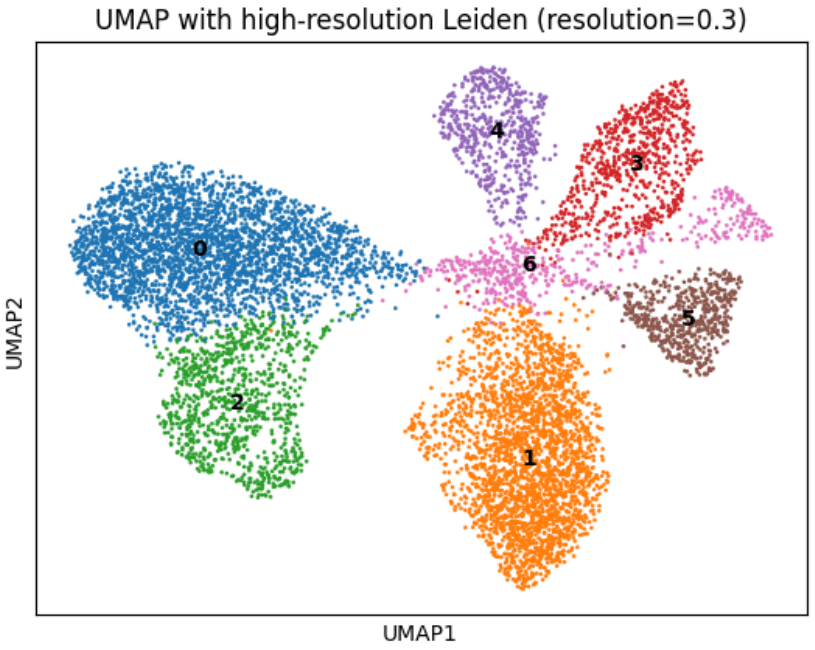
5. Neighbors graph, Leiden clustering, and UMAP - In this step, our goal was to group similar cells into clusters and assign them biological identities. We built a k-nearest-neighbor graph in PCA space (using the first 15 PCs), applied Leiden clustering - a graph-based community detection method that partitions the kNN graph into groups of highly connected cells - to define cell clusters, and computed UMAP, a dimensionality-reduction method for visualization that preserves local neighborhood structure, to visualize the resulting populations in 2D. To interpret the clusters, we used the marker-gene table to compute gene-set scores (mean expression of each cell type's markers) and assigned each cell the best-matching cell type label, then summarized the dominant label per cluster.



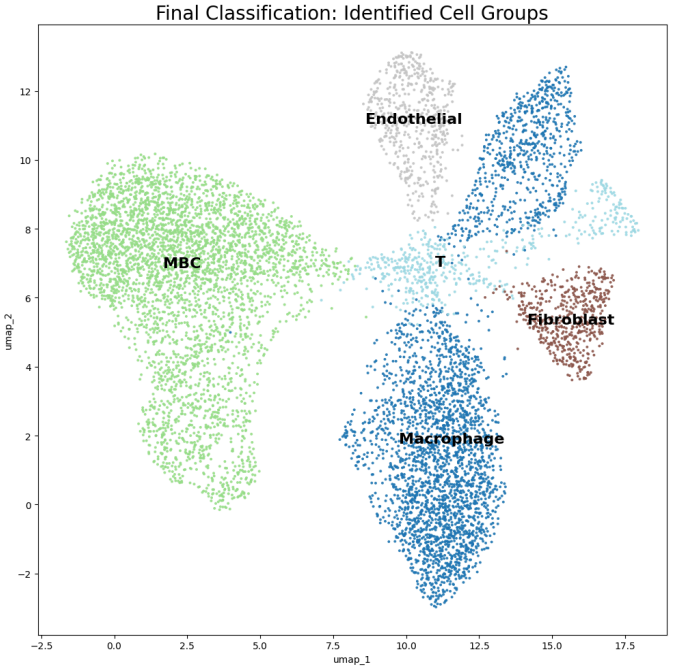
leiden_cluster	major_cell_type	n_cells
0	Macrophage	1277
1	MBC	962
2	MBC	934
3	Macrophage	862
4	MBC	799
5	MBC	675
6	Fibroblast	558
7	Endothelial	554
8	Macrophage	531
9	Monocyte	470
10	MBC	466
11	Hepatocyte	257
12	T	166
13	Monocyte	116



Then, we assigned biological identities to the high-resolution clusters and summarized how the clustering changed between Step 5 and Step 6. First, we used the provided marker-gene table to compute a marker gene-set score for each candidate cell type within each high-resolution cluster (mean expression of marker genes using `adata.raw`). Each `leiden_highres` cluster was labeled with the best-scoring cell type, and the resulting per-cell annotation was stored as `final_cell_group`.



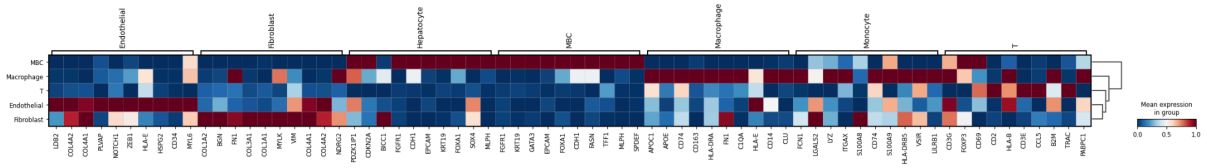
Next, we quantified how clusters from Step 5 relate to the refined Step 6 clusters by computing a cluster overlap matrix (cell-count crosstab). For each Step 5 cluster, we identified the Step 6 cluster containing the largest fraction of its cells and reported a purity score (largest overlap divided by the old cluster size), which indicates whether an old cluster remained mostly intact (high purity) or was split across multiple new clusters (lower purity). Finally, we plotted the UMAP colored by `final_cell_group` and added text labels at group centroids to present the final interpreted cell populations.



old	new	new_name	n_old	n_in_new	purity
0	1	Macrophage	1277	1277	1.00
1	0	MBC	962	962	1.00
2	0	MBC	934	932	1.00
3	1	Macrophage	862	862	1.00
4	0	MBC	799	791	0.99
5	2	MBC	675	667	0.99
6	5	Fibroblast	558	531	0.95
7	4	Endothelial	554	554	1.00
8	3	Macrophage	531	531	1.00
9	6	T	470	327	0.70
10	2	MBC	466	258	0.55

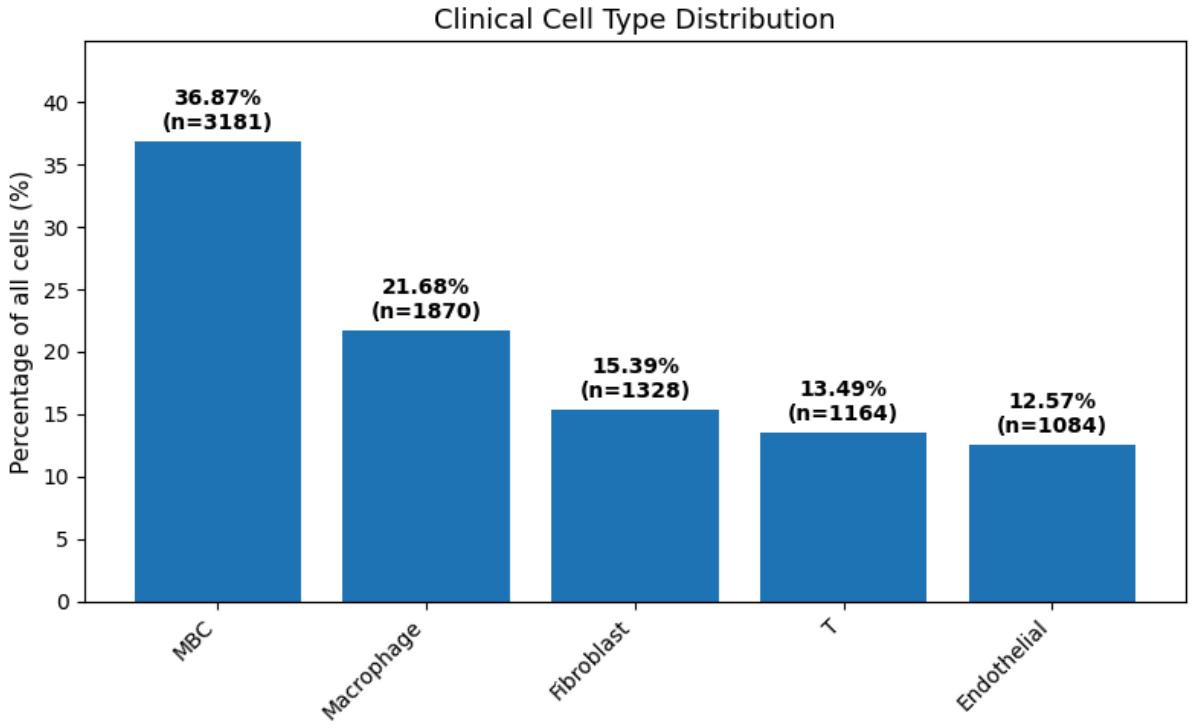
11	1	Macrophage	257	256	1.00
12	6	T	166	164	0.99
13	3	Macrophage	116	116	1.00

We generated a Matrixplot that serves as a final biological validation of our cell-type annotations by displaying the mean expression of top marker genes across the five defined clinical groups. The primary goal of this visualization was to confirm that each cluster identified during the analysis possesses a unique and consistent genetic signature, ensuring that our classification reflects the actual biology of the biopsy.

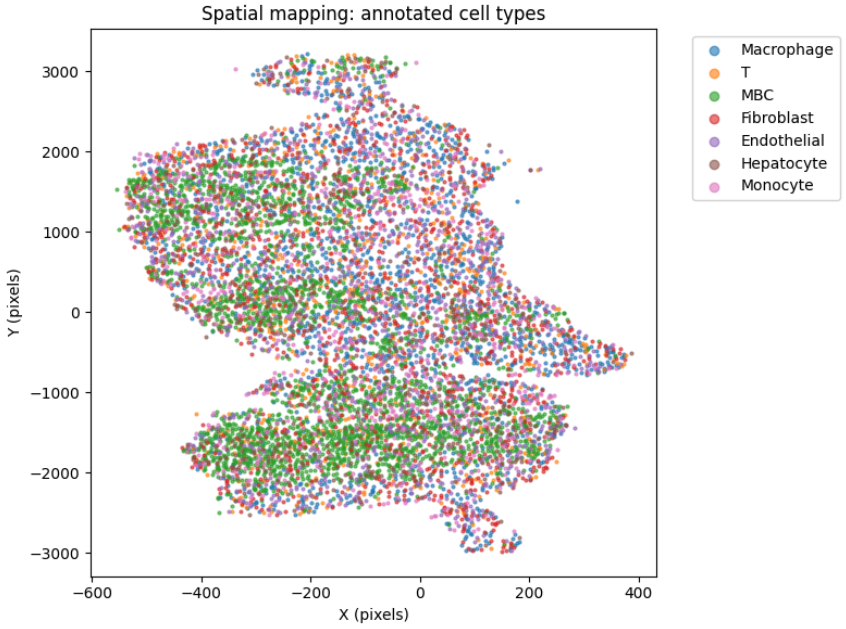


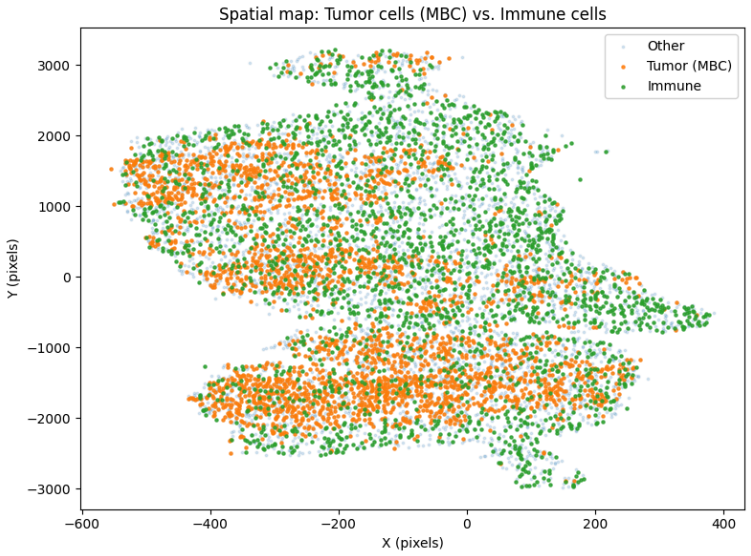
As observed in the plot, there is a clear diagonal pattern of high expression (dark red), where each group exclusively expresses its canonical markers - such as CD3G for T cells and KRT19 for MBC tumor cells - with minimal overlap between the clusters. This clean separation provides high confidence in our cell-type assignments and demonstrates that the tumor microenvironment has been accurately mapped, which is critical for evaluating the patient's potential response to immunotherapy.

6. Cell Type Distribution (Q1) – we performed a clinical 5-group re-classification to summarize the biopsy composition using a simplified set of major populations. We restricted the analysis to five groups and forced each cell to be assigned to exactly one of these groups by selecting the group with the highest score per cell. We then calculated the distribution of these five groups as both cell counts and percentages, as shown in the graph below. Finally, we computed the immune-cell percentage for Q1, which is defined as T cells + B cells + Macrophage. This calculation confirmed that immune cells constitute more than 10% of the biopsy sample.



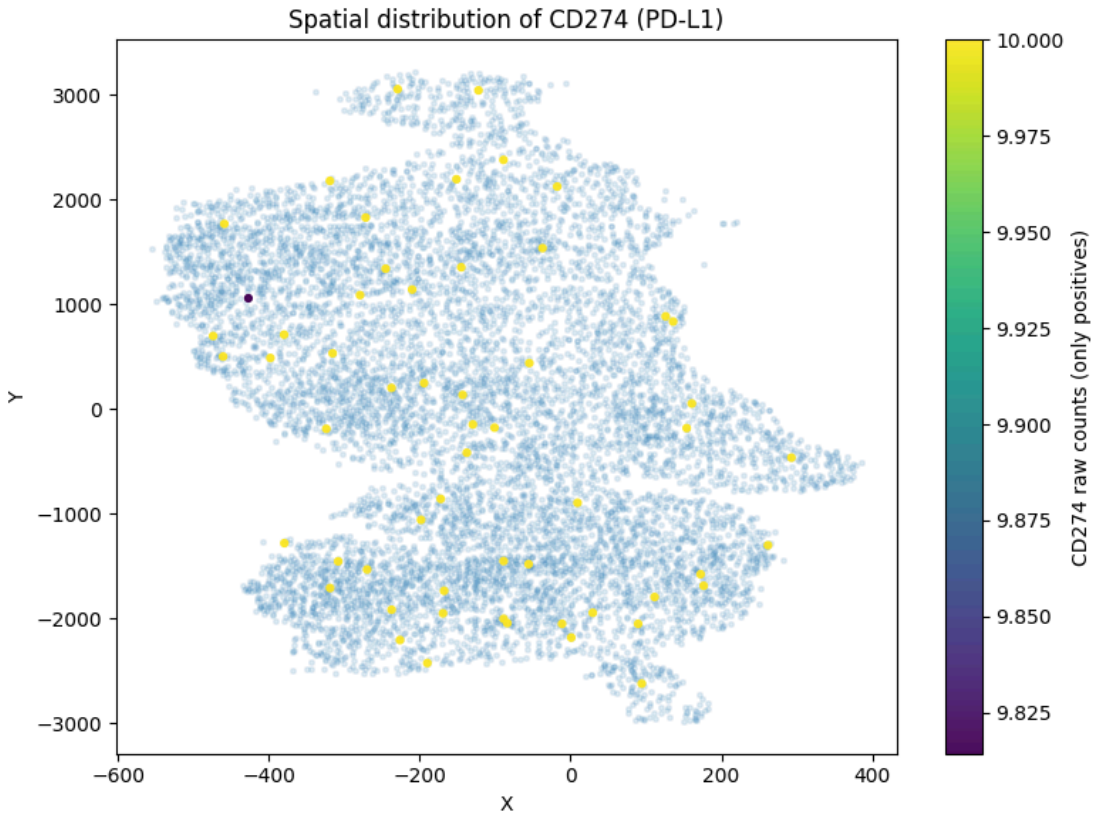
7. Spatial Distribution and Immune-Tumor Mixing (Q2) - to evaluate the spatial relationship between cells, we mapped the coordinates of each cell to its original position in the tissue. First, we visualized the distribution of all annotated cell types to ensure that the clustering was spatially coherent and to observe the general organization of the biopsy.





To address Q2, we generated a focused spatial map specifically highlighting the locations of tumor cells (MBC) and immune cells (T, B, and Macrophages). By inspecting this plot, we could qualitatively assess the mixing of the cells. Since immune cells were found to be distributed throughout the tumor areas rather than being isolated in a separate region, we concluded that the cells are mixed, satisfying the second requirement.

8. PD-L1 Expression Level (Q3) - we analyzed the expression of the PD-L1 gene to address Q3. Following the project instructions, which define expression as any detected count (raw count ≥1), we accessed the raw data layer (adata.raw) to ensure we were using unnormalized values. We calculated the percentage of cells expressing PD-L1 across the entire biopsy. Our analysis revealed that only 0.66% of the cells express CD274. This result is significantly below the 10% threshold required for a favorable immunotherapy response. To complement the quantitative analysis, we generated a spatial map highlighting only the PD-L1 positive cells.



In the background, all cells in the biopsy are shown in a light-gray to provide the structural context of the tissue. Positive cells (raw count ≥ 1) are plotted with a color scale indicating their specific raw count levels. The visualization confirms the statistical finding: PD-L1 expression is sparse and scattered across the biopsy. The lack of a concentrated or widespread PD-L1 presence indicates that the primary target for checkpoint inhibitor drugs is missing in this patient.

Conclusions

Our three answers reflect whether the key biological requirements for a PD-L1 checkpoint inhibitor are present in this biopsy.

First, **Q1 = Yes** (35.17% immune cells) means the tissue contains many immune cells. This matters because checkpoint inhibitors do not “kill” tumor cells directly, instead, they boost the activity of immune cells. If immune cells are absent, there is little for the drug to activate.
Second, **Q2 = Yes** (immune cells are mixed with tumor cells) indicates that immune cells are not confined to a separate region of the biopsy. They are spatially intermingled with the tumor population, which makes biological sense for a potential response: immune cells must be close enough to encounter tumor cells in order to attack them.

However, **Q3 = No** (PD-L1 positive cells = 0.66%, raw count ≥ 1) is the major limiting factor for this specific therapy. PD-L1 inhibitors work by blocking the PD-L1 “brake” that tumor cells use to suppress immune activity. If PD-L1 is not expressed, then there is very little “brake” to block, and the drug has limited opportunity to change the immune - tumor interaction. In our biopsy, only 0.66% of cells show any detectable CD274 expression by the assignment’s definition, which is far below the 10% threshold.

Therefore, even though the biopsy is immune-infiltrated and immune cells are spatially positioned to interact with tumor cells, the lack of sufficient PD-L1 expression suggests that this patient is unlikely to benefit from a PD-L1 inhibitor. Since the pathway this drug is designed to block is barely present in the sample, it is unlikely that treatment will have a significant therapeutic impact.