Gene ordering by spatial variability in human brain tissues using spatial transcriptomics

Candidate: Francesco Edoardo Vallone

Tutor: Prof. Andrea Sottosanti

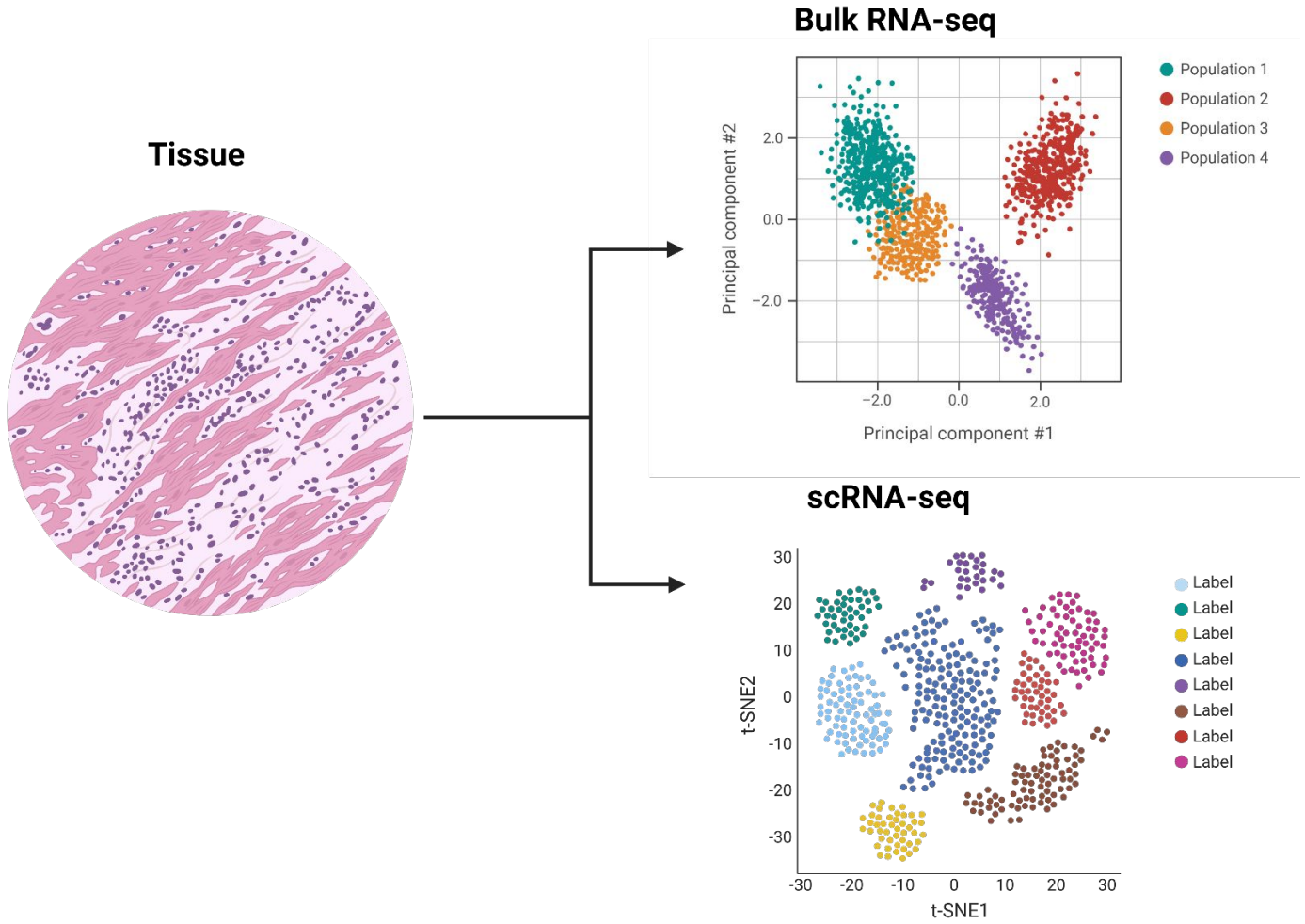
Università degli Studi di Padova

Dipartimento di Scienze Cardio-Toraco-Vascolari e Sanità Pubblica

UBEP

Unità di Biostatistica, Epidemiologia e Sanità Pubblica

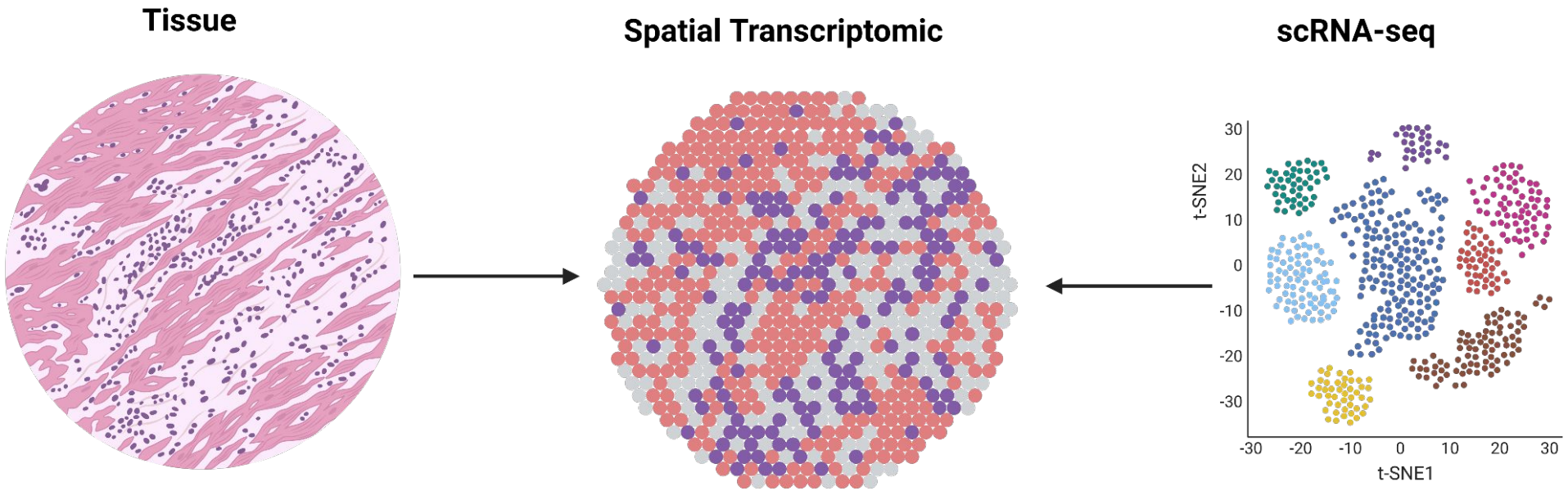
Spatial transcriptomic: an overview



scRNA-seq vs Bulk RNA-seq: Unveiling Tissue Pathophysiology

- **scRNA-seq** provides gene expression data at the single-cell level, unlike bulk RNA-seq which averages expression across cell populations.
- This allows the identification of **cellular heterogeneity** and distinct cell types.
- It has been crucial in understanding **specific cellular roles** in the pathophysiology of human tissues, revealing insights that bulk RNA-seq often misses.

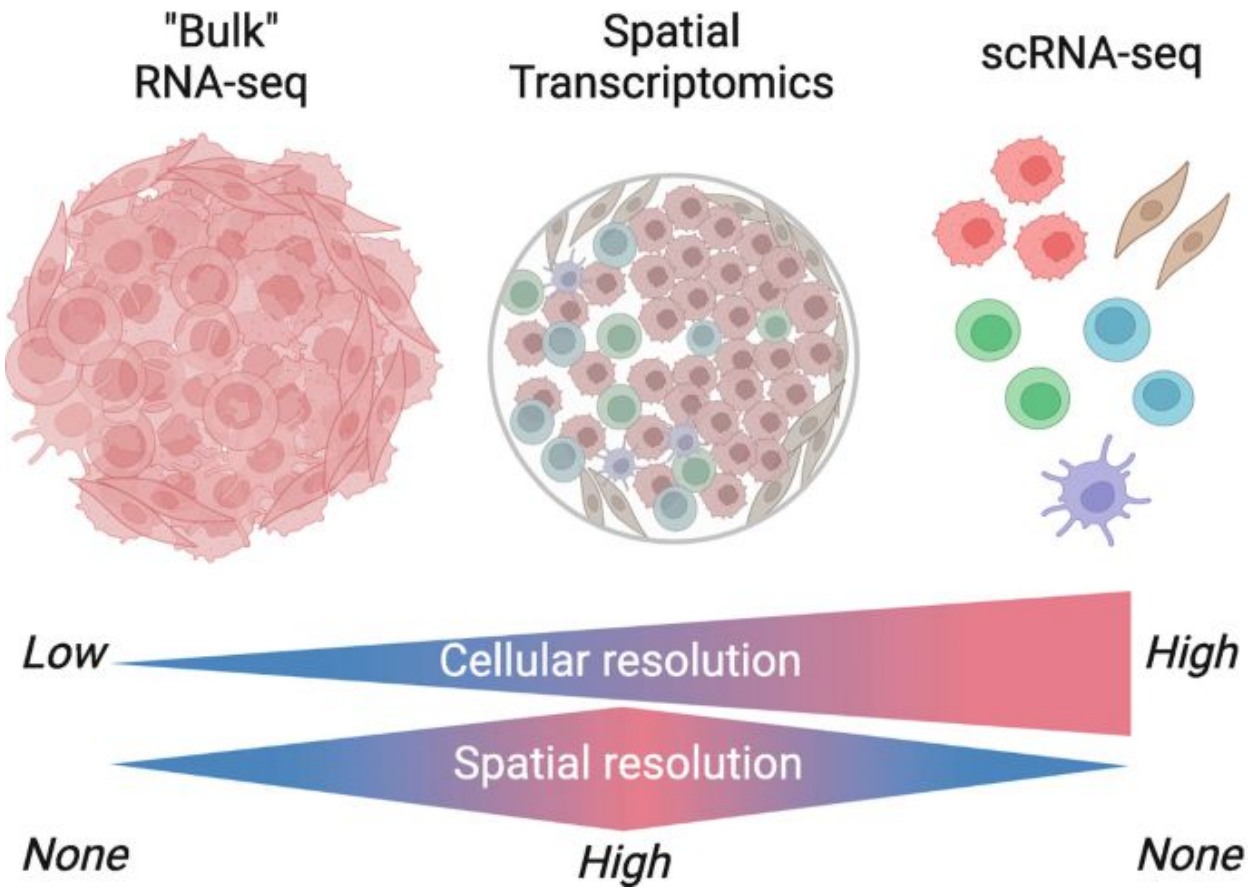
Spatial transcriptomic: an overview



Spatial Transcriptomics: beyond scRNA-seq

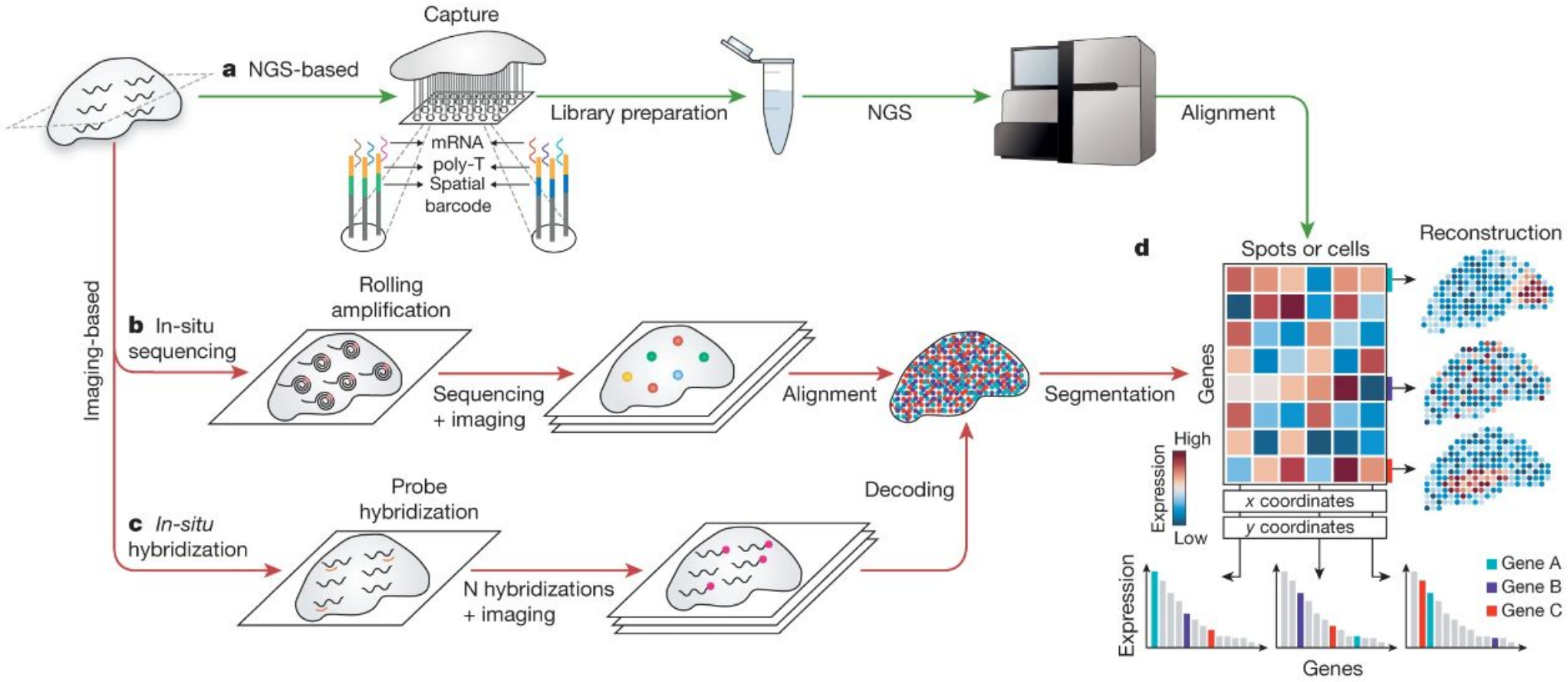
- Captures **gene expression** in the context of **tissue structure** and **cell location**.
- Reveals how **cell position** influences function and interactions, providing insights scRNA-seq cannot.

Spatial transcriptomic: an overview



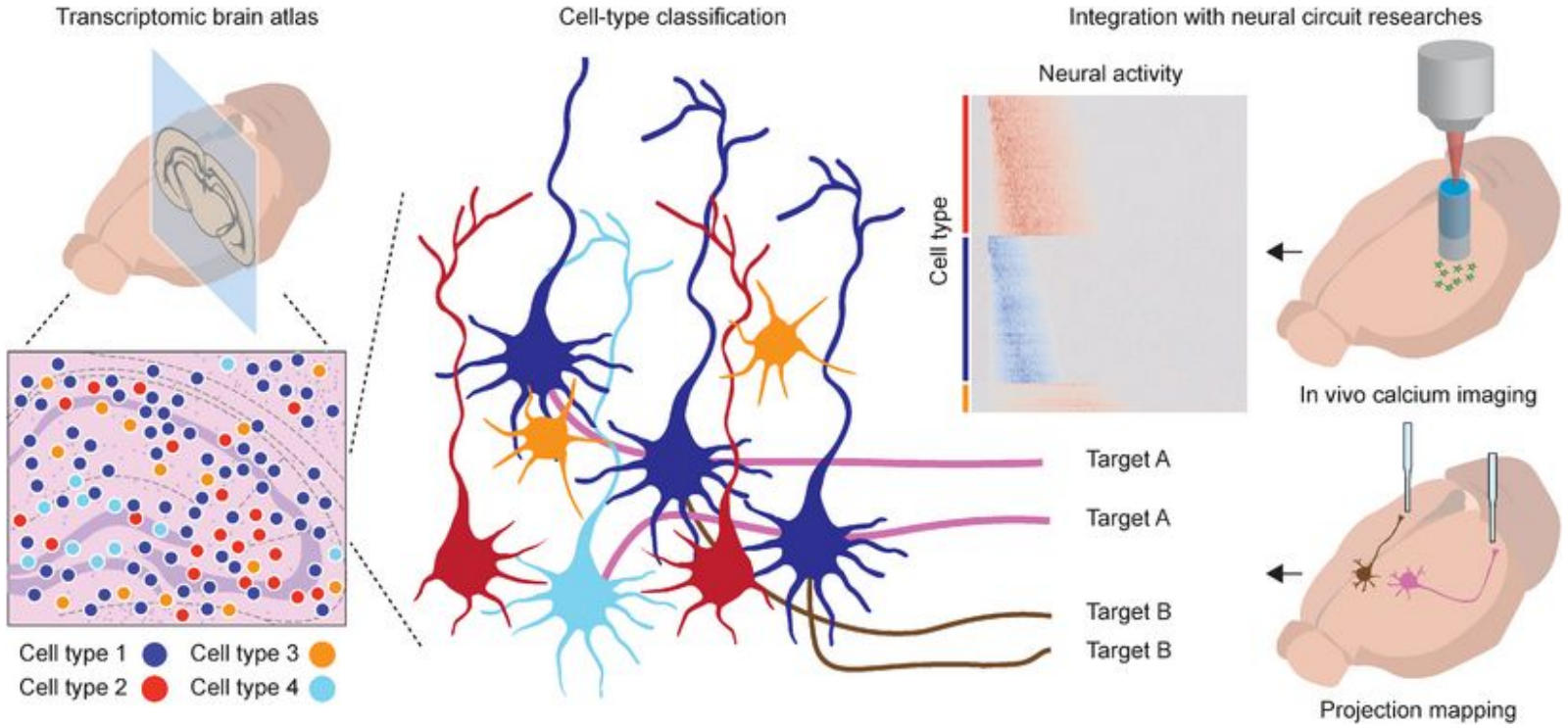
Spatial transcriptomic: an overview

Standard workflow for spatially resolved transcriptomics



https://doi.org/10.1038/s41586-021-03634-9

Spatial transcriptomic in neuroscience

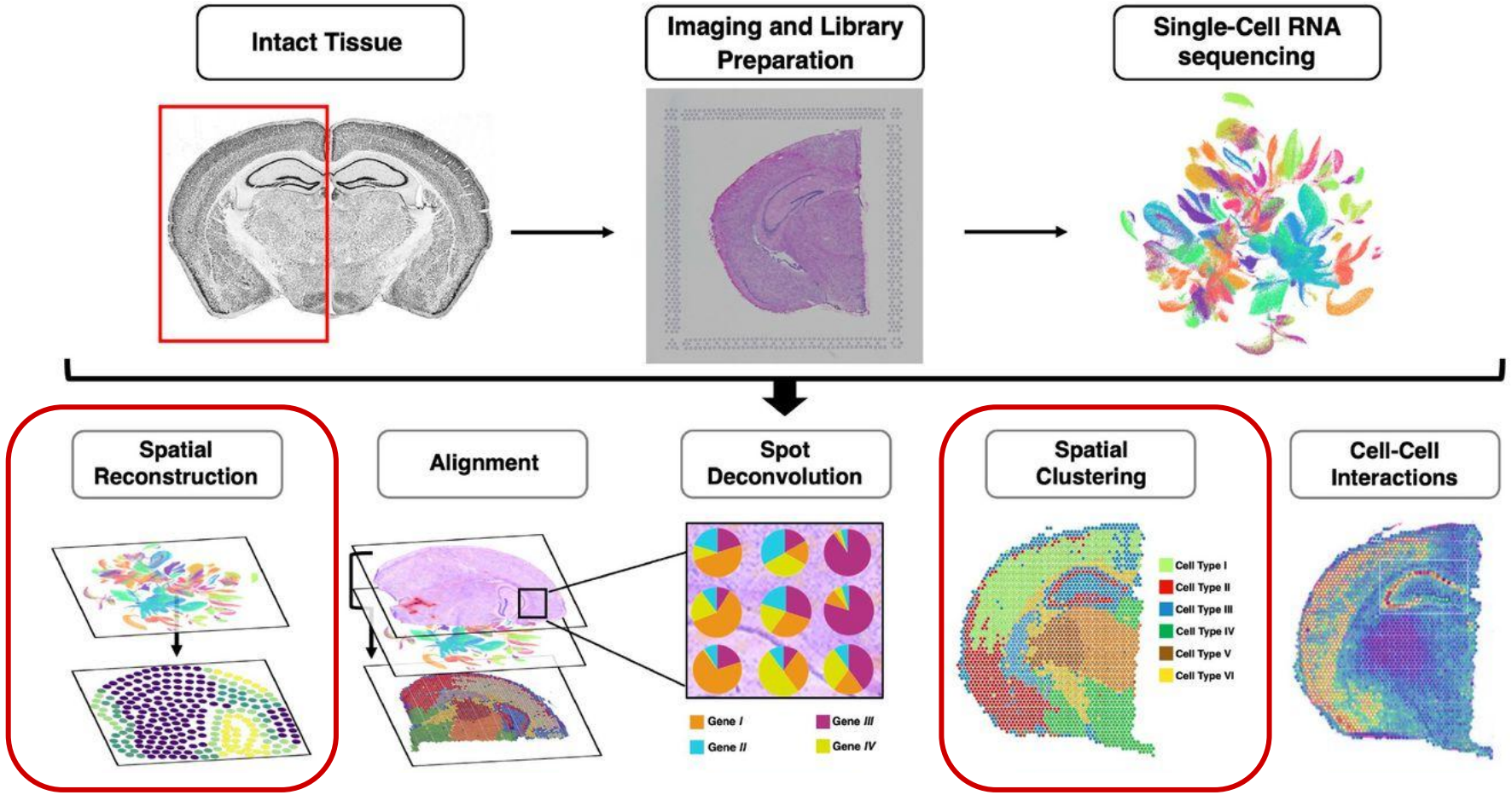


Importance of Spatial Transcriptomics in Neuroscience

- Enables study of gene expression while preserving spatial organization.
- Crucial for understanding complex brain circuits and cell interactions.
- Maps cell types/functions, revealing patterns linked to neurological diseases.

https://doi.org/10.1038/s41592-020-01040-z

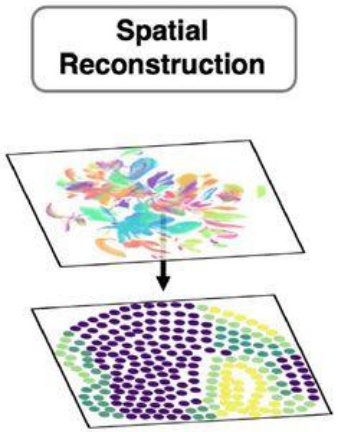
Aim of the project



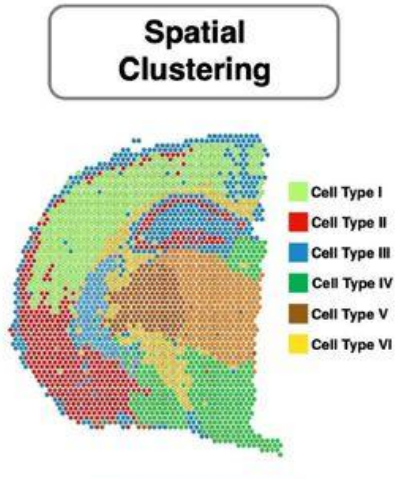
https://doi.org/10.1101/2022.02.28.482392

Aim of the project

1



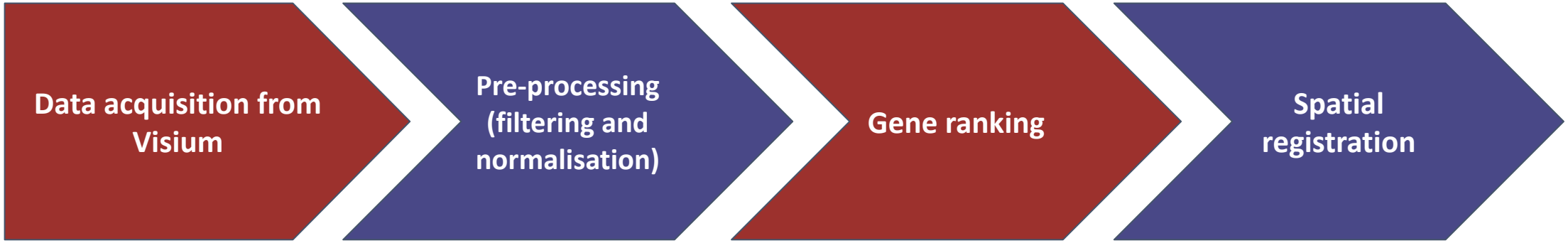
2



Compare **gene expression patterns** between two distinct layers of the human DLPFC

Università degli Studi di Padova
Dipartimento di Scienze Cardio-Toraco-Vascolari e Sanità Pubblica
UBEP
Unità di Biostatistica, Epidemiologia e Sanità Pubblica

Workflow of the project

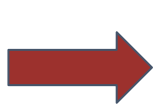


Methods

R packages used for the analyses

- **spatialLIBD**;
- SingleCellExperiment;
- SpatialExperiment;
- **nnSVG**;
- **PlackettLuce**;
- dplyr;
- ggplot2;
- rmarkdown

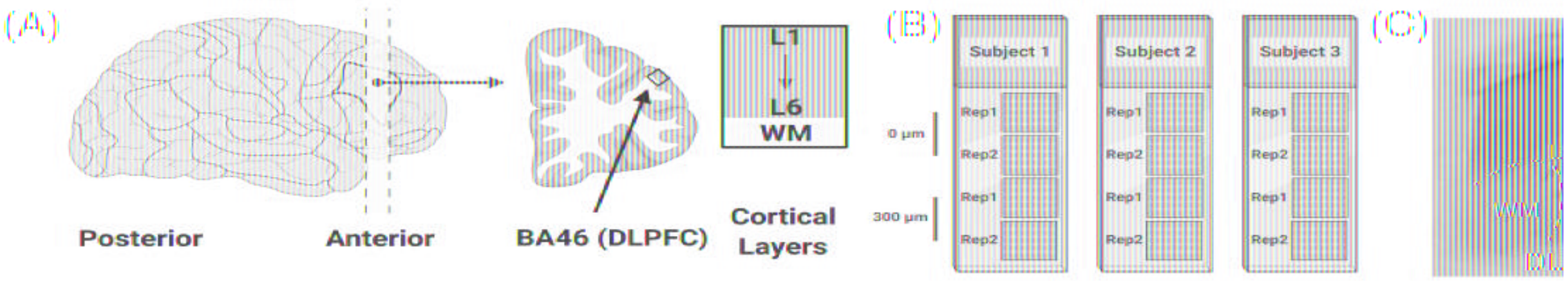
Data acquisition and spatial registration analysis



Gene ranking analysis

R version 4.4.1 (2024-06-14)
Platform: x86_64-redhat-linux-gnu
Running under: Fedora Linux 40 (Workstation Edition)

Data acquisition from Visium



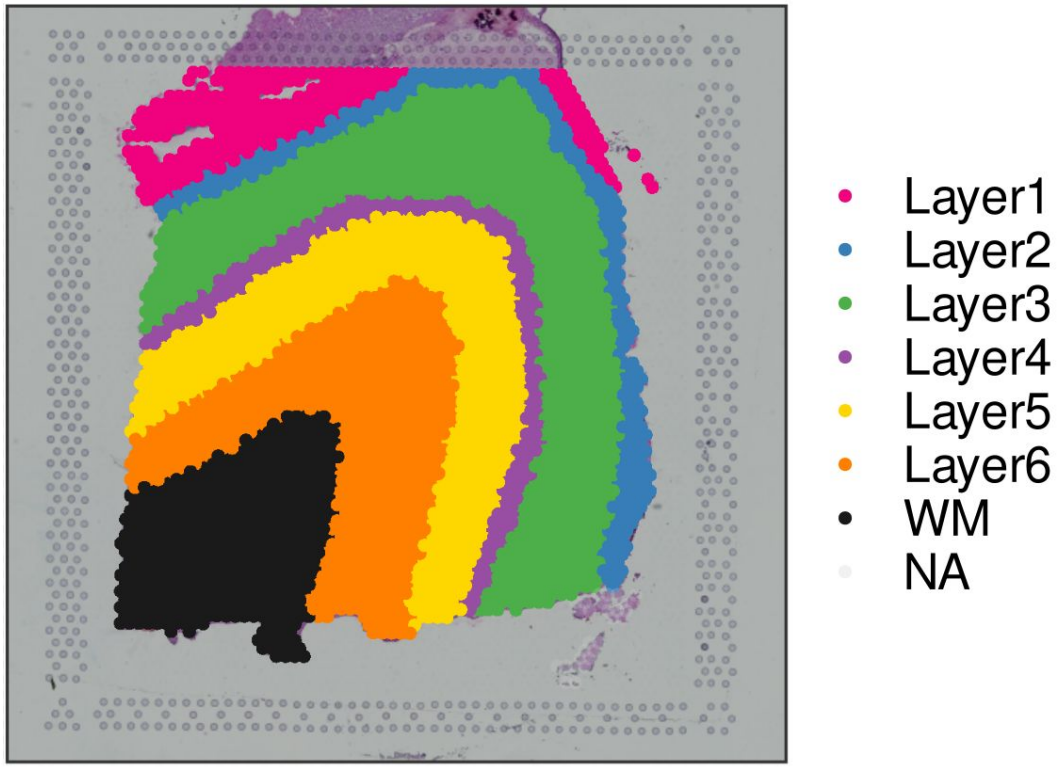
(A) This study focuses on a portion of the DLPFC comprising six neuronal layers and white matter.

(B) Data includes three subjects with two pairs of spatially adjacent replicates.

(C) Each DLPFC dissection (histology) spans all six layers plus white matter.

Data acquisition from Visium

151673 LIBD Layers



- Pre-processed DLPFC data were acquired from the **spatialLIBD** R package, loading the SingleCellExperiment object (**sce**).
- Data were subsequently converted into a **SpatialExperiment** object for seamless handling of both spatial coordinates and gene expression.
- Ensured proper integration of metadata (sample IDs, spatial coordinates, tissue images).
- Verified data consistency using **update functions** from **SingleCellExperiment** to maintain compatibility with the latest methods.

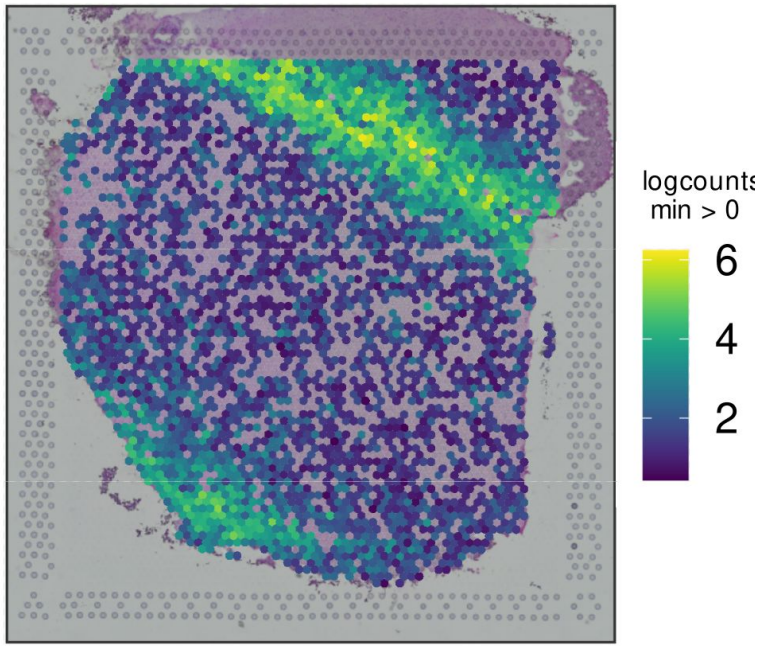
Pre-processing (filtering and normalisation)

- **Data Preparation**:
  - **Log-transformed counts** (calculated with **scran**) were already present in the spe object imported from the **spatialLIBD** package.
- **Spot Downsampling**:
  - A seed was set for reproducibility and 20,000 spots (from 47,681) were sampled from the dataset.
- **Gene Filtering**:
  - Genes with at least 3 counts and present in 70% of spots were retained (< 1000 final genes)
- **Duplicate Handling**:
  - Removed duplicate spots to ensure unique coordinates for analysis.

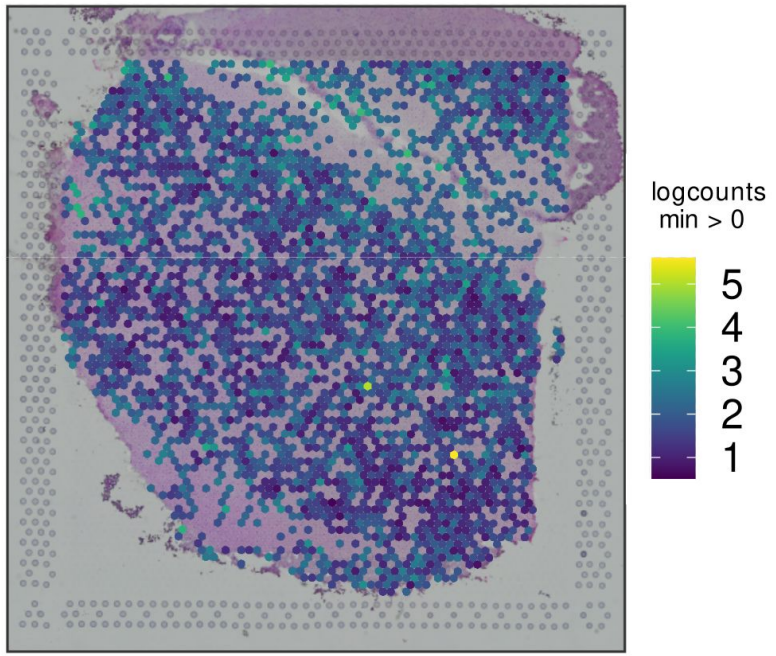
Gene ranking

Identifying and Comparing Spatially Variable Genes (SVGs) in White Matter and Layer 2 of the brain tissue

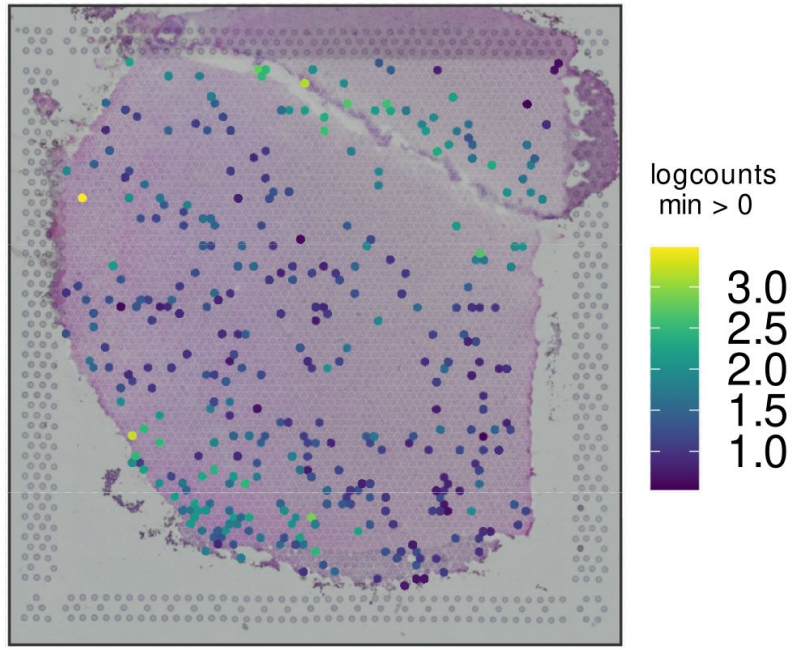
GFAP



NCDN



ELOVL1



Analytical steps:

- Identified **significant SVGs** in **White Matter** (adjusted p-value ≤ 0.05, top 15).
- Extracted same genes from **Layer 2** for comparison.
- Ranked genes in both layers using the **Plackett-Luce model**.

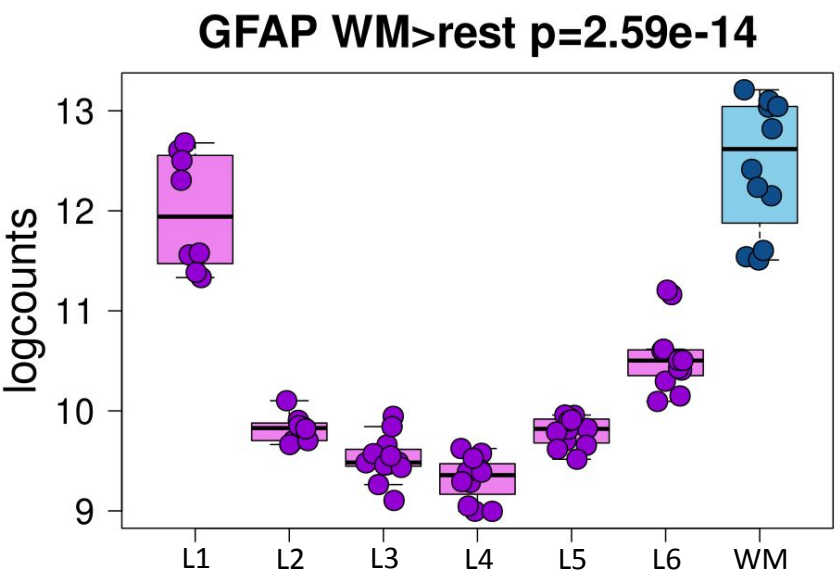
Università degli Studi di Padova

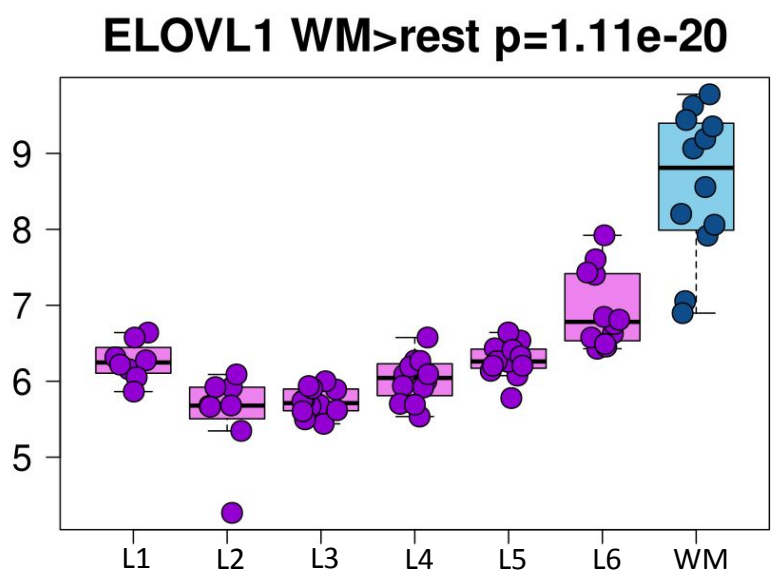
Dipartimento di Scienze Cardio-Toraco-Vascolari e Sanità Pubblica

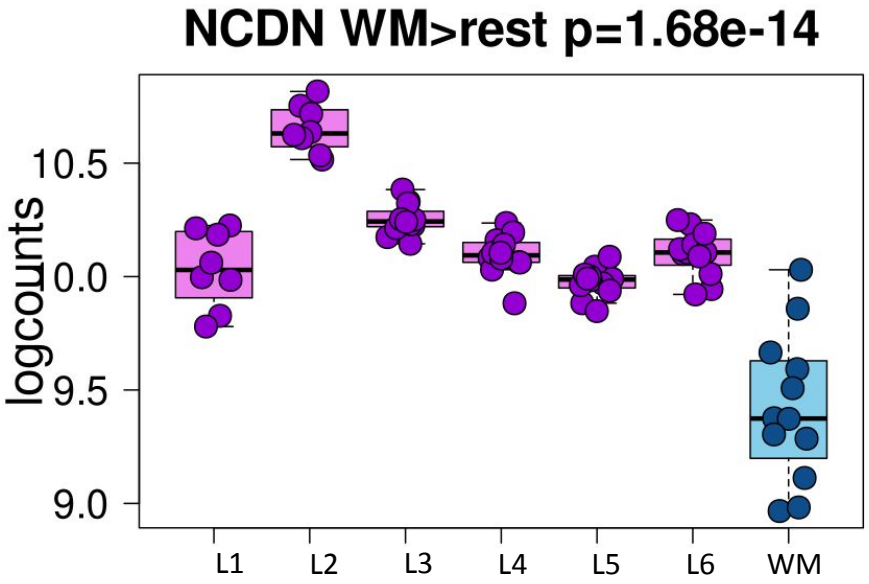
UBEP

Unità di Biostatistica, Epidemiologia e Sanità Pubblica

Gene ranking

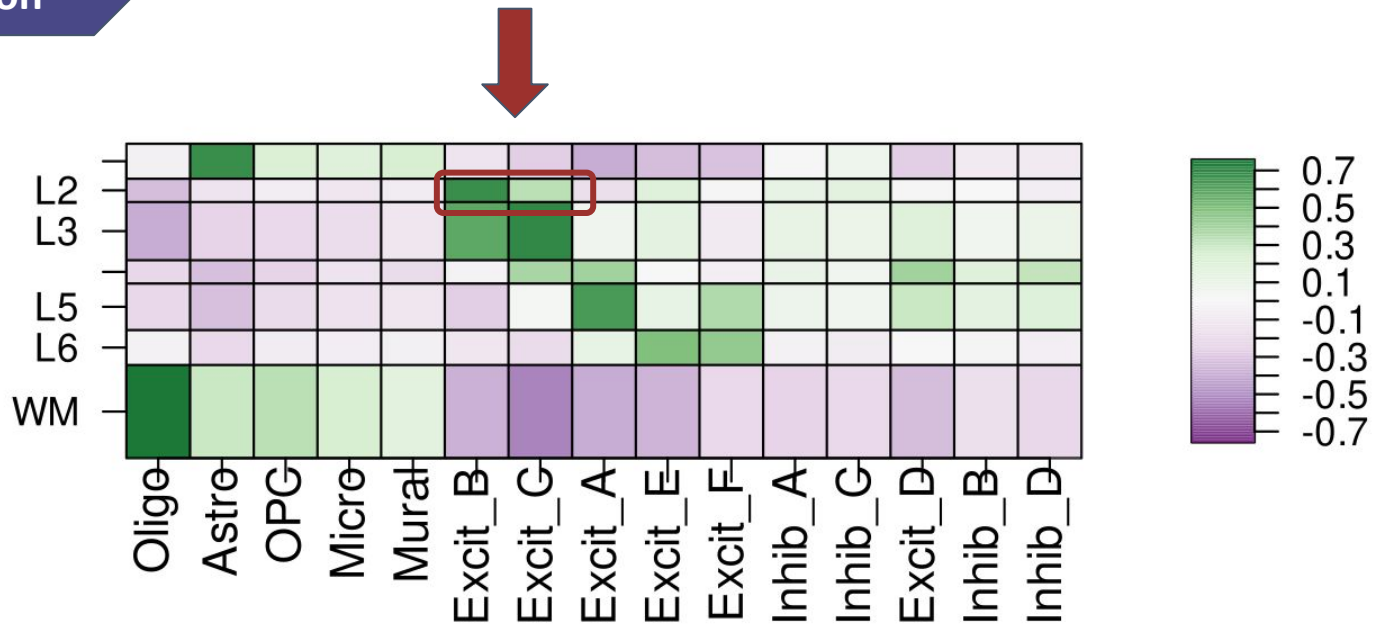






- Genes are ranked by their log coefficients, indicating how strongly they are expressed in the White Matter (WM) vs Layer 2, and enrichments were checked
- Among the genes, two exhibited opposite trends in their spatial expression:
  a. ***ELOVL1*****:** more prevalent in the **White Matter**.
  b. ***NCDN*****:** more prevalent in **Layer 2**.

Spatial registration



cluster	layer_confidence	layer_label
Oligo	good	WM
Astro	good	L1
OPC	good	WM
Micro	good	WM
Mural	poor	L1*
Excit_B	good	L2/3
Excit_C	good	L3
Excit_A	good	L5
Excit_E	good	L6
Excit_F	good	L6/5
Inhib_A	poor	L2/3/4*
Inhib_C	poor	L2*
Excit_D	good	L4
Inhib_B	poor	L4*
Inhib_D	good	L4

- Used **spatialLIBD** for aligning spatial transcriptomics data to DLPFC layers.
- Performed **pseudobulk analysis** to group cells by spatial location and aggregate gene expression.
- Mapped spatial data to pre-labeled cell types using known marker genes.

CONCLUSIONS AND DISCUSSION

- ***ELOVL1*** was highly expressed (ranked) in the white matter. Importantly, *ELOVL1* is known to be involved in myelination, which is essential for creating and maintaining the myelin sheath. (doi: 10.1002/emmm.201000061)
- ***NCDN*** was more abundant in Layer 2 of the brain, which was primarily occupied by excitatory neurons. NCDN was shown to play regulatory roles in cognitive functions and neuronal plasticity, highlighting its relevance in this layer. (doi: 10.1016/j.ajhg.2021.02.015)
- This approach enabled to **rank genes** across different brain layers, revealing insights into spatial gene expression patterns that traditional methods often miss.
- This methodology has **important implications** for both medical and basic research, helping to identify layer-specific gene dysregulation in neurological disorders and virtually other diseases.
- One **limitation** of this technique is its reduced ability to detect rare cell types or subtle differences in gene expression among similar cell populations.

Thank you for your kind attention