Alignment of Spatial Transcriptomics data with the `alignProMises` R package

Daniela Corbetta, Angela Andreella, Livio Finos, Davide Risso

daniela.corbetta@phd.unipd.it

Introduction

- **Spatial transcriptomics** data provide both **genomic and spatial information**
- **The structure of the brain differs between subjects** → Samples from different subjects cannot be compared and jointly analyzed since they are not aligned → challenge of multi-sample analysis

Aim of the analysis: rotate different samples from different subjects to absorb the unwanted variability caused by the misalignment

Alignment methods based on Procrustes theory: a statistical shape analysis that aligns matrices using **similarity** transformations

- **2 matrices → explicit solution**: $\hat{X}_1 = X_1\hat{R}$, where $\hat{R} = UV^\top$ and U and V derive from the **SVD** of $(X_1^\top X_2)$
- **More than two matrices → iterative algorithms**: the **ProMises model** and the **Efficient ProMises model** (Andreella and Finos, 2022) provide a unique solution.

ProMises model

Let $X_i \in \mathbb{R}^{n\times m}$, $i = 1, \dots, N$, be the matrices to be aligned. Every X_i is the **rotation of a common reference matrix plus an error term**:

$$X_i = (M + E_i)R_i^\top \quad \text{subject to} \quad R_iR_i^\top = R_i^\top R_i = I_m$$

- $E_i \sim \mathcal{MN}_{n,m}(0, \sigma^2 I_n, I_m)$.
- M is the **common mean** matrix with dimension $n \times m$.
- R_i is the **orthogonal rotation parameter**. It has **von Mises-Fisher prior distribution** with location parameter F and concentration parameter k → conjugate prior for the matrix Normal distribution.

The MAP estimate for R_i is $\hat{R}_i = U_iV_i^\top$, where U_i and V_i derive from the SVD of $X_i^\top M + kF$.

Limitations:

- high **computational load**
- matrices must have the **same dimension**

Efficient ProMises model

Idea: reduce the computational load of the ProMises model by a preliminary **dimensionality reduction**.

Thin-SVD of $X_i = L_iD_iQ_i^\top$, $Q_i \in \mathbb{R}^{m\times n}$.
Semi-orthogonal transformation: $X_i^* = X_iQ_i \in \mathbb{R}^{n\times n}$.
Model:

$$X_iQ_i = (M^* + E_i^*)R_i^{*\top} \text{ subject to } R_i^{*\top}R_i^* = I_n$$

- The reduced matrices are aligned by the ProMises model → MAP estimate of $\hat{R}_i^* = U_iV_i^\top$, where U_i and V_i derive from the SVD of $X_i^{*\top}M^* + kF$
- Project the aligned matrices $\hat{X}_i^* = X_i^*\hat{R}_i^*$ back to the original space with the inverse transformation $Q_i^\top$
- If the matrices have the same dimensions → alternative version: thin-SVD of $\hat{M} = \sum_{i=1}^N X_i/N = LDQ^\top \rightarrow X_i^* = X_iQ$

Suitable for matrices with **different number of columns**.

Package overview

4 main functions:

- `GPASub`: performs functional alignment of a matrix by the ProMises model with known reference matrix M;
- `ProMisesModel`: performs the functional alignment using the ProMises model with unknown reference matrix M;
- `EfficientProMisesSubj`: performs the functional alignment using the Efficient ProMises model;
- `EfficientProMises`: performs the functional alignment using the alternative version of the Efficient ProMises model.

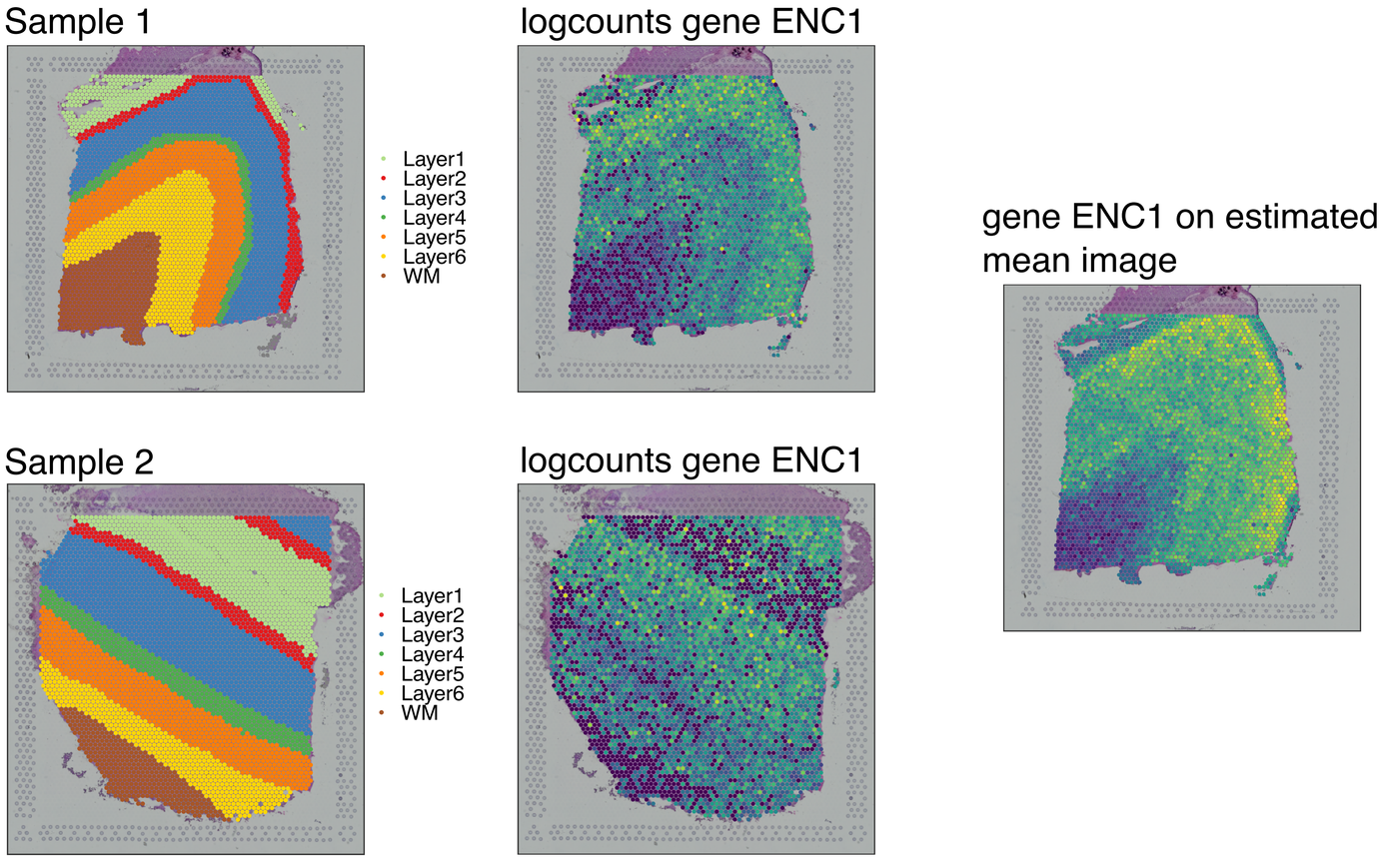
Case study

Data: 2 subjects, 4 samples per subject with different number of spots (Maynard et al., 2021).

- Different samples have **different orientation**;
- Match the different samples with the **Efficient ProMises model** → `EfficientProMisesSubj` function:

```
>out = EfficientProMisesSubj(data, t = 1,
maxIt = 100, Q = Q, k = k, scaling = F,
centering = F, singleQ = T, l = l)
```

- At the end of the alignment step, obtain the **estimated mean matrix** $\hat{M}^* = \sum_{i=1}^N X_i^*\hat{R}_i^*/N$;
- Project $\hat{M}^*$ in the space of an image: $\hat{M} = \hat{M}^*\hat{R}_1^{*\top}Q_1^\top$ → obtain a **de-noised image** → see genes expression levels on the estimated mean matrix;
- Gene ENC1 highly expressed in Layer 2 and Layer 3.



References

Andreella, A and Finos, L. (2022). Procrustes Analysis for High-Dimensional Data. *Psychometrika*, **87**, 1422 – 1438.
Maynard, K.R. et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, **24**, 425 – 436.