

The fEVE framework : a modular and descriptive framework for -omics ensemble clustering



ASLOUDJ Yanis, MOUGIN Fleur, THEBAULT Patricia
yasloudj@u-bordeaux.fr



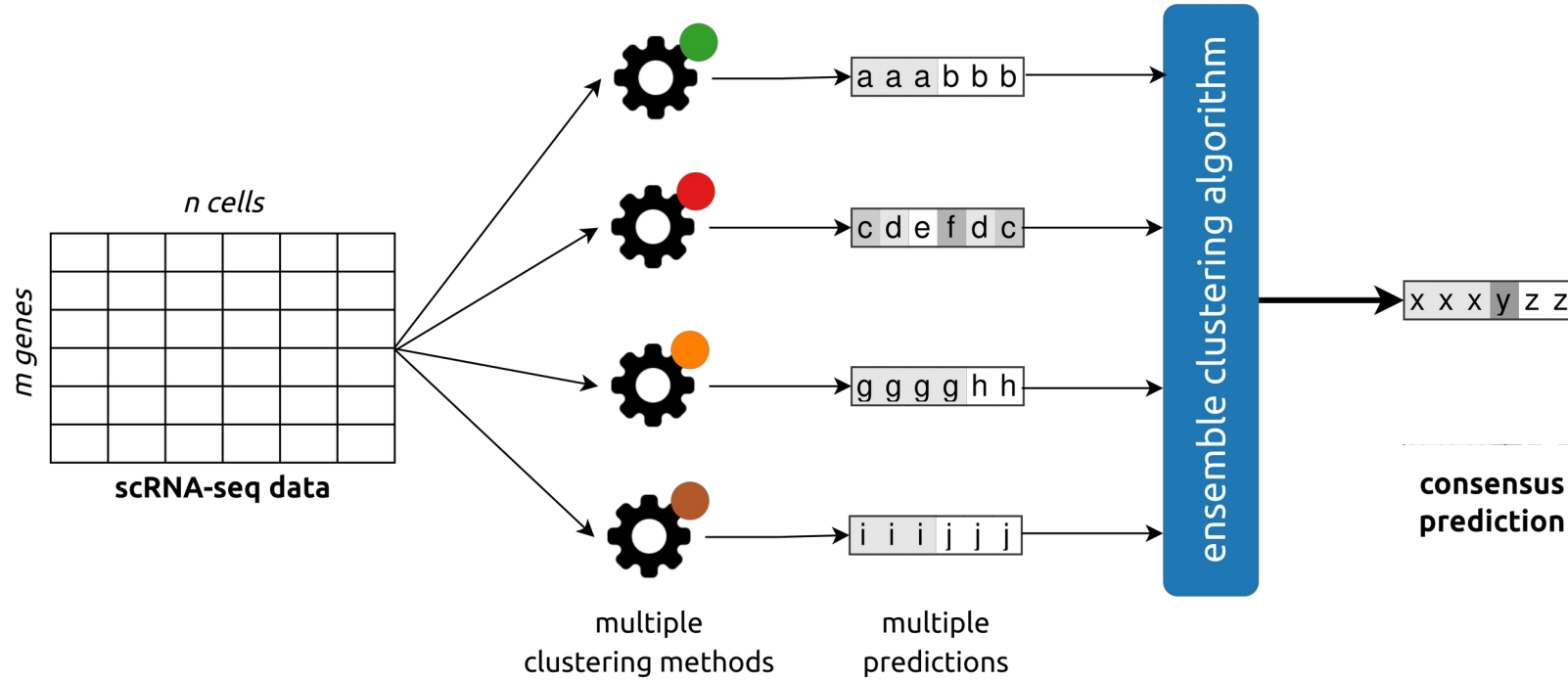
GDR Groupement
de recherche
BIMMM Bio-Informatique Moléculaire :
Modélisation et Méthodologie

■ Overview of scRNA-seq analyses (and my PhD)



■ Ensemble clustering algorithms predict robust clusters

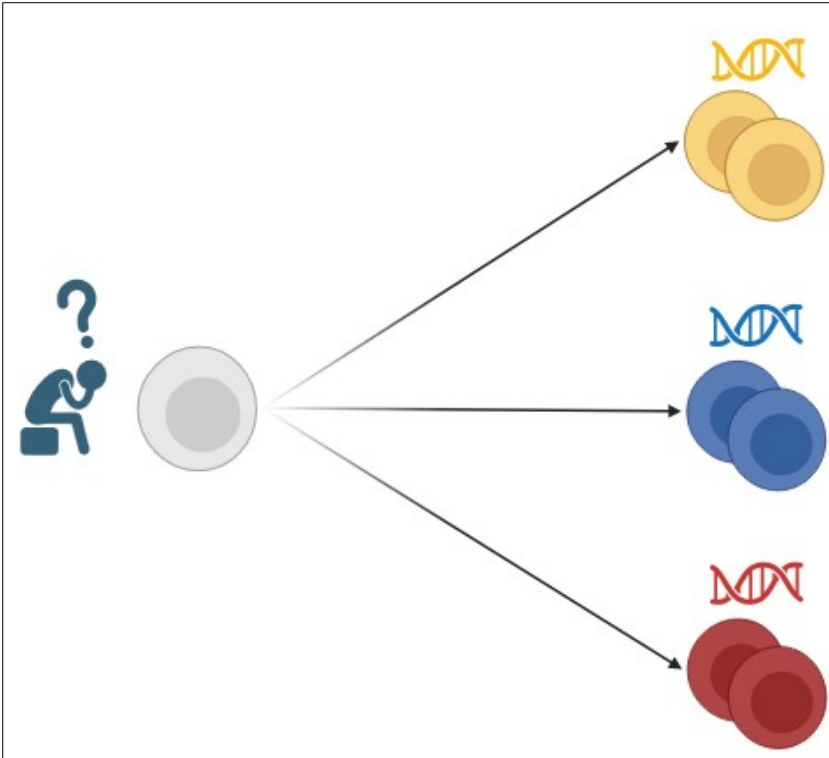
- In scRNA-seq, **hundreds of clustering methods** have been developed.
- **Ensemble algorithms** leverage multiple methods to predict robust clusters.



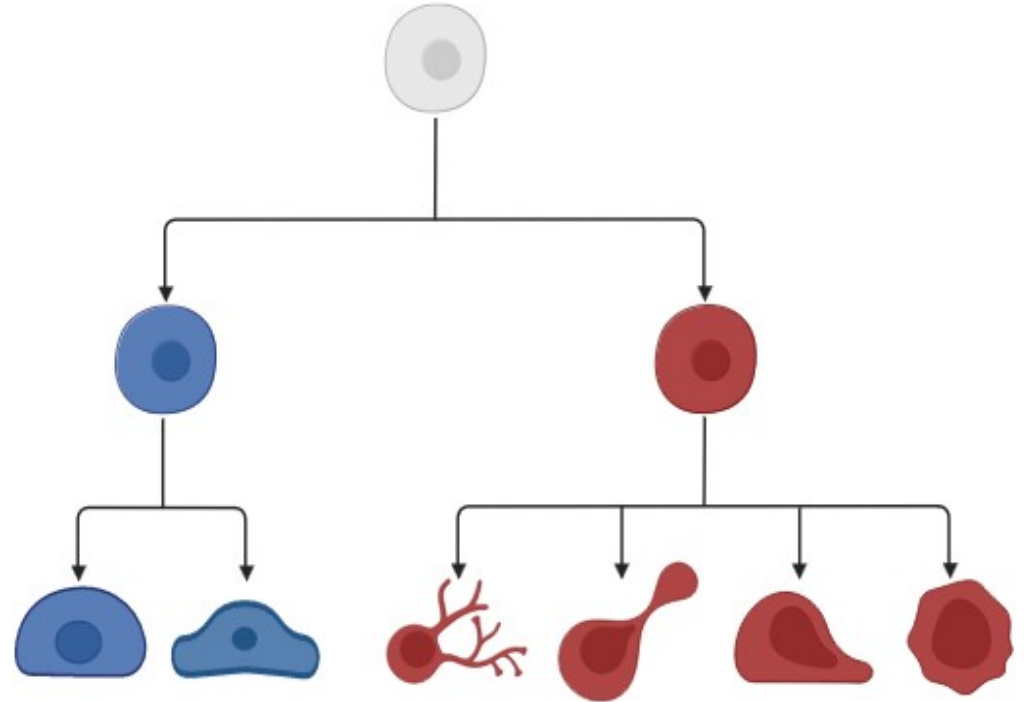
- Specifically, they minimize Δ , the differences between predictions.

■ These algorithms had yet to address some challenges in clustering

Quantifying uncertainty of measurements and analysis results



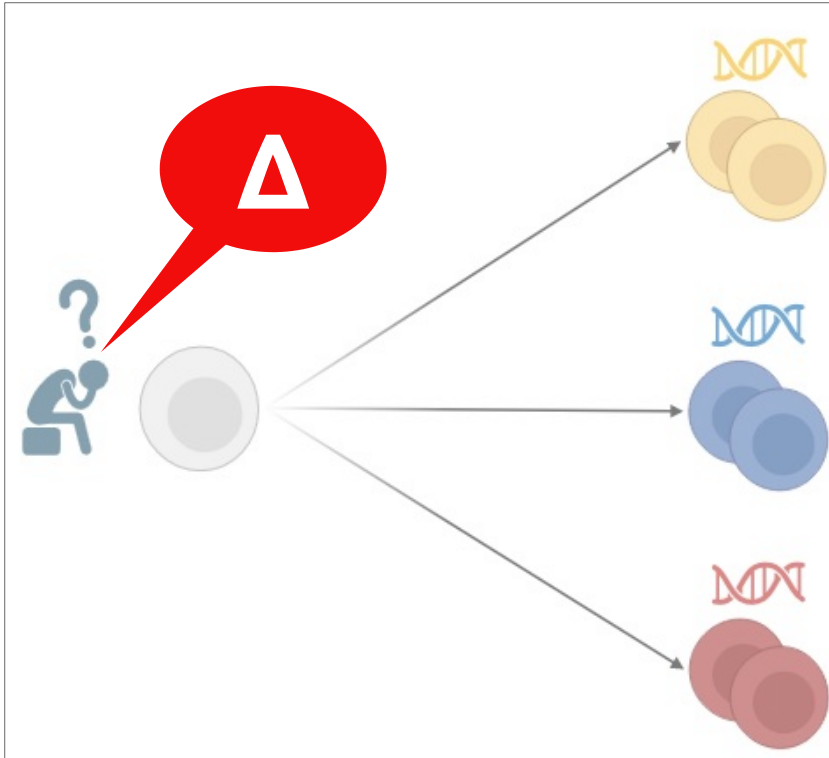
Varying levels of resolution



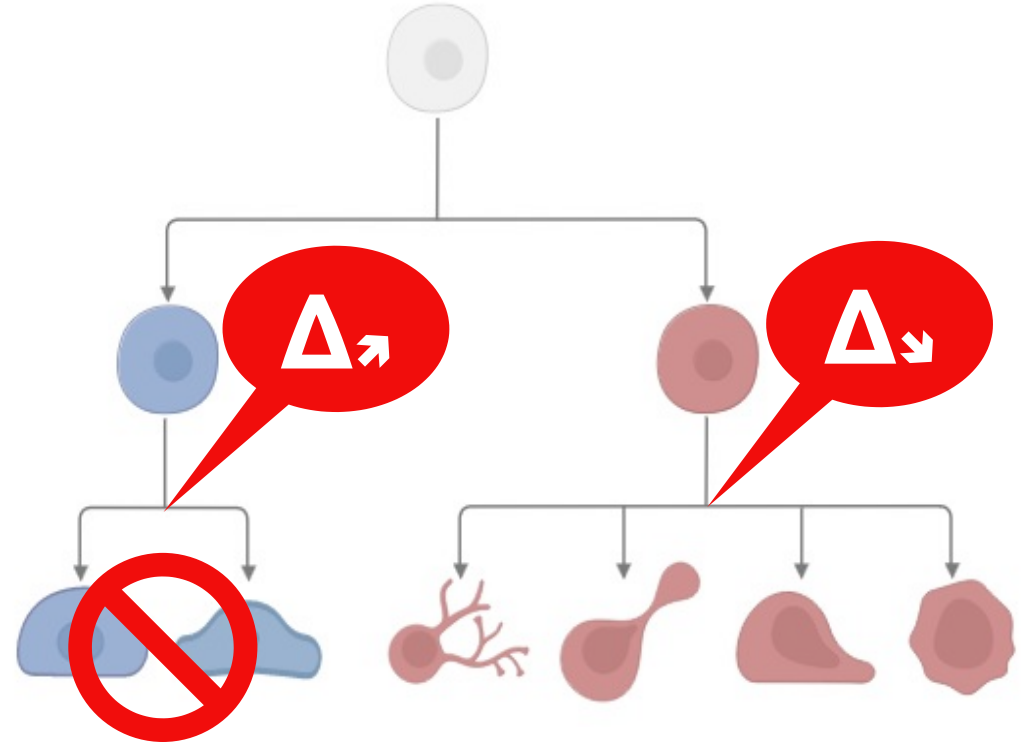
* Challenges from Lähnemann et al. 2020.

■ Can the differences Δ be leveraged to adress these challenges ?

Quantifying uncertainty of measurements and analysis results



Varying levels of resolution



* Challenges from Lähnemann et al. 2020.

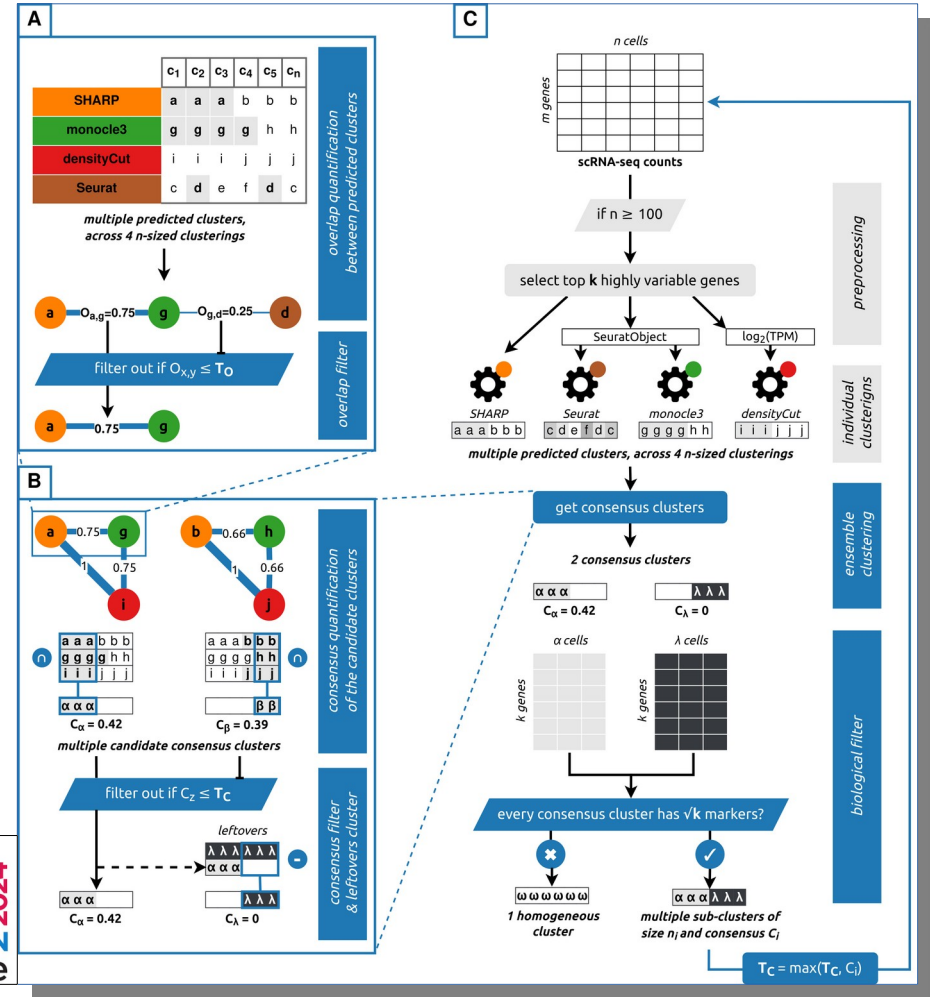
■ We had developed scEVE to investigate this question

Can we leverage the differences in prediction Δ instead of minimizing them?

Are Δ informative?

scEVE

JOBIM2025: call for submissions close at the end of the month



■ Overview of the scEVE algorithm

Feature selection

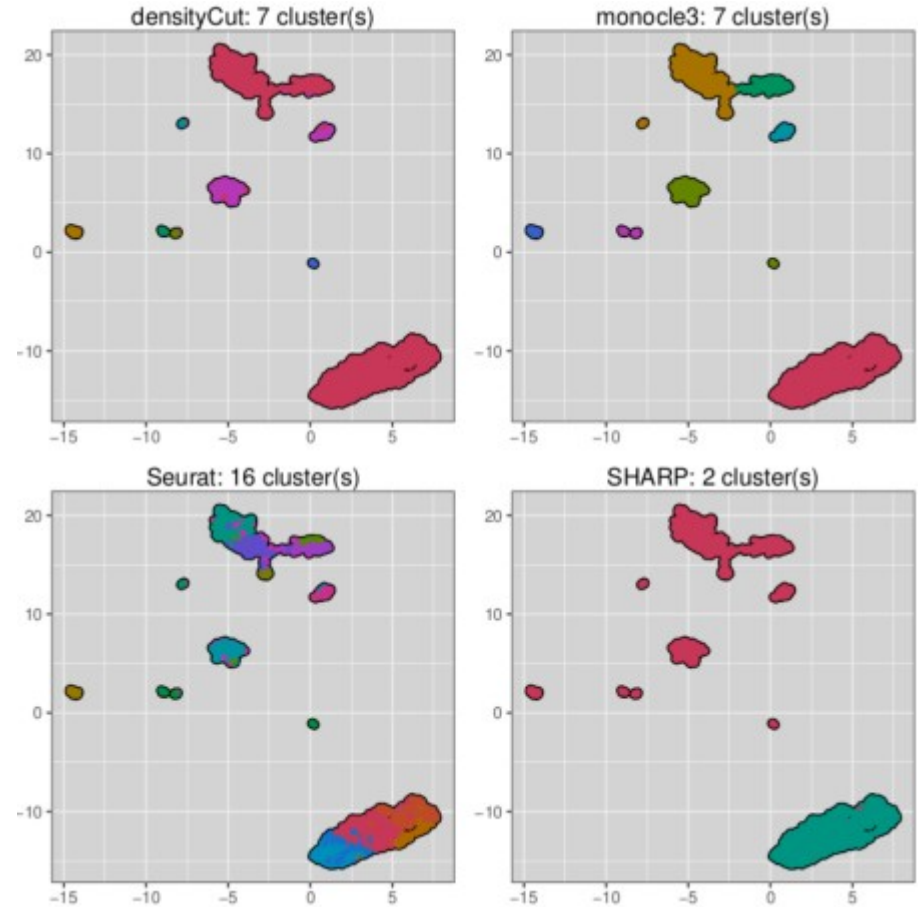
scRNA-seq datasets are **large** (20k features)

■ Overview of the scEVE algorithm

Feature selection

Prediction of base clusters

different methods yield ***different*** results
on ***different*** datasets

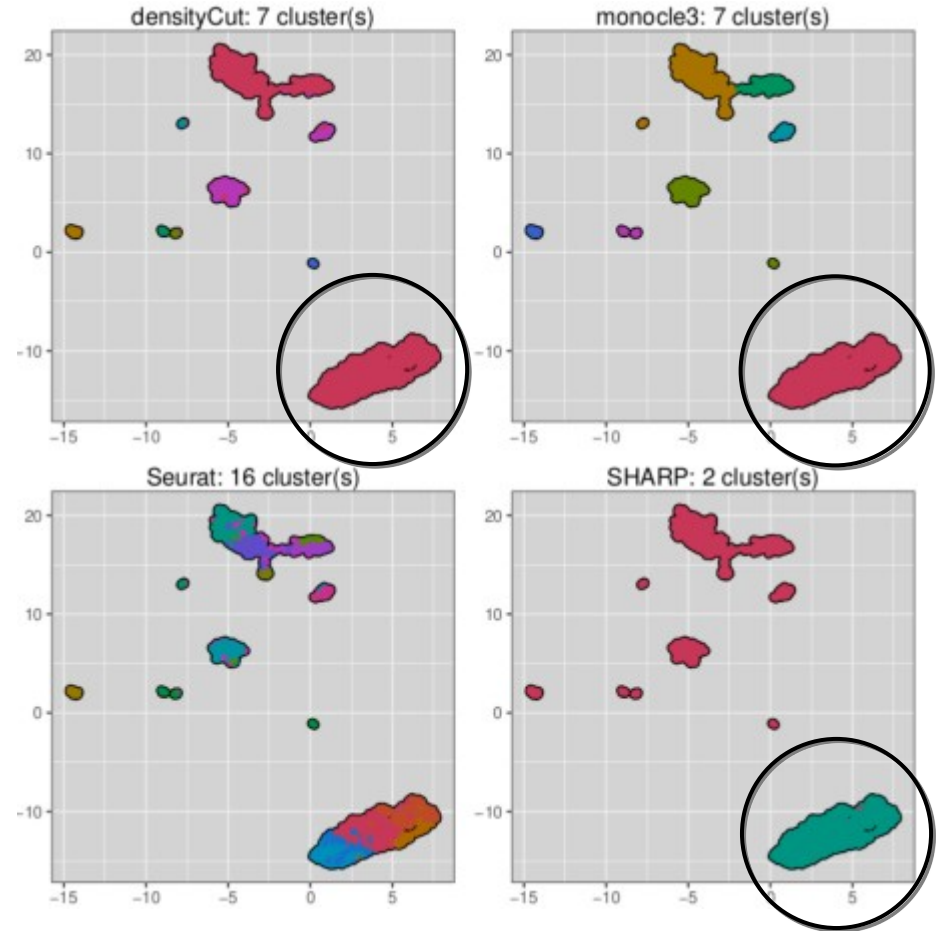


■ Overview of the scEVE algorithm

Feature selection

Prediction of base clusters

Detection of robust clusters



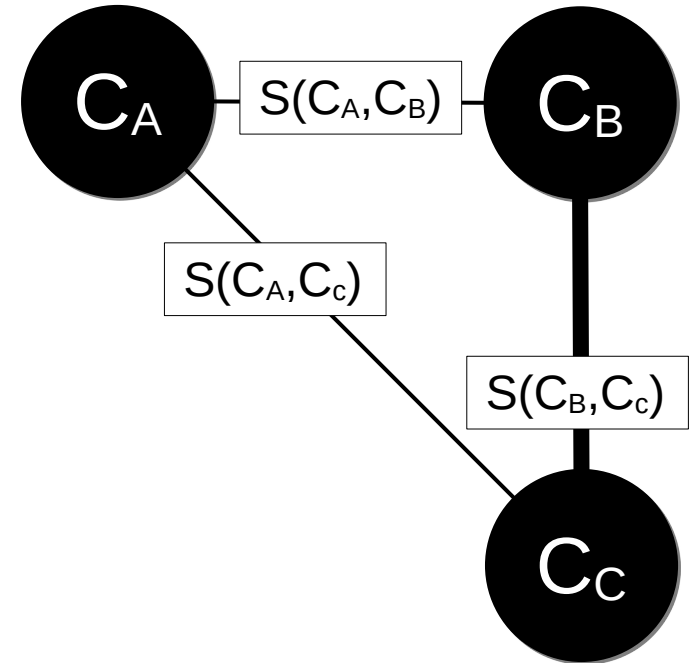
■ Overview of the scEVE algorithm

Feature selection

Prediction of base clusters

Detection of robust clusters

$S(C_x, C_y)$: **cells shared** by two clusters



■ Overview of the scEVE algorithm

Feature selection

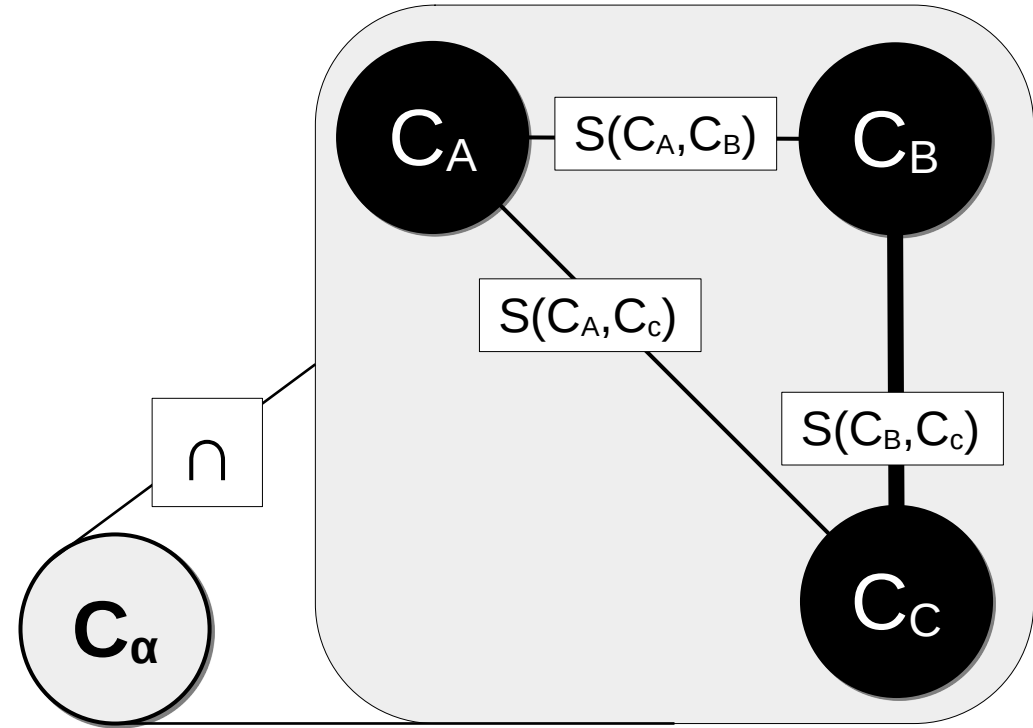
Prediction of base clusters

Detection of robust clusters

$S(C_x, C_y)$: **cells shared** by two clusters

C_α : $C_A \cap C_B \cap C_C$

robustness (C_α) : $f(S)$



■ Overview of the scEVE algorithm

Feature selection

Prediction of base clusters

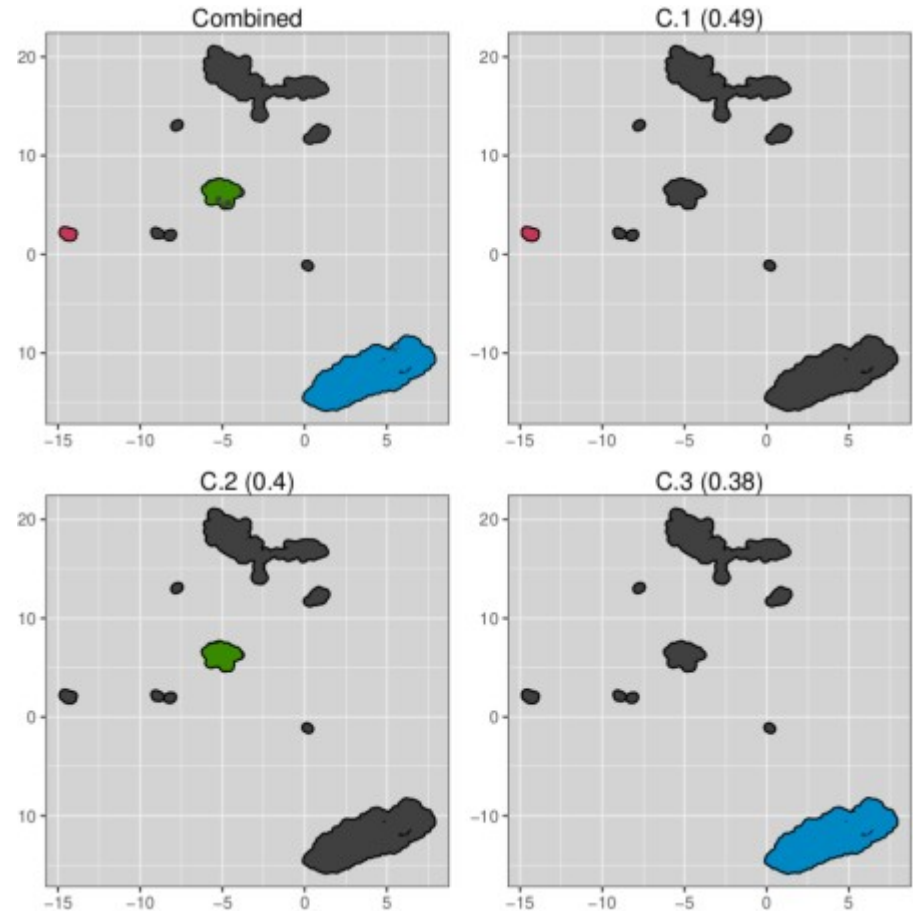
Detection of robust clusters

$S(C_x, C_y)$: **cells shared** by two clusters

$C_\alpha: C_A \cap C_B \cap C_C$

robustness (C_α) : $f(S)$

Cells unassigned to a robust cluster are merged in a **leftover cluster**.



■ Overview of the scEVE algorithm

Feature selection

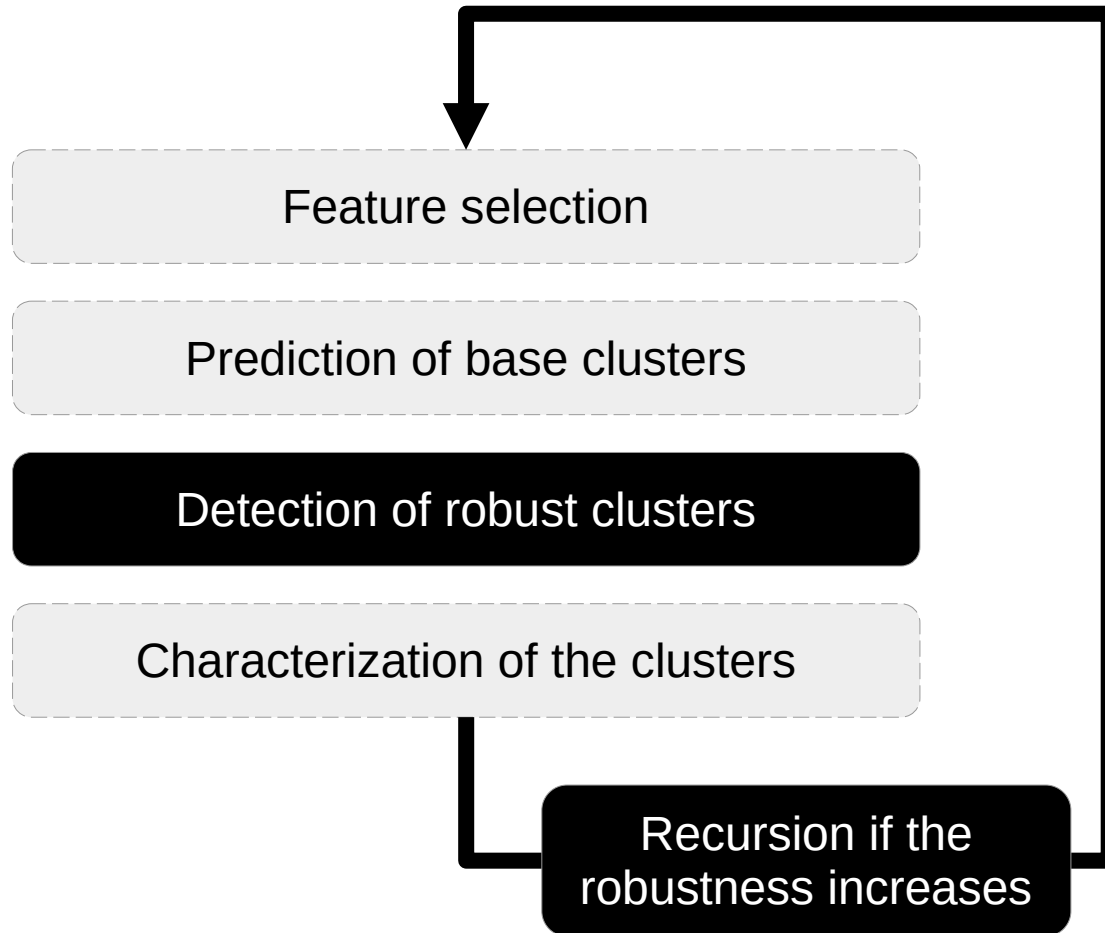
Prediction of base clusters

Detection of robust clusters

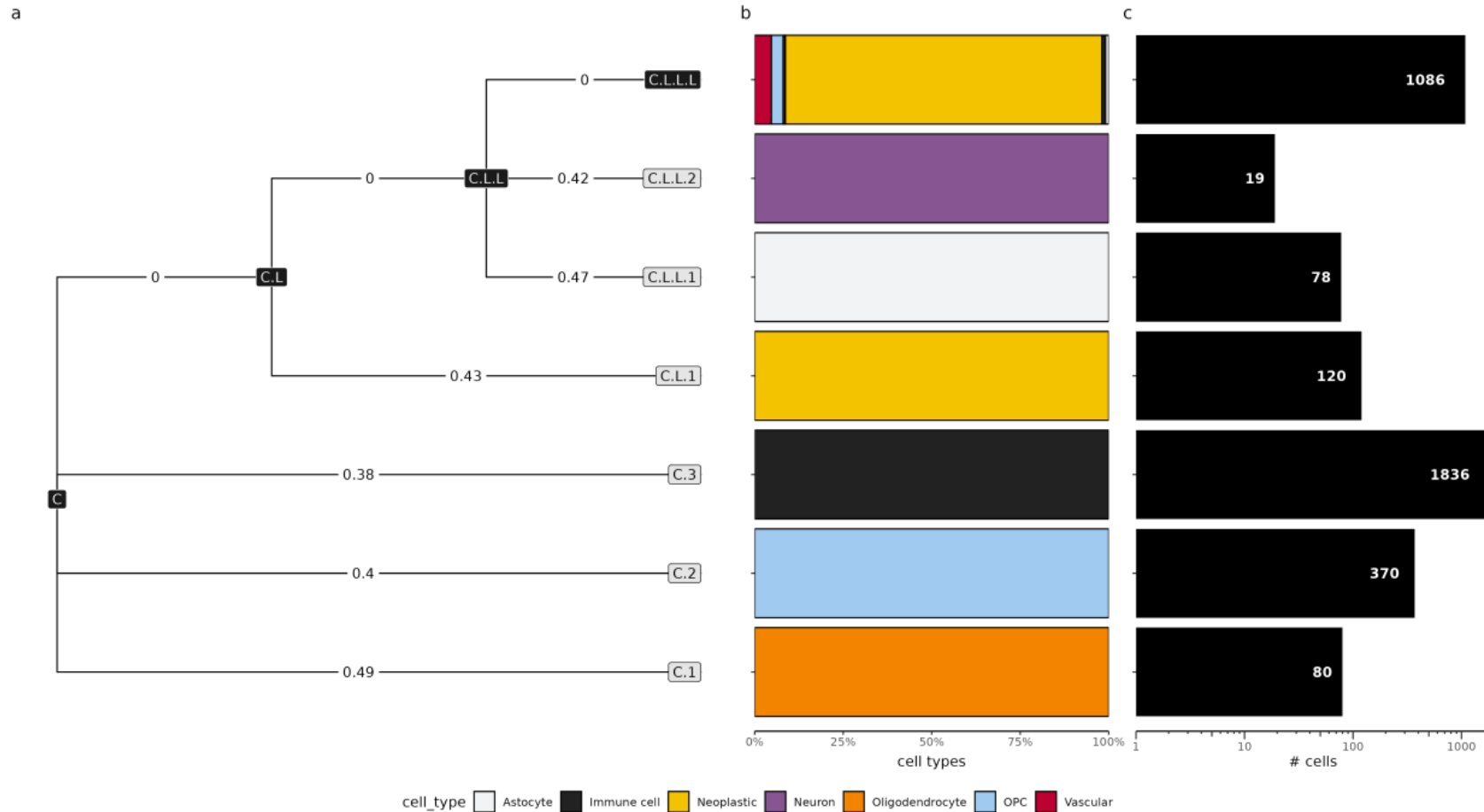
Characterization of the clusters

gene signatures are generated, or the clusters are filtered out

■ Overview of the scEVE algorithm

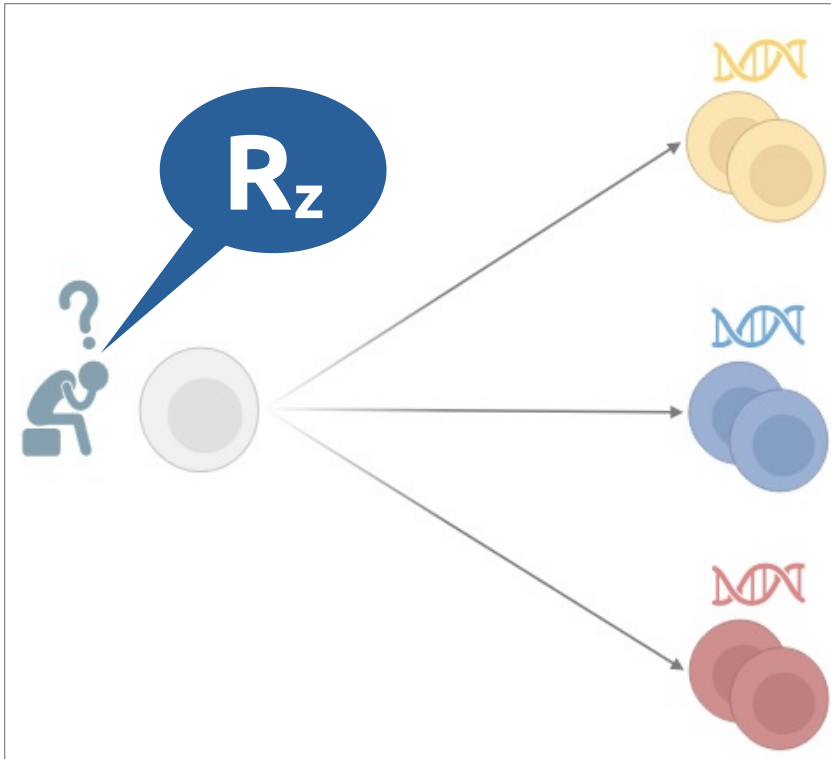


■ scEVE addressed conceptual challenges by leveraging Δ

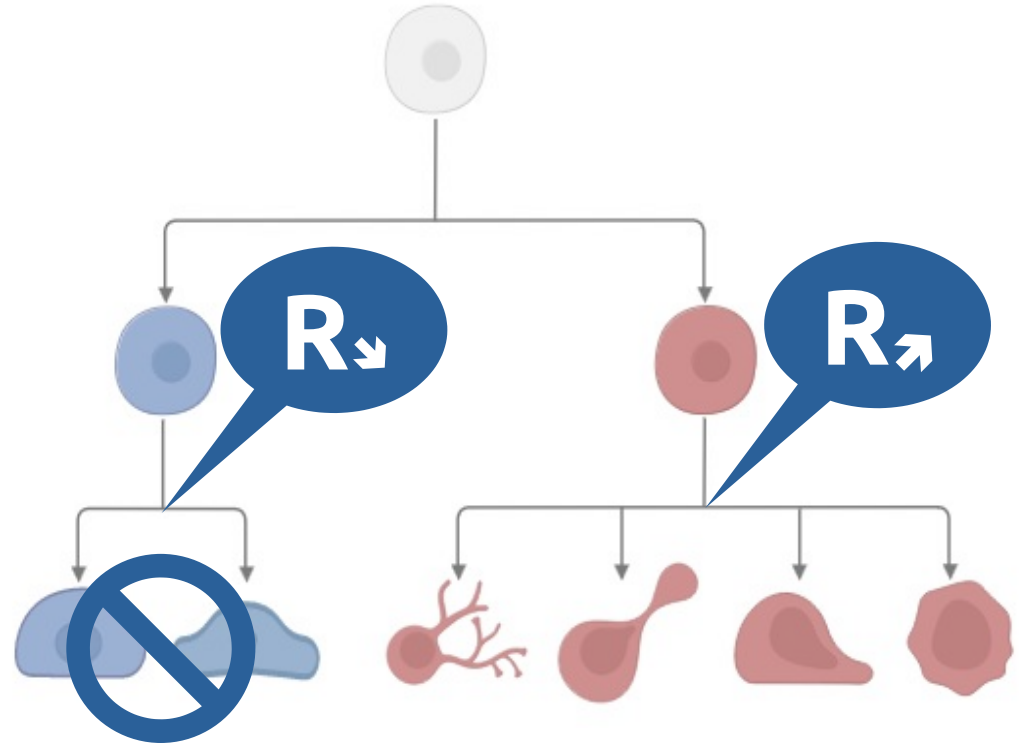


■ scEVE addressed conceptual challenges by leveraging Δ

Quantifying uncertainty of measurements and analysis results



Varying levels of resolution



* Challenges from Lähnemann et al. 2020.

■ The performances of scEVE are great...

Evaluation of scEVE with state-of-the-art (ensemble) scRNA-seq clustering algorithms

- scEVE had **good clustering performances** ; it even **ranked best** when its uncertain clusters were filtered out
- **Up to 10 times faster** than other ensemble algorithms
- **Up to 10 times less memory usage** than other ensemble algorithms

■ The performances of scEVE are great... but limitations exist

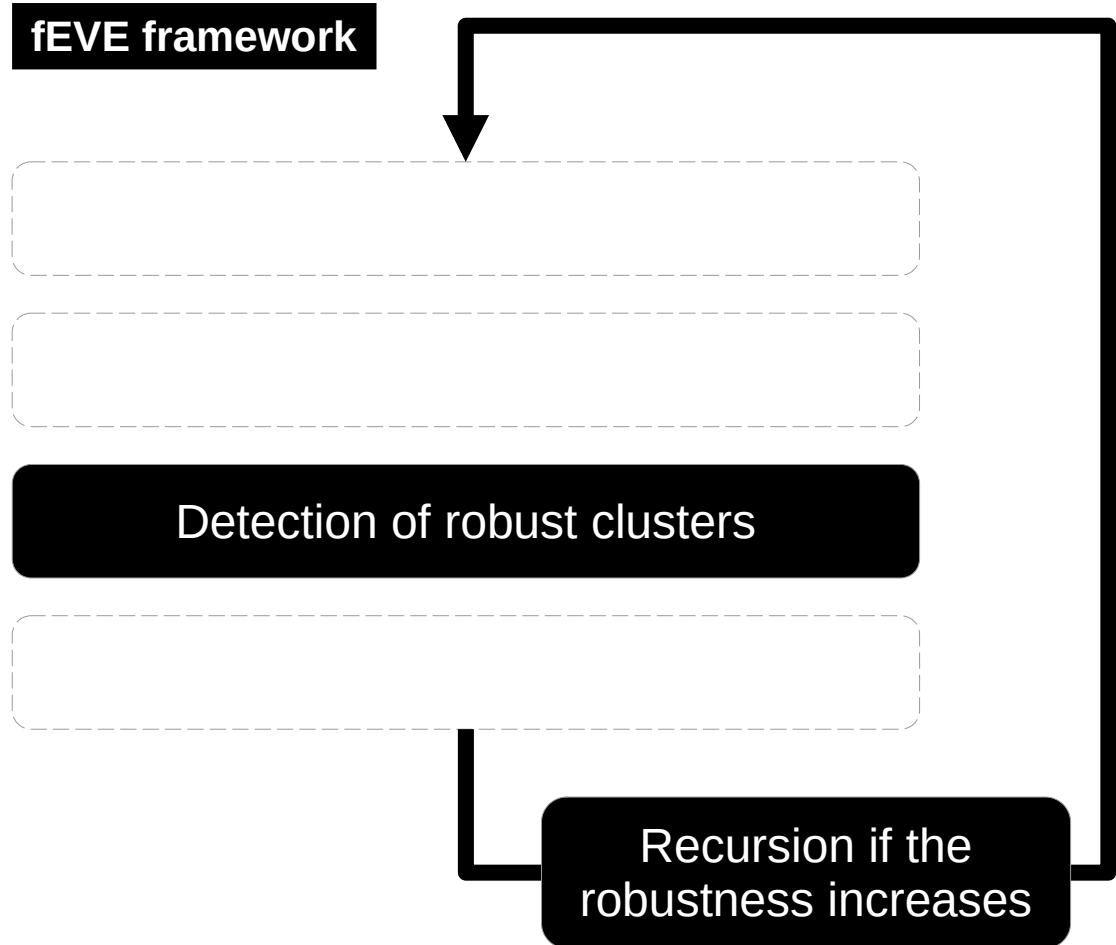
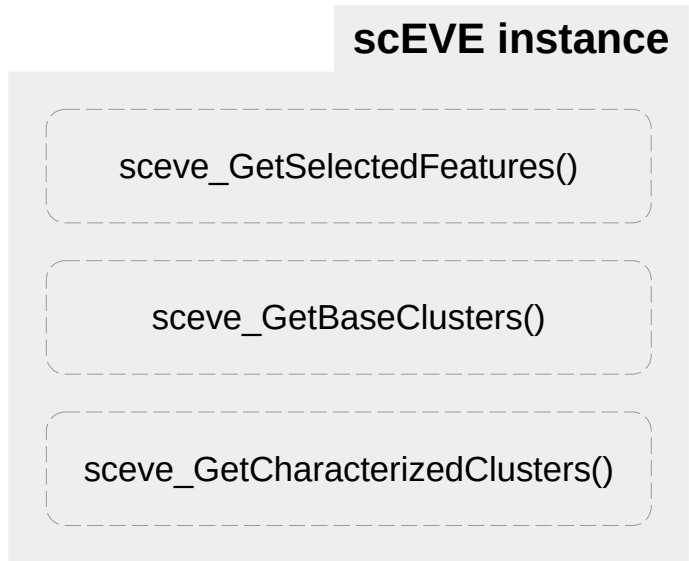
Evaluation of scEVE with state-of-the-art (ensemble) scRNA-seq clustering algorithms

- scEVE had **good clustering performances** ; it even **ranked best** when its uncertain clusters were filtered out
- **Up to 10 times faster** than other ensemble algorithms
- **Up to 10 times less memory usage** than other ensemble algorithms

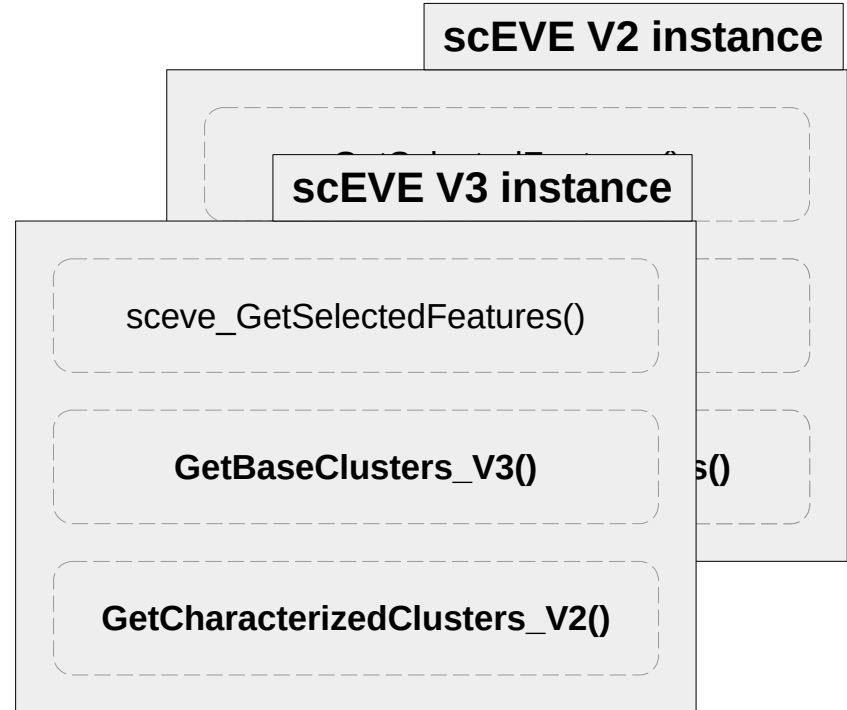
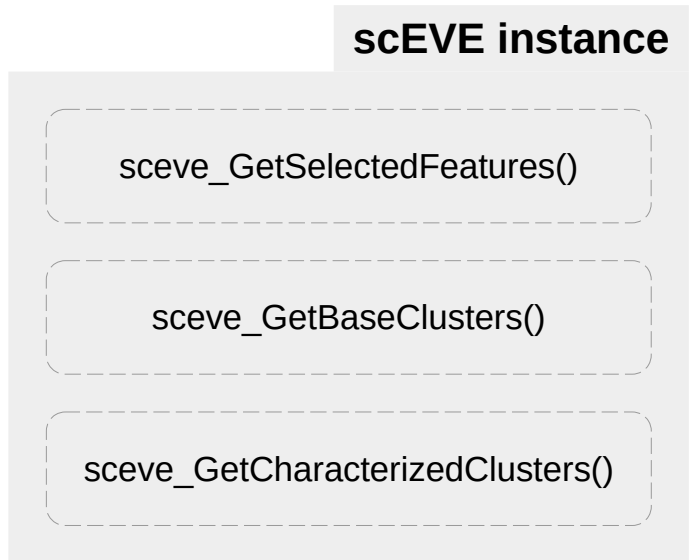
"There are many approaches to select features, predict base clusters or detect gene signatures"

"I found the scEVE algorithm interesting, but I don't work with scRNA-seq data"

■ The modular fEVE framework addresses these limitations

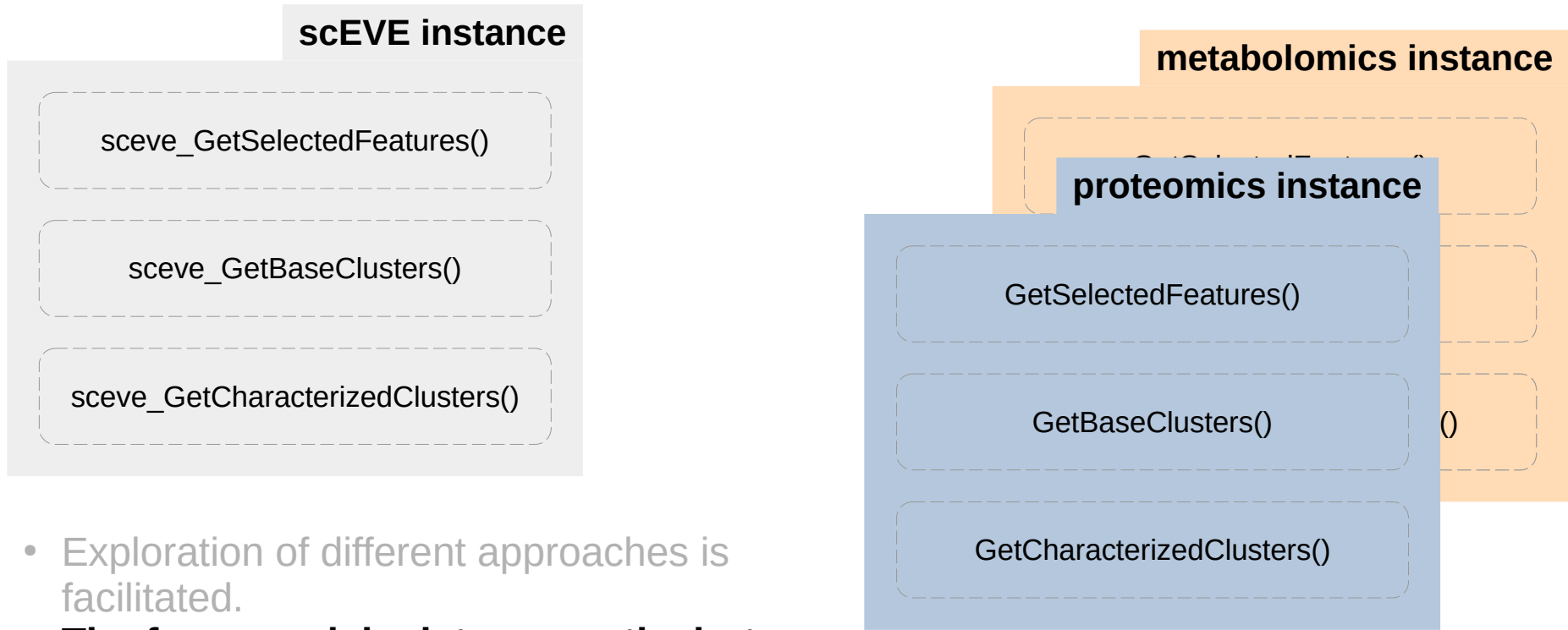


■ The modular fEVE framework addresses these limitations



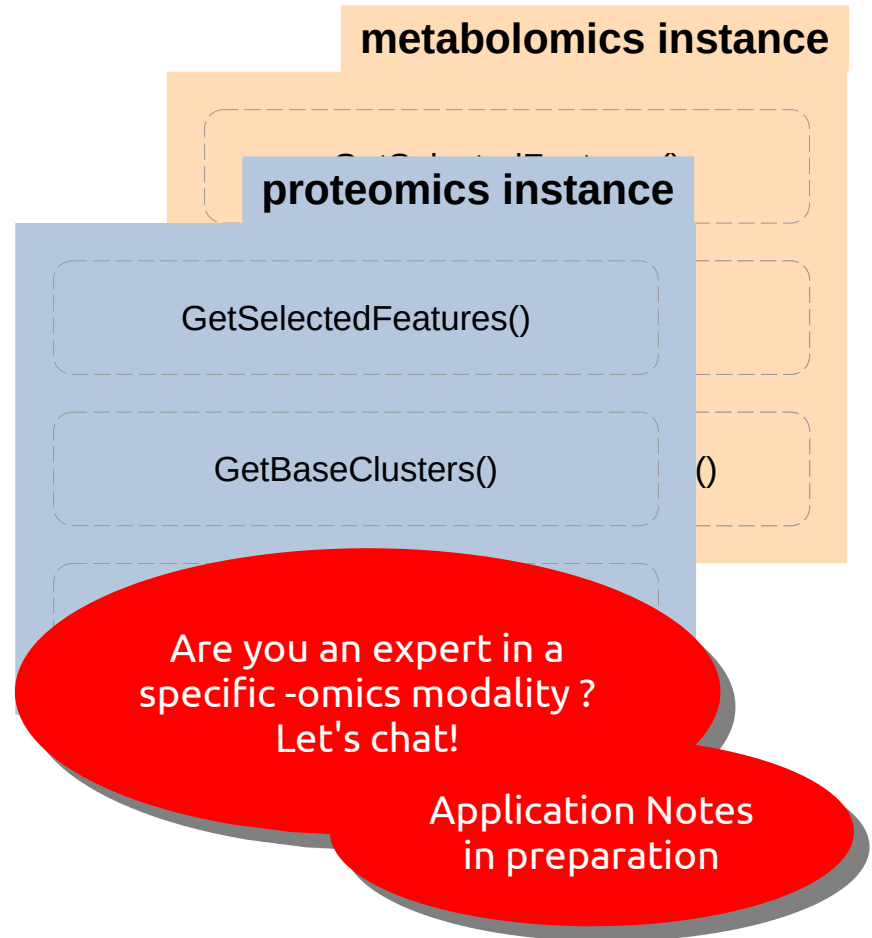
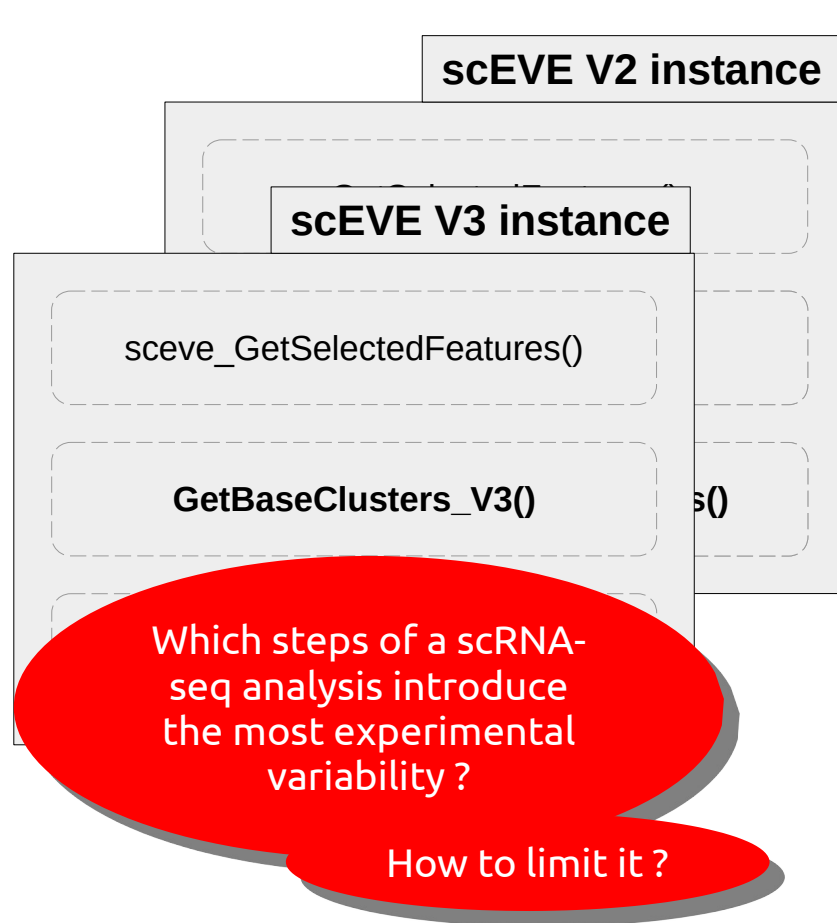
- **Exploration of different approaches is facilitated.**

■ The modular fEVE framework addresses these limitations



- Exploration of different approaches is facilitated.
- **The framework is data-agnostic, but its instances are not.**

■ The fEVE framework also raises new research questions



■ Take home messages

- We had developed a scRNA-seq ensemble algorithm able to **quantify the uncertainty of its results at multiple resolutions.**
(article is under revision)
- From it, we have developed **fEVE, a modular framework for ensemble clustering.**
It is **data-agnostic** and **easily customizable.**

Oh, and I'm looking for a new position in Fall 2025



<https://github.com/yanisaspic/fEVE>



yanis.asloudj@u-bordeaux.fr