Single Cell Transcriptomics

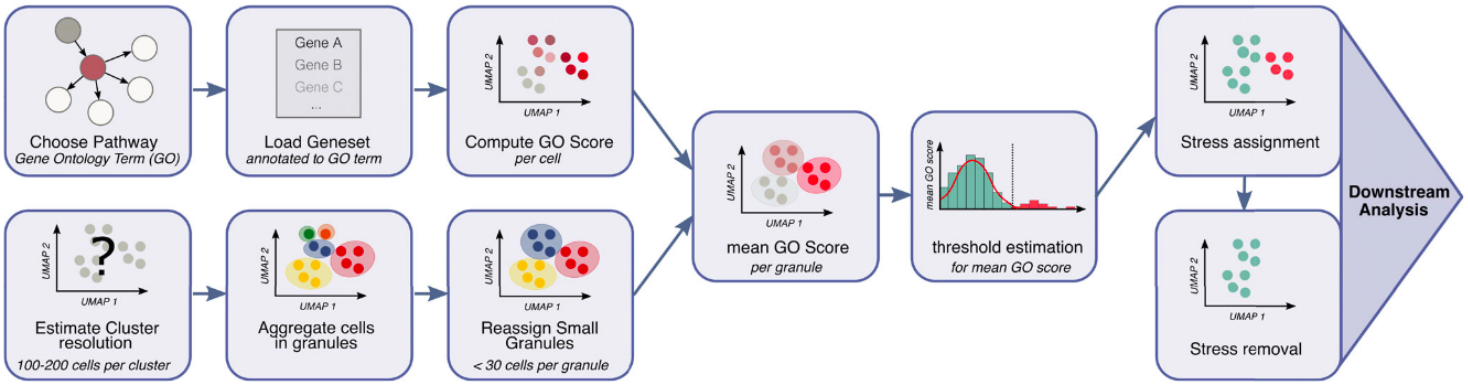
Newsletter August 2022

Paper 1

Vértesy, Á *et al.* [Gruffi: an algorithm for computational removal of stressed cells from brain organoid transcriptomic datasets](), *The EMBO Journal,* 2022

The first paper presents a method for filtering stressed cells from scRNA-seq organoid datasets. The removal of stressed cells based on a gene expression pattern helps to get a better analysis of fetal developmental trajectories which better resembles *in vivo* data. The method, granular functional filtering (Gruffi) combines multiple scoring to determine if a cell should be labeled as stressed: a GO-score is combined with cell clustering based on granules (small clusters), into multiple granule scores which are used to assign a stress label. As a result of using this method, the removal of stressed cells caused more clear trajectories of fetal development.

The authors tested Gruffi extensively on brain organoids but the algorithm can also be applied to other organoid systems. To show this Vértesy *et al.* applied Gruffi on retinal organoids where they adjusted the stress scoring by including a different stress selecting score (stressors and metabolic pathways of the retina). So, by adjusting the stress scoring according to the prevalent stressors and unique metabolic pathways specific for your system, Gruffi can be applied to any organoid dataset.



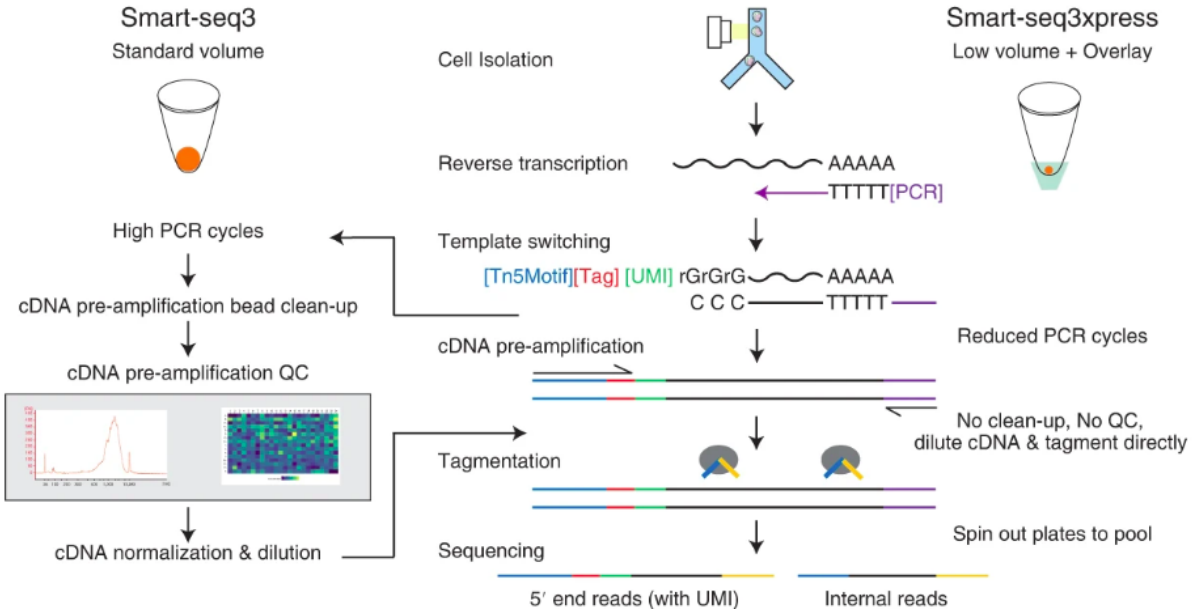
Paper 2

Hagemann-Jensen, M *et al.* [Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress](), *Nature Biotechnology,* 2022

The second paper is on a variant of Smart-seq3 but with reduced reagent usage and increased cellular throughput. Smart-seq3xpress is a scalable nanoliter implementation of Smart-seq3 and works with lower volumes due to the usage of an overlay (to cover the reagent with an inert hydrophobic substance).

They were able to obtain similar numbers of detected genes and molecules per cell on K562 and HEK293FT cells. When testing Smart-seq3xpress on 26260 human peripheral blood mononuclear cells (hPBMCs), higher gene detection was observed by the authors across cell types in comparison to Smart-seq2 and Smart-seq3. Additionally, more SNPs were covered in full-length Smart-seq3xpress data in a comparison of Smart-seq3xpress to droplet-based 10x Genomics on a matched hPBMCs donor. The full-transcript coverage also revealed cell type-associated isoform variation.

This method demonstrates a scalable solution for full-transcript-coverage scRNA-seq and allows for high-sensitivity scRNA-seq with isoform-specific and allele-specific resolution at a scale suitable for large-scale atlas building.

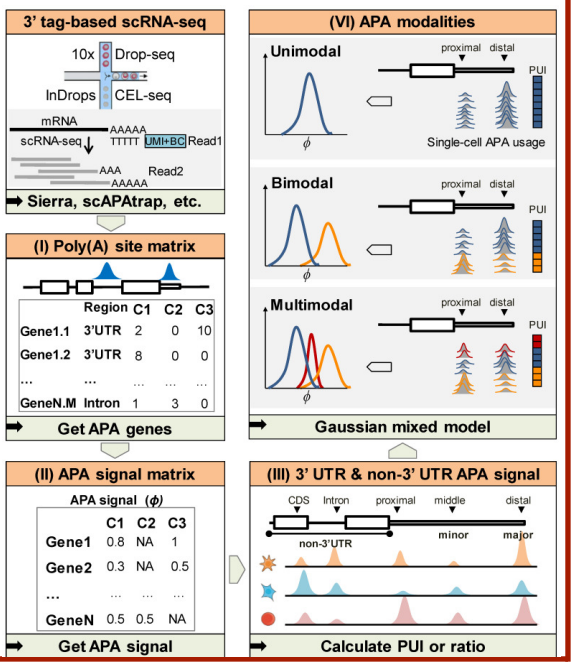


Paper 3

Qian, L *et al.* [scAPAmod: Profiling alternative polyadenylation modalities in single cells from single-cell RNA-seq data](), *International journal of molecular sciences*, 2022

The third paper is on a newly developed analysis framework based on a Gaussian mixture model, scAPAmod, which can be used to identify patterns of alternative polyadenylation (APA) (also called modalities of APA) from homogeneous or heterogeneous cell population at the single-cell level. 3' tag-based scRNA-seq data can be used with scAPAmod to distinguish the APA modalities. APA modality is a measure for cell-cell heterogeneity of APA usages in a cell population. This is different from APA dynamics which considers the differential use of APA sites between two cell populations or two cells.

The performance was evaluated using simulated data and scRNA-seq data. In the analysis of dynamic changes in the pattern of APA usage, it was found that the same gene has different patterns of APA usage in different differentiation stages. This study was able to profile the heterogeneous pattern of APA isoforms, in contrast to conventional analysis of single-cell heterogeneity, and could contribute to showing heterogeneity of single-cell gene expression with higher resolution.

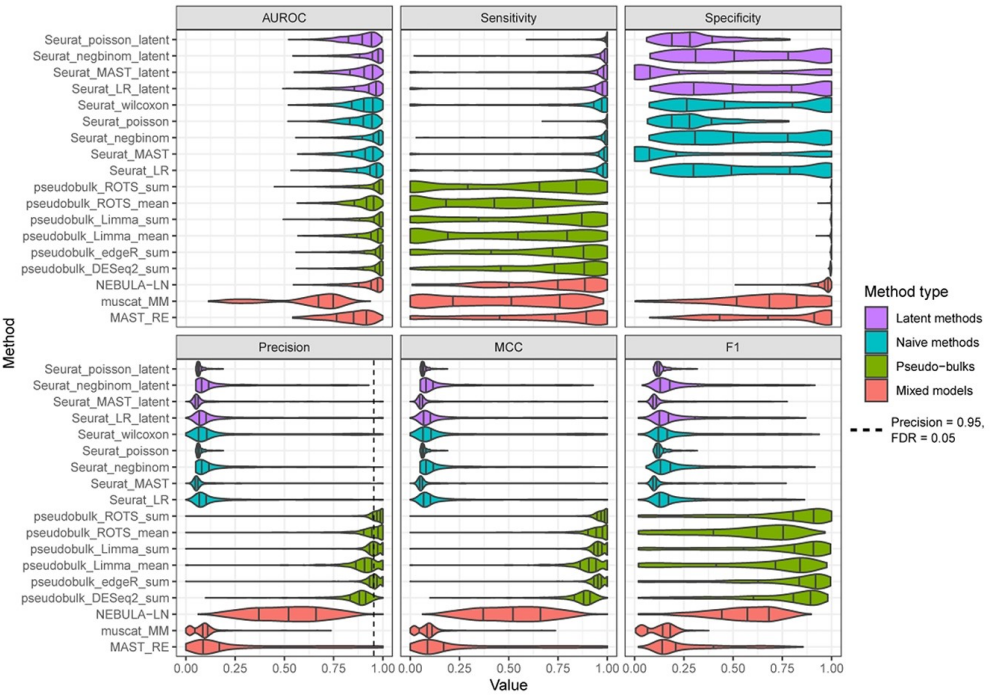


Paper 4

Junttila, S *et al.* [Benchmarking methods for detecting differential states between conditions from multi-subject single-cell RNA-seq data](), *Briefings in Bioinformatics,* 2022

The last paper describes a comparison of 18 methods for the identification of differential states (DS) between conditions from multi-subject scRNA-seq data. The comparison included three mixed models (MAST_RE, MUSCAT_MM, NEBULA-LN), six pseudobulk methods (edgeR, DESeq2, Limma and ROTS with sum and mean aggregation), and five naïve methods (Wilcoxon rank sum test, MAST, LR, negbinom, poisson). Scoring of the methods was based on the results of six gold-standard performance metrics: area under the ROC curve, specificity, sensitivity, precision, F1-score, and Matthew's correlation coefficient tested on multiple simulated datasets and a real dataset.

The results showed that the naïve methods were more susceptible to false positives than pseudobulk methods and mixed models. When looking at all performance metrics, the pseudobulk methods and mixed models performed better than the naïve methods, and the pseudobulk methods even outperformed the mixed models. The authors therefore recommended that scRNA-seq analysis pipeline developers should include pseudobulk methods and mixed models in their pipelines.



Next Single Cell Seminar

Date: TBA

If you would like to announce anything single cell related, being it job announcement, event, your published paper, technology development etc., please contact us.

Contact: kfr920@alumni.ku.dk