



TITLE: Effects of misclustering in multi-sample, multi-group scRNA-seq studies: a stability-based approach

SPEAKER: Gabriele Dolfi

COAUTHORS: Davide Risso

Università di Padova, Dipartimento di Scienze Statistiche

ABSTRACT: In scRNA-seq studies, statistical procedures have become increasingly convoluted. With the rise of complex experimental designs, the inferential results derived from such procedures can exhibit a lack of reproducibility and Type I error inflation. In particular, our work investigates FDR inflation in pseudo-bulk analysis for multi-sample multi-group studies, uncovering a major potential source of error termed 'differential misclustering'.

A stability-based approach is proposed to improve the robustness of the bioinformatic pipeline (e.g., clustering) and to adjust the final inference model to account for the aforementioned issue. Observation-level stability measures are employed as a tool to estimate the number of cell types in a tissue, to identify potential subtypes, and to construct aggregation weights for the pseudo-bulk model. The efficacy of the proposed approach was validated through simulations and applications to real data.

The results of this work underline the need for more robust statistical procedures in the context of scRNA-seq data analysis. On this basis, the proposed stability measures represent a valuable tool for establishing a reproducibility framework for complex scRNAseq experiments.

BRIEF SPEAKER BIO: Gabriele Dolfi is a recent graduate of the Master of Science in Statistical Sciences at the University of Padua. During his studies, he specialized in omics data analysis, both through his Master's thesis and his research at the BioAgingLab (Azienda Ospedale-Università di Padova).