

Title: Mapping scRNA-seq data onto cell type taxonomies

Author: Valentine Svensson

Abstract: The recent cell atlasing efforts has led to the definition of hundreds of new cell types. Maintaining interpretability of so many discrete classes is becoming a challenge, which can be addressed with taxonomic trees: broad cell types are successively divided into several fine grained types. No formulation has been made to link gene expression of individual cells to cell types in taxonomies. We present a generalized linear model using a novel likelihood defined on tree structured discrete categories. This model allows researchers to map new cells onto a taxonomy, evaluate the certainty of taxons, identify cell type markers, and investigate differential expression between cell types. With generalized linear models, accounting for nuisance factors (e.g. batch effects or technical variation) is straight forward. When trained on a global cell atlas, this model provides both a resource for comparative analysis and a summary of what is known about cell types globally.