ScTralign: a computational method to align and compare biological development trajectories across conditions from single cell RNA sequencing data

Understanding the biological processes underpinning lifecycle transition stages in parasites is crucial not only to further knowledge on the parasite itself, but also in identifying possible targets for therapeutics.

Single cell RNA sequencing (scRNA-seq) is becoming a more common approach for studying such life cycling transitions, as it is possible to capture gene expression profiles of individual cells at varying stages of the lifecycle and dissect them from the mixed populations into clusters of similar cells. A crucial element to gain full insight from such scRNA-seq experiments is to compare datasets from different conditions of development, such as comparisons of mutant and wild type parasites, and of different parasite species. A common way to try and integrate such datasets and identify differences between cell clusters across conditions. However, this may force similarity between cells, possibly obscuring processes unique to one of the datasets, and some expression differences might not be properly captured at the cluster level.

We present scTralign, a method that allows the alignment of two independently generated linear scRNA-seq trajectories irrespective of whether or not they have differing cell populations and process kinetics. The method creates a common pseudotime axis for the two datasets, separating cells which have no equivalent on the opposing trajectory from cells that are similar across both, thereby allowing variation to be preserved and identified between both datasets. The user can then identify genes that are differentially expressed between cells undergoing processes shared between the two conditions to see if expression patterns differ between the datasets, which might be causing later splits in development between the conditions.

We verify the process of scTralign with simulated scRNA-seq datasets, as well as applying it to two conditions of *Trypanosoma brucei* development to identify genes that allow slender parasites to transition into stumpy forms and to compare the lifecycle of *Plasmodium berghei* in different organs.

Compared with other trajectory alignment methods, scTralign correctly identifies the overall alignment of the real and simulated data, identifying where the processes begin to differ from one another. We also compare our method with the current methods used to extract differentially expressed genes across different conditions, showing scTralign captures more information from the dataset than contemporary methods.