Graph theoretic definition of cell types in high-dimensional single cell data

Will Macnair
Institute for Molecular Systems Biology, ETH Zurich, Switzerland - macnair@imsb.biol.ethz.ch

Laura de Vargas Roditi
Institute for Molecular Systems Biology, ETH Zurich, Switzerland - devargas@imsb.biol.ethz.ch

Manfred Claassen*
Institute for Molecular Systems Biology, ETH Zurich, Switzerland- mclaassen@ethz.ch

Sensible cell-type definition and assignment in heterogeneous cell populations in cancer and immune biology constitutes an important, though challenging task. Experiments with single-cell resolved readouts, such as mass cytometry, provide dozens of parameters simultaneously for a single cell, making data driven cell-type definition possible. However, the high dimensionality of such data constitutes a challenge for data analysis and requires development of computational methods enabling and interpretation of the multiparametric single cell data. Previous techniques [1, 2] have been developed for this purpose but they suffer from incorporate little to no biological prior knowledge, such as tree structures induced by differentiation mechanisms.

We present treeSNE enabling separation of functionally different cell types, while incorporating the potential differentiation and signaling mechanisms shaping the cell type repertoire. treeSNE learns a robust graph based metric on high-dimensional single-cell measurements, and visualizes these measurements by means of neighborhood relation preserving dimensionality reduction or graph layout techniques. This approach lends itself to automatic definition of cell-type composition of poorly characterized heterogeneous cell populations guided by differentiation mechanisms, such as primary tumors. We applied treeSNE to synthetic data, peripheral mononuclear blood cell and tumor samples and find that it recapitulates previously identified cell subpopulations. Furthermore, we investigate the context dependent mutual arrangement and graph theoretic properties of cell types in these cell populations and identified novel properties to sensitively and specifically detect rare and yet important cell types such as stem cells.

treeSNE separates functionally different cell subpopulations in large highdimensional datasets. We expect this approach to contribute to the elucidation of cell type composition of poorly characterized heterogeneous cell populations. Such characterization is valuable in a clinical context, where it can help improve accuracy and treatment options of cancer subtype classification.