ITM Web of Conferences **5**, 00010 (2015)
DOI: 10.1051/itmconf/20150500010

Current Challenges in Cell-Type Discovery Through Single-Cell Data

Laura De Vargas Roditi[1,3,a], Will Macnair[2], and Manfred Claassen[3]

[1] *Institute of Molecular Systems Biology, ETH Zurich, Auguste-Piccard-Hof 1, 8093 Zurich, Switzerland*
[2] *Swiss Institute of Bioinformatics (SIB), Zurich, Switzerland*

Abstract. Single cell sequencing and proteome profiling efforts in the past few years have revealed widespread genetic and proteomic heterogeneity among tumor cells. However, sensible cell-type definition of such heterogeneous cell populations has so far been a challenging task. Single cell technologies such as RNA sequencing and mass cytometry provide information precluded by conventional bulk measurements and have achieved significant improvements in multiparametricity at high cellular throughput. By combining these technologies with computational and mathematical techniques it is possible to quantitatively define cellular heterogeneity, uncovering distinct phenotypic profiles that can be utilized to, for example, characterize tumor heterogeneity with the potential to develop and improve therapeutic strategies.

1 Current Methods and Challenges in Analyzing Single-Cell data

Various approaches have been used to characterize cellular heterogeneity from single cell data (1,2), based typically on a definition of a cell type as a characteristic abundance profile of a set of molecular markers (gates). These profiles have been either defined from prior knowledge (3) or derived in a data driven fashion by means of conventional clustering techniques (4). Other techniques addressing this task cover generic dimensionality reduction (5) and clustering techniques that account for differentiation mechanisms in an algorithmic fashion (6). However, the high dimensionality and complex biological variability of these single cell data require development of novel computational approaches for quantification and interpretation of the results.

To date such approaches have not sufficiently taken into account the non-linear continuous nature of biological processes, which give rise to intermediate cell states observed during transition between persistent cell states. Conventional clustering techniques (7) make rigid implicit assumptions on the shape of cell subpopulations and not only ignore but are confused by the occurrence of such intermediate states(8). Recent computational approaches took advantage of intermediate cell states to define cell types from single cell data for linear and general bifurcated processes (8, 9). While the latter approaches permit a wide spectrum of geometric arrangements of heterogeneous cell populations, these approaches lack robustness and reproducibility due to the stochastic nature of the underlying algorithm. Dimensionality reduction techniques such as principal component analysis and t-distributed Stochastic Neighbor Embedding (t-SNE) shift the interpretation of nonlinearities and cell population

[a] e-mail: delaura@ethz.ch

clusters to visual inspection of distance-preserving projections of the high dimensional single cell data(5). Besides a first recent systems theory-based attempt to detect bifurcation events from single cell transcriptomics data of developmental processes(10), no constructive and robust approach has been able to objectively describe nonlinear geometries and trajectories for heterogeneous cell populations. Cellular differentiation, in particular, hematopoietic differentiation can follow a nonlinear bifurcated topology(11), where hematopoietic stem cells at the root give rise to a multitude of cellular types through division and differentiation following a branching pattern. While some attempts have been made to use the branching way in which cellular populations are generated in vivo to motivate computational techniques (12), this approach makes an incomplete use of this prior expectation: cells are clustered according to phenotype, however the technique does not recapitulate their bifurcated trajectory.

We believe that it is not only important to identify cellular types, but also the cellular trajectories taken through division and differentiation that lead to a particular cell type. This would have the potential to identify a "cell of origin" in tumors, which has been a long-debated subject motivating many attempts to identify a cellular type capable of giving rise to invasive cancer (13, 14). In order to achieve this, future single-cell analysis methods need to further explore reliable prior biological knowledge. As previously mentioned there has been an attempt to describe linear trajectories (15) however, differentiation processes are not strictly linear and it is important to be able to describe bifurcated trajectories as well. Furthermore, other processes that display heterogeneity may be of interest beyond differentiation, and description of cellular processes should also encompass topologies which are bifurcated and cyclic. Notably, tumor samples are expected to exhibit a variety of different topological cell type arrangements which may be unknown. Future techniques will face the challenge of making robust predictions about cellular types while being flexible enough to accommodate various possible topologies describing different cellular processes.

References

[1] Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, Hicks J, Wigler M. 2011. Tumour evolution inferred by single-cell sequencing. Nature 472: 90-4

[2] Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, Sim S, Okamoto J, Johnston DM, Qian D, Zabala M, Bueno J, Neff NF, Wang J, Shelton AA, Visser B, Hisamori S, Shimono Y, van de Wetering M, Clevers H, Clarke MF, Quake SR. 2011. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotech 29: 1120-7

[3] Maecker HT, McCoy JP, Nussenblatt R. 2012. Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol 12: 191-200

[4] Aghaeepour N, Finak G, Flow CAPC, Consortium D, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH. 2013. Critical assessment of automated flow cytometry data analysis techniques. Nat Methods 10: 228-38

[5] Amir el AD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, Shenfeld DK, Krishnaswamy S, Nolan GP, Pe'er D. 2013. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol 31: 545-52

[6] Qiu P, Simonds EF, Bendall SC, Gibbs Jr KD, Bruggner RV, Linderman MD, Sachs K, Nolan GP, Plevritis SK. 2011. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotech 29: 886-91

[7] Lo K, Brinkman RR, Gottardo R. 2008. Automated gating of flow cytometry data via robust model-based clustering. Cytometry Part A 73A: 321-32

[8] Qiu P, Simonds EF, Bendall SC, Gibbs KD, Jr., Bruggner RV, Linderman MD, Sachs K, Nolan GP, Plevritis SK. 2011. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol 29: 886-91

[9] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32: 381-6

[10] Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, Yuan G-C. 2014. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proceedings of the National Academy of Sciences 111: E5643-E50

[11] Bendall SC, Simonds EF, Qiu P, Amir el AD, Krutzik PO, Finck R, Bruggner RV, Melamed R, Trejo A, Ornatsky OI, Balderas RS, Plevritis SK, Sachs K, Pe'er D, Tanner SD, Nolan GP. 2011. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332: 687-96

[12] Levine Jacob H, Simonds Erin F, Bendall Sean C, Davis Kara L, Amir E-ad D, Tadmor Michelle D, Litvin O, Fienberg Harris G, Jager A, Zunder Eli R, Finck R, Gedman Amanda L, Radtke I, Downing James R, Pe'er D, Nolan Garry P. 2015. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 162: 184-97

[13] Blanpain C. 2013. Tracing the cellular origin of cancer. Nat Cell Biol 15: 126-34

[14] Polak P, Karlic R, Koren A, Thurman R, Sandstrom R, Lawrence MS, Reynolds A, Rynes E, Vlahovicek K, Stamatoyannopoulos JA, Sunyaev SR. 2015. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature 518: 360-4

[15] Bendall SC, Davis KL, Amir el AD, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Pe'er D. 2014. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157: 714-25