Stress The System: single-cell RNA sequencing on a panel of 9 different cell types from in-vivo, in-vitro, and clinical samples

Kunderfranco P(1)[†] , Mazza EMC(1), Carriero R(1), Soldà G(1,2), Martini E(1), Kallikourdis M(1), Condorelli G(1), Di Pasquale E(1,3), Crasto S(1), Pozzi D(1,2), Matteoli M(2,4), Calcaterra F(1,5,6), Mavilio D(1,5,6), Correnti M(1), Raggi C(1,7), Chiereghin C(1), Duga S(1,2), Savino B(1), Locati M(1,6), Lodato S(1,2), Mancinelli S(1), Scamardella E(1), Lugli E(1), Cibella J(1), Peano C(1,3,§)

(1) Humanitas Clinical and Research Center, Rozzano, Milan, Italy;
(2) Humanitas University, Rozzano, Milan, Italy;
(3) Institute of Genetic and Biomedical Research UoS of Milan, National Research Council, Rozzano, Milan, Italy;
(4) Institute of Neurosciences, National Research Council, Rozzano, Milan, Italy;
(5) Unit of Clinical and Experimental Immunology, Humanitas Clinical and Research Center, Rozzano, Milan, Italy;
(6) Department of Medical Biotechnologies and Translational Medicine (BioMeTra), University of Milan, Italy, Milan;
(7) Dipartimento di Medicina Sperimentale e Clinica, Università di Firenze, Italy
§ Corresponding author

[†] Email: clelia.peano@humanitasresearch.it

Motivation

Single-cell RNA sequencing (scRNA-seq) is transforming our understanding of developmental biology and gene regulation [1, 2], although substantial computational obstacles remain. In particular, integrated analysis of different scRNA-seq datasets, comparison of multiple cell populations and subpopulations and the integration of measurements and parameters still remains challenging. Thus, new datasets are needed to enable the application of new computational methods [3, 4] and to facilitate comparative analysis in order to guide future experimental and data-analysis settings towards gold standard procedures and pipelines. The aim of this study was to apply scRNA-seq on different cell populations from in-vivo, in-vitro and clinical samples in order to define experimental and data analysis guidelines to obtain good qualitative, quantitative, and reproducible single cell RNA sequencing data from different cell types.

Methods

scRNA-seq has been performed by using the 10x Genomics Chromium Single-cell system and the Single Cell 3' library preparation protocol on 8 different cell populations: CD8 naïve T cells isolated from healthy donors' peripheral blood, primary culture of hippocampal neurons established from E18 mouse embryos at 5 days of in vitro development, single cell suspension derived from dissociated somatosensory areas of the murine cerebral cortex at birth (P0), non-cardiomyocyte

cells from wild-type mice hearts, human Endothelial colony-forming cells (ECFCs) isolated from peripheral blood of adult healthy controls, tumor cells from a human intrahepatic cholangiocarcinoma cell line (HUCCT1), cardiomyocytes derived from human induced pluripotent stem cells (iPSCs), CD45+ immune cells FACS-sorted from human glioblastoma tumoral tissue; prostate-derived cells from human semen. After library preparation samples have been sequenced on an Illumina NextSeq 500 sequencer generating 26bp reads for 10x barcodes and UMI sequencing and 98 bp reads for transcript insert sequencing. We used Cell Ranger version 2.1 (10x Genomics) to process raw sequencing data and the Seurat suite version 2.0.0 [3, 4] for downstream analysis. For clustering, we first reduced dimensionality by principal-component analysis (PCA). We selected variable numbers of principal components (PCs) using either a permutation-based test or heuristic methods implemented in Seurat and performed clustering using methods implemented in Seurat.

Results

In total 20 samples have been sequenced at single-cell level by obtaining on average 100 million reads per sample, more than 70% of the samples showed a perfect single-cell behavior having more than 70% of the reads assigned to single cells after Cell Ranger demultiplexing step. We observed a cell viability reduction mostly related to the time occurred from sample sorting and loading into the Chromium single cell system, thus mostly evident in in-vivo and clinical samples. On average 4250 single cells have been recovered from each sample and a median number of detected genes per cell of 1770 has been obtained. The sequencing saturation was inversely proportional to the number of cells loaded into the Chromium and to the level of transcription characterizing different cell types. Different biological replicates of the same samples were analyzed, and different number of the same cell populations have been loaded in the system in order to perform a titration experiment and obtain an estimation of the sensitivity of the system respect both to the identification of different cell subpopulations and to the reproducibility of the results. We observed a good level of correlation between the abundance of cell subpopulations (R2= 0.76) when we loaded into the system 3000 and 1500 cells of the same biological samples. Moreover, the average gene expression level in the same cell subpopulations identified through the tSNE clustering strongly correlated (R2= 0.81), irrespectively of the number of cells loaded. Then we compared those samples showing the same level of sequencing saturation in order to derive from the scRNA-seq data an indication about the level of transcription of the different cell types, thus showing that the median number of detected genes per cell inversely correlated with the transcriptional behavior of the cells. We then focused our attention on a subset of samples in which we compared the same cell sample in two different conditions and by applying Seurat algorithm we were able to identify conserved cell types across conditions, allowing for comparative analysis to identify shifts in cell type proportion, as well as cell-type-specific transcriptional responses to different conditions. Results of this study will be useful to the scientific community interested in applying scRNA-seq technology to analyze different types of cells to correctly design future experiments and to obtain reliable and reproducible data.

References

1. Johnson, M.B. et al. Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human

cortex. Nat. Neurosci. 18, 1–30 (2015).
2. DeLaughter, D.M. et al. Single-cell resolution of temporal gene expression during heart development. Dev. Cell 39, 480–490 (2016).
3. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. NatBiotechnol. 2018 Apr 2.
4. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May 21; 161(5):1202-1214.