Name: Sen Wang

Instructor: Professor Alshaykh

EC 601 A1: Product Design in Electrical and Computer Engineering

Multimodal Single-Cell Integration

In this project, the research direction is focused on the integration of multimodal single-cell analysis. In order to have a better understanding of this project, it is first necessary to introduce the definition of multimodal single-cell analysis. As the name suggests, multimodal single-cell analysis (MSCA) is the simultaneous analysis of multiple data types in the same cell. In the last decades, thanks to the mastery of single-cell genomics, this has given researchers the possibility to measure DNA, RNA, and proteins in individual cells. This has given researchers the ability to discover new disease-associated cell types and cell-targeted therapeutic interventions. In recent years, as experimental equipment has been optimized, it has given researchers the possibility to observe multiple genomic patterns in the same cell. Meanwhile, as multimodal single-cell data become more abundant, multimodal single-cell analysis algorithms become increasingly important.

Before introducing MSCA, some introduction to multimodal single-cell data is necessary. Due to the small size of single cells, the measurements are discrepant. These variabilities often arise from differences in sampling depth between cell molecules and technical problems in processing cells in batches. In the meantime, the biological activity of a single cell is dynamic, not static. Nevertheless, most of the single-cell data nowadays are framed in static features. Notably, genetic regulation underlies the dynamic cellular processes that allow organisms to develop and adapt to changing environments. Researchers in existing single-cell data science often use "pseudo-time algorithms" to model dynamic biological processes. However, can this fully simulate actual biological processes? This is still a controversial issue. Above all,

a multimodal single-cell analysis algorithm that can encompass both temporal and biological dynamics is essential.

As mentioned above, researchers can infer biological differences between individual cells through single-cell multimodal analysis, thus expanding the understanding of cellular heterogeneity and its impact on cellular function. In early diagnosis and applications, by combining advanced medical detection devices, single-cell multimodal analysis greatly advances clinical research and applications, especially regarding early diagnosis and disease monitoring, by tracking the precise composition and distribution of specific cell populations with dynamic changes in cells. According to the literature, several projects have now integrated MSCA platforms to analyze patient-derived samples, for example in oncology, immunology, and neurology.

Let us focus our attention on oncology first. It is well known that tumors are composed of a large number of different cell types and cell states. Due to the heterogeneity between and within tumors, treatment resistance and tumor recurrence remain the most significant challenges in cancer treatment. Among them, recurrence remains the biggest challenge in cancer treatment. At the single-cell level, cells have different responses to drug therapy, and even a tiny number of drug-resistant cells in a tumor can affect the therapeutic effect. Because drug-resistant cells may be rare in tumors, there is an urgent need for susceptible methods to detect tumors. Therefore the single-cell multimodal analysis is of unique potential to reveal biological variants within tumors and to link specific acquired mutations to molecular outcomes. This is because, with single-cell multimodal analysis techniques, researchers can obtain the dynamic response of cells within the tumor upon receiving a therapeutic agent rather

than just a static time response. This enables early detection of treatment-resistant cells and subsequent rapid diagnosis of medical outcomes. Meanwhile, this technology can also significantly reduce the time researchers need to analyze samples.

Second, the topic is now shifted to immunology. The immune system is very complex, consisting of many types of cells, and has previously been defined by researchers primarily through cell surface markers and scRNA-seq. Although cell surface markers provide a good indication of cellular protein expression, a complete picture encompassing multiple cellular parameters is desirable to gain insight into the immune system. For example, lymphocytes communicate with their environment in various ways, including cell-cell interactions and the recruitment of new cells through chemokine secretion. The specific functions of various cell types in this ecosystem are known. However, only extensive data on the secretome are available, while cell-to-cell changes and their effects have not been fully explored to date. Note that once again, we refer to the dynamic exercise between biological systems. Researchers have recently discovered that by culturing cells in individual compartments in a multimodal analysis platform, it is possible to analyze not only RNA and intracellular proteins but also individual secretomes and then transcriptomes or genomes. This is the essence of single-cell multimodal analysis.

Last but not least, let's look at the application of single-cell multimodal analysis in neurology. Researchers believe that MSCA has the potential to unravel complex neurological interactions. In addition, it will be possible to study cell development and differentiation by following different cell trajectories to predict cell lineages and visualize the way these different cells are organized in neural tissues. Another parameter that may be integrated into neurology is the analysis of the

electrophysiological properties of cells to identify new cell subclasses and establish cell heterogeneity. These types of approaches are key stepping stones to assessing the full complexity of the nervous system. This analysis technique is certainly exciting news for neurology.

Since a mainstream algorithm for the integration of single-cell multimodal analysis has not yet emerged, let us present some of the two more important available methods. They are horizontal integration and vertical integration, respectively. Horizontal integration strategy. The horizontal integration strategy defines features as anchor points in unmatched experiments of the same type. This task is commonly faced in large-scale scRNA-seq projects where data are generated in multiple batches and techniques, as differences introduced by technical factors can lead to systematic biases in the distribution of observed RNA expression quantities (and even cell type composition). Researchers have found that the source of these technical differences can obscure relevant biological variability if not taken into account, thus complicating the interpretation of downstream analyses. Often, horizontal integration is formulated as a batch problem to eliminate undesired technical variation in experiments while preserving true biological variability within and between experiments. With the increasing number of reference maps at single-cell resolution, this is arguably one of the most essential steps in the single-cell analysis workflow.

Vertical integration strategy. Vertical integration strategies take advantage of the explicit assignment between molecular lineages in matched multimodal experiments, thus defining cells or groups of cells as anchors between data modalities. Vertical integration methods can be further divided into local and global approaches. Local analysis refers to the association between specific features of different

molecular layers, usually intending to detect putative interactions between them. Global integration, on the other hand, uses a full range of measurements to determine broader cellular states such as cell cycle phases and pluripotency potential. Global analysis typically identifies patterns of covariation across genomic features and layers.

In conclusion, single-cell studies in the past have been limited to static studies. Researchers could only analyze a single molecular structure at a single time in a single cell. It is also because of this limitation that it is difficult for researchers to break through some difficult medical problems, such as tumors. From my perspective, the single-cell multimodal analysis will hugely impact the biomedical field. Moreover, I hope to continue to study this topic in depth.

Reference

[1] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck III, W. M., Zheng, S., Butler, A., ... & Satija, R. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573-3587.

[2] Argelaguet, R., Cuomo, A. S., Stegle, O., & Marioni, J. C. (2021). Computational principles and challenges in single-cell data integration. Nature biotechnology, 39(10), 1202-1215.

[3] Källberg, J., Xiao, W., Van Assche, D., Baret, J. C., & Taly, V. (2022). Frontiers in single cell analysis: multimodal technologies and their clinical perspectives. Lab on a Chip, 22(13), 2403-2422.

[4] Dou, J., Liang, S., Mohanty, V., Miao, Q., Huang, Y., Liang, Q., ... & Chen, K. (2022). Bi-order multimodal integration of single-cell data. Genome biology, 23(1), 1-25.