Understanding the cell-to-cell heterogeneity in a multi-cellular organism through computational and analytical approaches

1. Background

Life starts from a single cell that divides as well as differentiates to develop into different parts of a multi-cellular organism. With the emerging sequencing technologies, it is evident that cell can possess different transcriptomes, epigenomes and proteomes even though they are derived from same cell line or present within a single tissue. Cell populations derived from the same tissue of an organism are heterogeneous. Cell-to-cell variability is responsible not only for diversity in cellular states but also affects interactions among numerous other distinct cells. Therefore, studies conducted at the single-cell level are needed to uncover the underlying complexity of biological systems, and to gain greater understanding of biological processes [1].

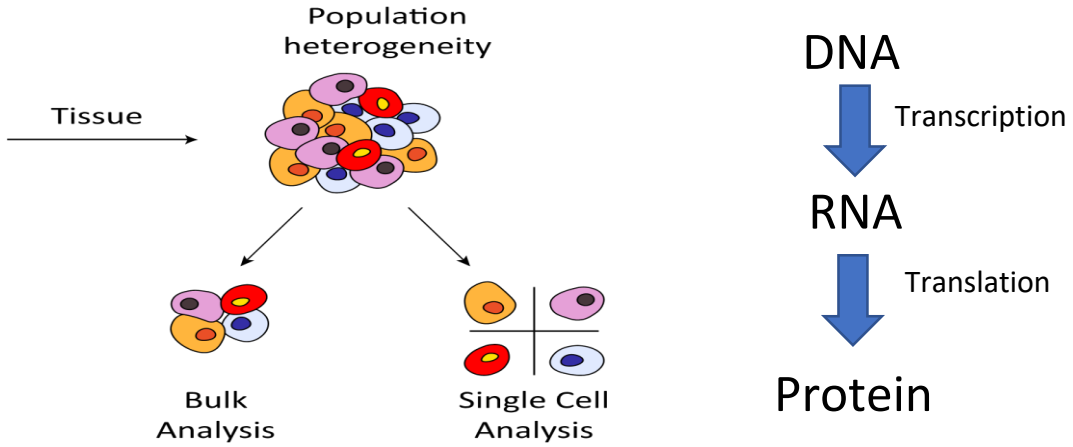

Figure 1. This single cell shows the process of the central dogma of biology, which are all steps researchers are interested to quantify (DNA, RNA, and Protein)

One of the important application areas of single-cell analysis is in cancer studies [2]. Cancer is a group of diseases which have a potential to spread to other parts of the body. It is difficult to prove a particular cause of cancer due to the absence of specific fingerprints. The diverse characteristics of cells resulted from mutation, differences in gene regulation, stochastic variation, or environmental perturbations; this is reflected at the omics level. These diverse characteristics of cell is increasingly considered as the cause of failure of cancer treatment and disease recurrence [3], and becomes important to understand tumor initiation, progression, metastasis and therapeutic responses. In this regard, single cell analysis may be able to functionally differentiate among normal and a cancer cell and also identify cells at various developmental stages.

Single Cell Analysis

The traditional RNA seq data is derived from bulk RNA, taken from a large number of cells. This leads to averaging effect, thus missing important signals that may be present in individual cells. Recently, technology for sequencing single cells has become available, and expression data at the resolution of

single cells is being produced. Such data sets are large both in terms of the number of cells (rows), as well are the number of genes (features/columns). The dimensionality often runs into tens of thousands.
This cell-to-cell heterogeneity can be understandable from the concept of RNA velocity [4], the gene expression state with respect to the time derivative. It can be directly estimated by distinguishing between unspliced and spliced mRNAs in common single-cell RNA sequencing protocols. This is a high-dimensional vector that predicts the future state of individual cells on a timescale of hours. Similarly, the goal of this project is to design a method for identifying the velocity of the pathways in a single cell that helps to understand the behavior of a cell with respect to time. Pathway velocity possess the idea of active or inactivation of different portion according to the change of expression of diverse biological molecules.

Single cell has introduced huge challenges and opportunities in the field of biology as well as technology [5,6]. In this regard advanced machine learning (ML) techniques including deep learning can be applied on multi-omics data.

2. Objectives

In this project we will address the following two objectives:

- To develop algorithms for estimating the pathway velocity from RNA velocity using single cell data for gaining a deeper understanding of cancer biology
- Effect of modification at protein levels in pathways of single cells

- **To develop algorithms for estimating the pathway velocity from RNA velocity using single cell data for gaining a deeper understanding of cancer biology**

Generally, to understand the molecular state of the unrevealed biological processes, up or downregulated genes are mostly identified by the researchers. On the other hand, biological pathways bridge the sequence and functional gap and provide a fully understandable molecular basis of disease [7]. The challenge in biological science is that, how a single cell processes information and responds to an abnormal situation. Our understanding is mostly based on the phenotypical changes of the cell population. However, cell-to-cell behavioral differences contribute in the overall nature of the population but not represent the behavior of a single cell. In order to understand the cellular heterogeneity, pathway analysis played an important role. From the idea of RNA velocity with respect to time derivative of the gene expression state the velocity of the pathways in a single cell can be derived. For an example, if a cluster of cells is considered those are closely placed in tissue and a window shift operation is performed on them it will provide the insight of the changing of pathways according to the change of expression of RNA molecules responsible for the particular pathway. Let us assume, we have cell 1, cell 2, cell 3, cell 4 and the size of the window is 2. If the window is shifted with respect to time the pathway may vary from cell to cell. This experiment can be performed in both the condition of cell i.e. normal as well as disease in a particular time if the activation of pathways varies in two different condition of cells that will provide an idea of the cause of that particular disease. The schematic diagram of the proposed method is described in Figure 2.

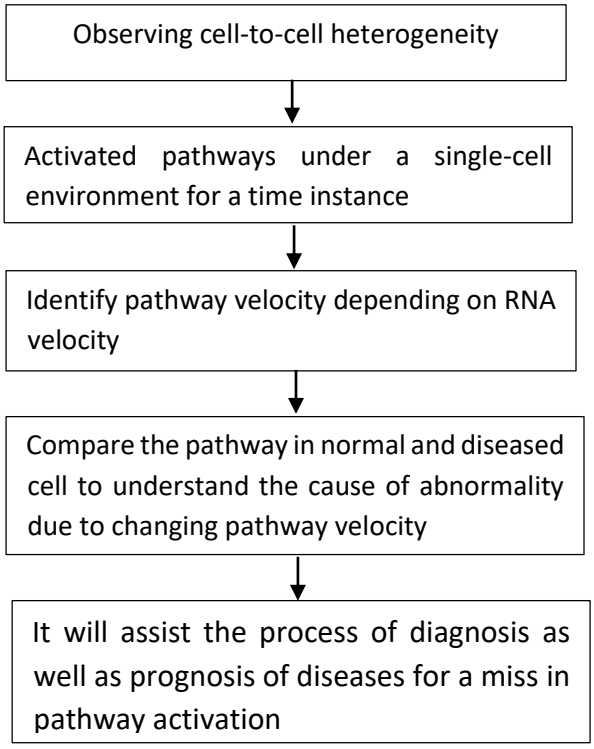

Figure 2: A schematic summarization of the proposal on single cell dependent cellular environmental variation in terms of pathway velocity based on RNA velocity

- **Effect of modifications at protein levels in pathways of single cells**

As mentioned in the last objective, the deviation due to RNA velocity is being observed and framed to predict the changes in case of diseased condition. Changes in mRNA expression and copy number variation are not always helpful for observing the changes in pathways in a single cell. Individually gene expression profiling has made a huge progress whereas protein expression profiling or modifications are not being observed properly in case of cell to cell heterogeneity in terms of pathway analysis. However, mapping gene expression profiling directly to protein expression in a single cell is not always biologically relevant. Due to post translational modifications [PTM] at protein levels, proteomic changes in terms of protein expression may occur. The heterogeneity in the expression level of the protein and their post translated isoforms may have implication for progression of diseases, resistance of applied therapeutic approaches, etc. For example, expression of tumor suppressor p53 and their phosphorylated isoform Serine-15 differ significantly across different cells in the same tumor tissue for colorectal cancers. PTM changes are directly associated with the structural changes of the proteins in a cellular environment [8]. Under structure-function paradigm, controlled protein-protein interactions should be affected by the above mentioned structural changes in protein sequences under single cellular environment. These functional variations affect different classes of pathways. Using computational intelligence, huge number of protein transcripts can be observed and analyzed. We want to design protein expression profiles considering the structural and functional changes

which are associated with variation at pathway level of single cell. Finally, these variations may also help to analyze the cell to cell heterogeneity even under disease condition only.

3. Data Sources

We want to use the publicly available dataset on single cell of 10x genomics [https://www.10xgenomics.com/10x-university/single-cell/].

Moreover, our lab has ongoing research collaboration with following institution/research lab from where we will try to get experimental data:

1. CISR-Indian Institute of Chemical Biology (IICB), India
2. Indian Statistical Institute (ISI), India
3. Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi), India
4. NIMHANS, Bangalore, India
5. Chemistry and Chemical Biology, Harvard University, USA
6. Thomas Jefferson University, Philadelphia, USA
7. Department of Biological Sciences, Middle East Technical University (METU), Turkey

4. Conclusion

Single-cell analysis has recently become an exciting and fast expanding field that has immense potential to improve the understanding of fundamental biological problems. In turns it help us to better understand the complexity of human disease. In this project, we will first try to estimate cell-to-cell heterogeneity depending on the mechanism like RNA velocity. Subsequently, we also try to understand the cell-to-cell heterogeneity based on protein expression profile in terms of structural and functional changes. These studies will help to facilitated better and biologically relevant prognostic, diagnostic and therapeutic outcomes.

5. References

1. D. Sengupta, N. A. Rayan, M. Lim, B. Lim and S. Prabhakar. Fast, scalable and accurate differential expression analysis for single cells. bioRxiv. 2016.
2. A. P. Patel and et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 344, 1396–401, 2014.
3. A. Kreso and J. E. Dick. Evolution of the Cancer Stem Cell Model. Cell Stem Cell. 13:3. 275-291. 2014.
4. G. L. Manno and et. al. RNA velocity of single cells. Nature. 560. 2018.
5. D. Sinha, A. Kumar, H. Kumar, S. Bandyopadhyay and D. Sengupta. dropClust: efficient clustering of ultra-large scRNA-seq data. Nucleic acids research. 46:6, e36-e36. 2018.
6. A. Mongia, D. Sengupta and A. Majumdar. deepMc: deep Matrix Completion for imputation of single cell RNA-seq data. bioRxiv. 2018.
7. S. Enayat, M. S. Ceyhan, A. A. Basaran, M. Gursel and S. Banerjee. Anticarcinogenic effects of the ethanolic extract of Salix aegyptiaca in colon cancer cells: involvement of Akt/PKB and MAPK pathways. Nutrition and Cancer: An International Journal. 65:10451058. 2013.
8. A. J. Magness and et. al. Multiplexed single cell protein expression analysis in solid tumours using a miniaturised microfluidic assay. Converging Science Physical Oncology. 3. 2017.

List of paper published/communicated by author

Journal

- Understanding the Evolutionary Trend of Intrinsic Structural Disorders in Cancer Relevant Proteins as probed by Shannon Entropy Scoring and Structure Network Analysis. BMC Bioinformatics. (Communicated).

Conference

- Identifying Potential Hubs for Kidney Renal Clear Cell Carcinoma from TF-miRNA-Gene Regulatory Networks. IEEE ASPCON 2018. Jadavpur University, India. (accepted).

Poster

- Understanding the Evolutionary Trend of Intrinsic Structural Disorders in Cancer Relevant Proteins as probed by Shannon Entropy Scoring and Structure Network Analysis. P-29. International Conference on Bioinformatics 2018. New Delhi. India. (accepted).