

### scRNA analysis pipeline

Single cell analysis is gaining momentum in the scientific community, as it offers cellular resolution that can help decipher complex biological systems. This increase in biological resolution leads to a corresponding increase in analytical complexity, demanding high-end bioinformatics expertise. To alleviate these constraints and provide complete bioinformatics support for scRNA-seq analysis, we have tested a range of tools and developed a standardized pipeline that can handle analysis of data from many different types of UMI-based scRNA-seq. Important factors to be considered for a seamless analysis are,

- Sequencing platforms provided software packages are usually hamstrung by proprietary algorithms that prohibit analytical customization and often mask the ways in which data are being manipulated.
- Novel approaches and tools have to be constantly explored, tested and updated within the pipeline.
- Scalability to datasets generated from various protocols.
- Data specific analysis and customization such as cell-cycle stage, pseudotime, normal vs tumor cells analyses, etc.
- Efficient debarcoding to attain maximum usable data.

### Dimensionality reduction

A gene-count matrix can contain millions of cells, making it extremely difficult to understand the underlying patterns. Reducing the number of dimensions by grouping cells with similar expression patterns is required to refine downstream analyses.

- Principal component analysis (PCA) is one of the most widely used methods for such dimensionality reduction.
- Accounting for relatively high number of zeroes (due to undetected genes per cell) in the single cell data, PCA-dependent methods to accommodate these "zero inflated" datasets is required.
- For clustering, the t-stochastic neighbour embedding (t-SNE) method, a non-linear clustering approach, is widely used to identify different clusters or sub-populations.

For the test case data, each sample represents expression values at different time-points during the mESC differentiation process, and as a result the specific time of collection of cells can lead to similar gene expression across cells at any given time point, causing clustering based on time of collection. Upon eliminating cell-cycle specific expression patterns and cell-type specific markers analysis revealed two distinct cell types: extraembryonic endoderm (XEN) cells and neuro-ectodermal (NEU) cells.

### Dynamic analysis - Pseudotime

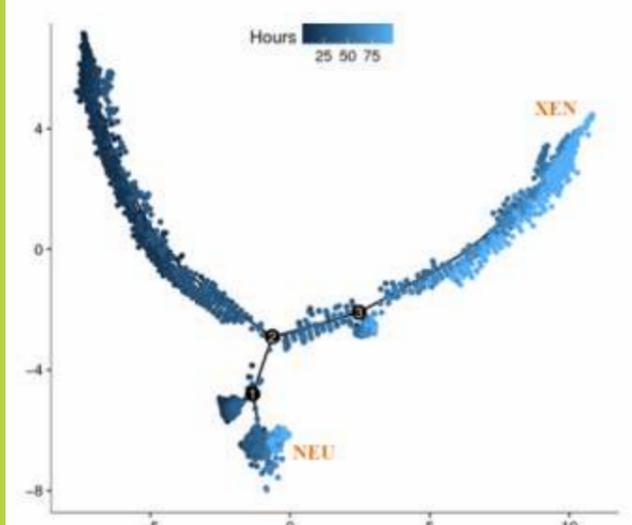
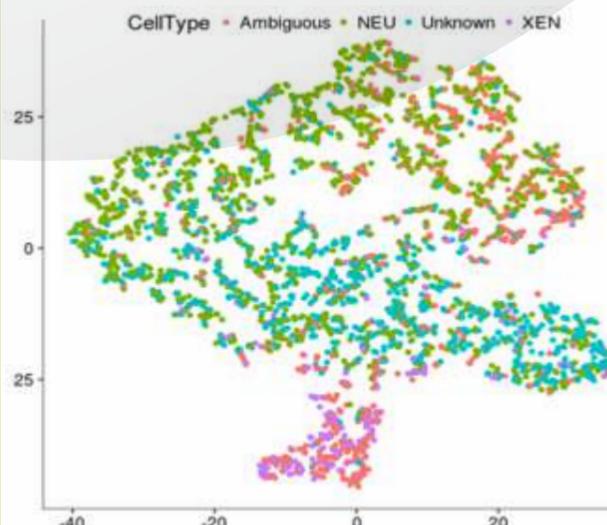
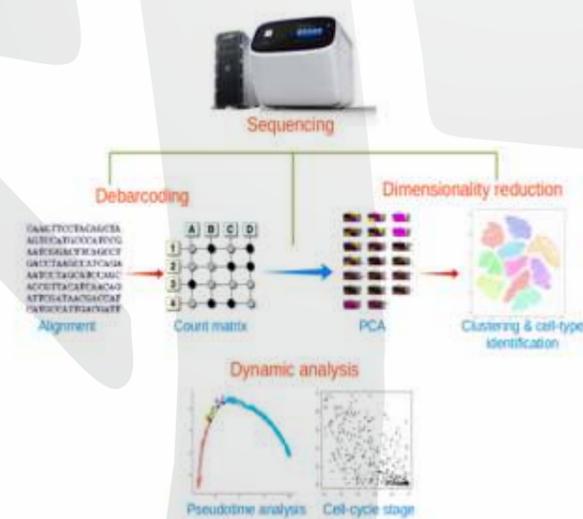
For the test case data, samples represent a time series data of induced cell differentiation which can reveal the gene expression patterns underlying differentiation of these cells over time so that the point at which cells truly begin to differentiate into distinct populations can be identified.

Two main branches specific to each cell-type (XEN and NEU) are observed.

The branching seems to happen between 48h-60h, and it reflects the commitment of cells to a particular lineage (colour range).

Additional small branched clusters could suggest the existence of additional subtypes or early commitment of cells to a particular lineage, requiring further investigation.

*Due to dynamic nature of analytical developments and lack of a well-established pipeline, scRNA-seq analyses demand advanced bioinformatics support. Using this test case analysis, we have highlighted our ability to handle scRNA-seq analyses, taking raw data and producing robust results that are ready to publish. We are constantly exploring novel innovations in this field in order to ensure that our clients receive cutting edge Bioinformatics services.*



**A network of domain experts in Bioinformatics to provide high quality solutions**

### Services Offered

- Accreditation compliant Clinical Genomics pipeline
- Single Cell Transcriptome Data analysis
- NGS data (Exome, Transcriptome, Epigenome) analysis
- Bioinformatics Tools and Database developments

Clinical Genomics Specialist

Big Data analytics & Data Modelling specialist

Applied statistics specialist

NGS analysis, Tool/Database development specialist

LifeBytes, Bangalore

Machine learning specialist

NGS data & Single-Cell transcriptome specialist

LifeBytes Head office, Brisbane

Artificial Intelligence & Proteomics specialist