Motor Neuron iPSC time-course scRNA-seq

Statement of Work

Jenny L. Smith, Computational Biologist, sr.

Research Scientific Computing,

Seattle Children’s

2023-05-04

# Background

Induced pluripotent stem cell technology is used to generate patient-specific stem cells that can undergo directed-differentiation to multiple lineages in culture. Diseases being explored include Duchenne muscular dystrophy, X-linked myotubular myopathy and autistic syndrome disorder. The iPSC model is employed in the current investigation into the physiologic effect and mechanism of astrocyte-mediated neuromodulation in normal and ALS mutant condition. This would identify molecular drivers of the observed phenotypic change using human iPSC-derived spinal astrocytes and motor neurons.

We ran a longitudinal single-cell RNAseq with the differentiation motor neurons from iPSCs and got the gene expression table using Cell Ranger (10x Genomics) of each timepoint. We collected cells from undifferentiated iPSC, day3 of differentiation, day5, day8… until day 45. We aim to identify the regional identity and subtypes of iPSC motor neurons and transcriptomic (dis)similarity with native MNs in the fetal spinal cord.

# Data provided

* Counts matrices for each time-point scRNA-seq
  + Genetically normal hIPSC cells
  + Cellranger pipeline
  + Single replicate 10x sample per time-point
  + Later time-points have scRNA-seq and microscopy to validate state of cellular differentiation
* Sample metadata sheet

# Scope

We will assist with computational analysis including QC, normalization, and cell-type annotation. Down-stream analyses for trajectory inference and time-series analyses will be completed.

This work will be supported by hourly charging of $170/hr. All scripts, figures, and data will be shared using a Miscrosoft Azure container. Estimate about **8-10 months** to completion, where we consider completion the agreed-upon final publication ready figures and methods section paragraph(s).

To help facilitate communication and collaboration on research findings, we will set-up a recurring meeting every 3 weeks to review the data analyses and results. Frequent email communication is always encouraged, and additional meetings can be scheduled as needed for questions and clarifications regarding the results/analysis. We will also create a google document to curate a set of running notes for continued tracking of progress and next steps.

Based on the scope of this work, this collaboration project would fall in the **moderate service level**, which is 10% effort for minimum of 6 months, and then extended to 8-12 months if needed. The 10% effort translates to 4 hrs per week (2 business days per month).

# Goals/Deliverables

* Merging individual scRNA-seq samples (time-points) into a single object. Check for batch effects, doublets, ambient DNA fraction, and other QC metrics.
  + Use cellranger raw counts –> doublet detection (scDblFinder, simulation) –> soupX ambient dna detection –> Seurat object
  + Gene quantification normalization and data dimension reduction (PCA, UMAP)
  + Seurat object will be shared to UW to allow for con-current transcriptome analyses
  + **Time Estimate**:
    - I’d estimate it would take 3 business days to complete: *6 weeks*.
    - If batch correction is required, and if time-points need to be excluded based on the QC checks, this would take some additional time: expect *10 weeks* to final object.
* Cell-type annotation using public references
  + Seurat cell-type annotation
  + Integrate > 1 public scRNA-seq reference dataset as reference
  + **Time Estimate**:
    - Since a custom reference will need to be generated to include the expected motor neuron cell types from spinal cord, this will take some time to integrate, and QC check the reference set.
    - Reference Set: I’d estimate it would take at minimum *4 business days* to identify, process, and integrate the reference data-set.
    - Query Set: I’d estimate it would take *2 business days* to complete the reference cell-type annotation on the iPSC model scRNA-seq dataset and investigate cell-type frequencies and quality of the reference mapping.
    - Total: *12 weeks* to have an annotated scRNA-seq dataset.
* Track the gene expression changes of the same type of cells along the differentiation. Are there cell pops that disappear and appear over time?
  + Utilize [Monocle](http://cole-trapnell-lab.github.io/monocle-release/) for pseudotime trajectory analysis and assist with troubleshooting
  + Utilize [psupertime](https://academic.oup.com/bioinformatics/article/38/Supplement_1/i290/6617492) for time-series analysis of gene-expression
  + **Time Estimate**:
    - Monocle will be run first, and has known troubleshooting issues. It will also require a data object type conversion and may require re-normalization of the data. Estimate *5 business days* to complete along with QC.
    - Psupertime requires data object type conversion, and may have troubleshooting time. Estimate *5 business days* to complete along with QC.
    - Total: *20 weeks* to final time-course and trajectory analysis.
* Assist with figure generation
  + Examples are provided in this [publication](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8353162/).

# Caveats and Unknowns

* Need to find a better in-vivo reference for spinal cord tissue, since the [Allen Institute motor cortex](https://azimuth.hubmapconsortium.org/references/#Human%20-%20Motor%20Cortex) dataset is not fully sufficient for the cell types expected to be present in the iPSC model. There is the possibility of using the human data from the [Spinal Cord Atlas](http://spinalcordatlas.org/more_info.html).
* A technical issue with this dataset (in singlets) is that the gain/loss of the cell pops overtime may be difficult to track due to biological variability between wells.
* What is our expected number of cell-pops per well? What is the general variability in induced pluripotent cells?

# Cost Estimate

* Expecting a total of **38 weeks** (9.5 months) of effort with 4 hours per week at $170/hr would cost $25,840.00 over the life of the project.
* Billed at **~$1,360.00 bi-weekly**.

# Collaborators

* David Mack, UW ISCRM Principal Investigator
* Changho Chun, UW ISCRM phD candidate
* Lisa Maves, SCRI Principal Investigator