

# Spatiotemporal properties of NDD genes during brain development for biomarker discovery

Urwah Nawaz

February 25, 2020

- section 1: brief overview of expression studies used in the brain
- section 2: bulk to scRNA-seq -> discuss some important findings of bulk RNA-seq and discuss limitations of bulk RNA-seq
- section 3: single-cell studies and explanation of technologies used on the human brain + also discuss with NDDs -> end with some downsides of using scRNA-seq and its limitations for using on postmortem brains
- section 4: deconvolution techniques to be able to integrate the bulk data
- 

## 1 Studying the brain using expression studies

The brain is functionally organized into regions, which are distinguished by distinct compositions of molecularly defined cell types and region-dependent patterns. These anatomical, and functional differences are reflected in the brain transcriptome. Furthermore, the brain undergoes protracted periods of development, refinement, and maturation, spanning the early

fetal periods to adolescence. During these windows of development any alternation in gene expression can lead to neurodevelopmental dysfunction in the brain. To link gene expression to function, several RNA-seq studies have analysed brain samples from distinct stages of development to understand the complexity of the brain. In this section of the review, we will discuss the application of using expression datasets to study functions of the brain. We will also discuss how in recent years rna-seq studies have been leveraged by single-cell techniques and discuss all publicly available resources

## **1.1 From bulk-expression to single-cell**

Over the past few years, RNASeq has gained increasing popularity and supplanted microarrays as the primary highthroughput choice for quantifying the whole transcriptome. RNASeq, based based on nextgeneration sequencing, directly determined the cDNA sequence. It effectively surmounts the limitations of microarray: a priori sequence knowledge is not required; different layers including exon, isoform and genelevel quantification can be performed; and large dynamic range can be obtained. As the human brain exhibits a more diverse splicing program than other tissues. RNASeq technology has become an ideal tool to interrogate the role of splicing events as well as transcriptomic changes in the development of neurological disorders.

Format: around 5-7 gene studies using bulk RNA-seq and then limitations of bulk rna-seq approach –> discuss single-cell rna-seq

## **1.2 Single-cell RNA-seq analysis on brain tissues**

Single-cell RNA-seq gained popularity after Tang et al demonstrated its feasibility to characterize expression profiles of individual cells using NGS technologies. Since this publication, a great interest was developed in utilising sc-RNAseq for transcriptomic analyses due to its identify rare cell populations which could have not been detected from bulk RNA-seq analysis.

Several techniques have been developed over the years to study the human brain.

One of the first scRNA-seq analyses on the human brain was performed by Darmanis et al (2015). Darmanis et al used Fluidigm C1 chip protocol to profile the expression of over 400 cells from the adult human cortical brain samples and developing brain samples of 16-18 weeks pc embryos. The study was able to identify inhibitory and excitatory neurons as well as major glial subtypes. Additionally, it was also able to infer expression differences between pre- and postnatal brains.

Several other studies also used the Fluidigm C1 chip protocol to study the brain. Nowakoski et al (2017) profiled cells from the developing human cerebral cortex and leveraged gene expression analysis using weighted gene co-expression networks. Similarly, Polioudakis et al (2019) also profiled roughly 40,00 cells from the developing brain to create a high resolution atlas of the developing human cortex.

While several different technologies exist for single-cell resolution capture of tissues, these studies were limited to fetal development. Additionally single-cell

Similarly, Krishnaswami et al. (2016) established one of the first protocols for snRNA-seq on human post-mortem brain tissue using fluorescence assisted nuclei sorting (FANS) to place individual neuronal nuclei, selected based on NeuN expression, into the wells of a microplate and used a SMART-seq approach for generating libraries. This work demonstrated the feasibility of snRNA-seq in archived post-mortem tissues, such as those accessible through brain banks.

Lake et al. (2016) performed Fluidigm C1 chip capture of NeuN positive FAN sorted neuronal nuclei to profile more than 3000 neurons from various cortical regions of a single healthy subject. They identified numerous inhibitory and excitatory neuronal subtypes which were in broad agreement with the cell-types described by Darmanis et al. (2015), but revealed finer subtypes powered by the larger dataset, such as layer-specific and region-specific excitatory neuron subtypes.

Habib et al. (2017), created an snRNA-seq method they termed DroNc-seq and applied it to several post-mortem human prefrontal cortex (PFC) and hippocampus samples, in addition to mouse brain tissue. DroNc-seq incorporated several adjustments to the Drop-seq [6] protocol, including an alteration to the dimensions of the microfluidic device to allow for capture of nuclei, which are smaller than cells, and inclusion of intronic reads in analyses due to the preponderance of pre-mRNA in the nucleus. They were the first to demonstrate the feasibility of droplet-based high-throughput snRNA-seq in archived post-mortem human brain tissue [8]. Furthermore, there was good correspondence in cell-types between the mouse and human datasets and with the findings of Lake et al. (2016).

Lake et al. (2018) independently designed an adaptation of Drop-seq for snRNA-seq. Their modifications included heat-based lysis of nuclei and incorporation of intronic reads, similar to Habib et al. (2017). Moreover, they performed an assay for single-nucleus chromatin accessibility based on combinatorial barcoding, in addition to snRNA-seq, and used their snRNA-seq findings to refine clustering of single-nuclei based on chromatin accessibility [47]. Although tissue was obtained from healthy subjects, the cell-type specific chromatin accessibility information generated was used to indirectly assess cell-type involvement in neurological and psychiatric diseases.

While most studies have focused on the cortex, Welch et al. (2019) performed high-throughput snRNA-seq on more than 40,000 nuclei derived from archived substantia nigra samples from 7 healthy donors. They developed and applied a single-cell data analysis tool called LIGER for aligning the data from multiple individuals into a consolidated dataset. Clustering driven by inter-individual variability is a recurring problem in snRNA-seq datasets, and LIGER was able to mitigate this effect. They identified the expected subtypes of glial and neuronal cells, including dopaminergic neurons. Moreover, they were able to pin-point subject specific effects: including activation of microglia in one subject who experienced traumatic brain injury (TBI) at the time of death, and distinct microglial and astrocytic signatures in another subject with histological signs of amyloid deposits upon post-mortem examination

[48]. Thus, not only will their dataset serve as a reference for future snRNA-seq studies of the substantia nigra, but their software, which can effectively combine results from multiple datasets for joint analysis without losing dataset-specific components of the information, will be widely applicable in future single-cell sequencing studies.

While the massive capacity of high-throughput snRNA-seq is enticing, some questions require a more targeted approach as exemplified by a recent human brain snRNA-seq study from the Allen Institute [49]. Boldog et al., (2018) identified a new subtype of inhibitory neuron, dubbed the rosehip neuron, which seems to be uniquely found in the human cortex. The information from snRNA-seq was complemented by morphological and electrophysiological data from surgical tissue as well as corroborated with fluorescent in situ hybridization (ISH). The data from this study is part of a larger human mid-temporal gyrus dataset [50], generated by the Allen Institute from both post-mortem samples and surgical tissue and it provides an excellent resource for benchmarking data produced by high-throughput platforms. Uniquely, this dataset accounted for cortical layer location during the dissection and extraction of nuclei.

### **1.3 RNA-seq methods including deconvolution of bulk-expression**

### **1.4 Large scale transcriptomic resources**

## **2 Current data availability**

Table 1: Single-cell RNA-sequencing studies of the human brain

| <b>Authors</b>   | <b>Year</b> | <b>Tissues</b> | <b>Method</b> | <b>No. of Cells/Nuclei</b> | <b>Age</b> | <b>No. of individuals</b> | <b>Data availability</b> |
|------------------|-------------|----------------|---------------|----------------------------|------------|---------------------------|--------------------------|
| Nowakowski et al | 2017        | V1, PFC        | scRNA-seq     |                            |            |                           |                          |

Table 2: Single-nucleus RNA-sequencing studies of the human brain

| Authors | Year | Tissues | Method | No. of Cells/Nuclei | Age | No. of individuals | Data availability |
|---------|------|---------|--------|---------------------|-----|--------------------|-------------------|
|---------|------|---------|--------|---------------------|-----|--------------------|-------------------|