Mapping Neural Diversity: A Molecular Analysis of Neuronal Identity in Mouse Primary Visual Cortex

Single-cell mRNA-Seq with the MiSeq™ and HiSeq™ Systems is enabling Allen Institute researchers to classify V1 neural cells.

Introduction

Michio Kaku, American futurist, physicist, and author, wrote, "the human brain has 100 billion neurons, each neuron connected to 10,000 other neurons. Sitting on your shoulders is the most complicated object in the known universe." His words are no exaggeration. In fact, to date, scientists can't even tell us how many different types of neurons we might find in a relatively simple mammalian brain, like that of a mouse, let alone a human one.

The Allen Institute for Brain Science, an independent, nonprofit research organization dedicated to understanding the inner workings of the brain, is trying to change that. Bosiljka Tasic, PhD, a researcher in the Cell Types program at the Allen Institute is developing pioneering molecular analytic methods, analyzing transcriptomes, and epigenetic landscapes of individual neurons to define neuronal identity within the mouse visual system. She spoke with iCommunity about the Allen Institute's mission and how single-cell sequencing with Illumina sequencing systems is transforming neuronal classification efforts.

Q: What is the mission of the Allen Institute for Brain Science?
Bosiljka Tasic (BT): The mission of the Allen Institute is to accelerate the understanding of how human brain works. To make an impact towards this ambitious mission, we are studying how information is coded and processed within the wide variety of cells that constitute the mouse brain. Mouse is the most accessible mammalian model system, and many of the principles of information processing are conserved among mammals. The plan is to define the basic information-processing framework by studying the mouse visual system, and then compare it with primates and humans to understand what is uniquely human.

To accomplish this, we have an organizational structure and scale that differs from standard academic labs. We bring together many scientists with different expertise and organize them in overlapping teams in a truly multidisciplinary fashion. We also provide the primary data from these studies on our website as encyclopedic resources to the community. One of our newest resources focuses on cell types in the primary mouse visual cortex (known as V1 or VISp), including gene expression data, physiology, and morphology of individual neurons that constitute this brain area.

Q: What types of cells make up V1?
BT: V1 is the main region in the cortex that processes visual information. We had limited understanding of the extent of cell diversity within this cortical area. We knew the major types: excitatory and inhibitory neurons, and non-neuronal cells, and some further subdivisions within each of these major categories. While the cortex has been studied by many people, a comprehensive description of cellular diversity and the correlation of different types of information at the single-cell level did not exist. Most studies were usually based on several parameters and not on a highly multidimensional data set. For example, if you use three genes to classify cells, you might get one picture. However, if you look at all the genes, you might obtain a very different view.

Q: Why is single-cell transcriptome sequencing a valuable tool for these studies?
BT: There are several reasons why single-cell transcriptome sequencing is advantageous. It's a genome-wide technique, meaning you analyze to the best approximation all the genes that are expressed in a cell. It gives you a view of cellular diversity based on many genes. We can detect thousands of genes per

Bosiljka Tasic, PhD is a researcher in the Cell Types program at the Allen Institute in Seattle, Washington.

cell using single-cell mRNA sequencing and get a highly multidimensional data set for every cell. Before this, people would characterize cells based on a couple of genes at most. If you happened to be looking at the key genes, that might be sufficient. However, we didn't know what the important genes were.

"We can detect thousands of genes per cell using single-cell mRNA sequencing and get a highly multidimensional data set for every cell."

Q: What have you discovered about V1 neuronal cells using single-cell mRNA-Seq?

BT: We didn't know when we started our single-cell mRNA-Seq study what we were going to see. We collected close to 1700 cells from the V1 of transgenic mice that passed our quality control criteria with good sequencing information. We were able to define 49 types in the V1 cortical area, with 42 being neuronal types, and 7 being non-neuronal types.[1] Out of the 42 neuronal types, we found that about half are GABAnergic neurons, while the other half are glutaminergic neurons.

What's nice about this cell classification analysis is that it is unbiased. When we performed clustering and decomposition of this single-cell data set into groups, we were blind to where the cells came from in the V1 (which cell layer and which transgenic line). We performed two parallel clustering approaches and then a third validation layer of analysis to determine how robustly we could classify each single cell into a type given the gene expression signatures we uncovered. We assigned excitatory, inhibitory, or a non-neuronal identity type post hoc based on known or new marker genes.

Q: You found some clusters that you characterized as "fuzzy". What does that mean and what are the implications?

BT: When working with our bioinformaticians, I wanted to be able to examine the gene expression signatures we obtained for our cell types, and then ask how well they matched any of those types for every cell. That examination, which uses a repetitive machine learning approach, showed that sometimes certain cells would classify into one cluster and sometimes into another. We call these "intermediate" cells and they provide us with a view that you don't normally see in research papers. These are cells that have ambiguous identities, and they are more prevalent among certain cell types. Some clusters are "connected" by many intermediate cells. Other clusters occupy a singular position in this multidimensional gene expression space and don't have any intermediate cells between them and other clusters.

These data suggest that not all cell types have a rigidly discrete identity. Some cell types might not be clearly separable, and may in fact be part of phenotypic continua. This is not a foreign concept for neuroscience, where we know that neurons can

change in response to activity or experience, and possess the plasticity needed to modify their behavior during the lifetime of the animal.

Q: What techniques did you use previously to analyze the transcriptome?

BT: We performed qRT-PCR in parallel with next-generation sequencing (NGS). However, the problem with qRT-PCR is that you don't always know what genes you want to look at. The issue with any non-genome-wide method is that you can spend a tremendous amount of time selecting genes and still not get the right answer. If you don't base your gene selection on prior genome-wide knowledge, your selection is going to be biased and won't provide a good representation of the complete transcriptomic landscape.

mRNA-Seq provides genome-wide gene expression profiles, enabling us to specifically select genes that best exemplify the divisions in the transcriptomic landscape. Some of the genes were known before, but many that we now use as best markers for individual neuronal types we didn't know about before identifying them with NGS. Instead, we used genes that were present in the literature—but many genes were not tested or detected because every method has its own sensitivity issues. With NGS, we identified many new, previously undiscovered markers whose expression we subsequently confirmed with alternate methods.

"With NGS, we identified many new, previously undiscovered markers whose expression we subsequently confirmed with alternate methods."

Q: How did you perform cell isolation?

BT: Cell isolation was a significant hurdle. Isolation of adult live neurons is hard because the adult nervous system tissue is not easy to dissociate into suspension. Cells are highly interconnected, and in the cell isolation process, axons and dendrites are torn apart and many cells don't survive. In addition, to access some rare cell types, we would need to profile many cells. So, we decided to use transgenic Cre lines as the cell source, where specific groups of cells are labeled with fluorescent proteins. Then we optimized our procedure for making cell suspension of adult brain and established fluorescence-activated cell sorting (FACS) for single-cell isolation. All this enabled us to isolate those rare populations, as we could sample rare cell types more frequently than we would be able to in an unbiased fashion. Therefore our cell sampling was not unbiased, but we deliberately chose and isolated cells from transgenic lines that could label potentially rare cell types. We obtained extremely reliable single-cell isolation using this approach.

Q: Which sequencing systems are you using for these studies?
BT: We performed library validation sequencing in our laboratory by relatively shallow sequencing with a MiSeq System. Then we outsourced deeper library sequencing to several local core laboratories that have HiSeq Systems.

> "Having the MiSeq System in-house enabled us to rapidly develop and change our methods, and confirm that we had good libraries before we sent them out to core labs for deeper sequencing on the HiSeq System."

Q: Why did you choose the MiSeq System?
BT: We needed a fast-turnover system in-house to perform library validation especially while we were developing our processes. Having the MiSeq System in-house enabled us to rapidly develop and change our methods, and confirm that we had good libraries before we sent them out to core labs for deeper sequencing on the HiSeq System. It was a good combination for us.

Q: What kind of library prep kits are you using?
BT: We tested several approaches, and selected Clontech's SMARTer as it could reliably amplify samples from single cells. For our V1 study, we used SMARTer Version 1. SMART-Seq 4 is now available and we're using it for our new studies. For library prep from cDNAs obtained by SMARTer, we used the Nextera™ XT Library Prep Kit. It allowed us to use small amounts of cDNA for NGS library preparation and bypass sonication.

> "Until we had single-cell transcriptomic analysis, we didn't have a way of taking a heterogeneous tissue and defining the molecular types within it in an unbiased manner."

Q: What were the parameters of your sequencing runs?
BT: We initially went overboard on sequencing depth, because we didn't know what depth was required to obtain good resolution for distinguishing new cell types. We sequenced cDNA from our early single-cell samples to about 20–30 million reads, sometimes even higher. We then performed read subsampling and clustering *in silico*, and decided that, for most cells, obtaining 5–10 million total reads per cell was sufficient.

This kind of depth is not necessary if one wants to distinguish neurons from non-neurons and excitatory from inhibitory cells. For those studies, one can use much lower depths—100,000 reads per cell or less is sufficient. However, if we want to distinguish related neuronal subtypes, with the cell numbers we were able to obtain, we definitely needed deeper sequencing.

Q: What information has single-cell sequencing uncovered that other approaches didn't enable you to see?
BT: Single-cell sequencing has contributed to the classification of cell types, and not only in the cortex. Until we had single-cell transcriptomic analysis, we didn't have a way of taking a heterogeneous tissue and defining the molecular types within it in an unbiased manner. Single-cell transcriptomics allows you to decompose tissue into types without first asking "What genes do I need to use?" Instead, we can look at all the genes.
Before single-cell sequencing, we also had no idea what would constitute a reasonably comprehensive index of the different cell types. Were we talking about thousands of types? What is the order of magnitude? Our studies suggest that we're talking about 50 different V1 cell types and some of them might be fuzzy. I can't claim our work is truly comprehensive—and there are probably rare cell types we did not identify. So it is possible that in the end, there will be dozens more within this brain area.

Finally, single-cell sequencing has allowed us to define markers that are specific for particular cell types. That has immense implications for building tools that can access those specific types. Now we have a recipe; for example: Gene A plus Gene B will give us specific access to a cell type Z. Before we were only guessing. These new tools will enable us to define the function of different cell types and investigate how they work within neural circuits.

> "Comparing single-cell mRNA-Seq data from different cortical areas will enable us to ask new questions relating to conservation and uniqueness of cell types."

Q: What's next in your research?
BT: Our study was, in a way, a pilot study. It showed us that we can perform single-cell mRNA-Seq with adult cortex cells in a well-defined anatomical region. We now want to profile other well-defined regions, especially other cortical areas. Comparing single-cell mRNA-Seq data from different cortical areas will enable us to ask new questions relating to conservation and uniqueness of cell types.

We have a few collaborations with external academic labs that are adopting our approach for classifying cells in their favorite brain regions. Using this technique, we can decompose any region of the brain that might have vastly different functional roles from V1. By building specific genetic tools, we can ask what is the function of certain cell types in the specific behaviors we're interested in studying. It's very exciting.

Learn more about the products and systems mentioned in this article:

MiSeq System, www.illumina.com/systems/miseq.html

HiSeq System, www.illumina.com/systems/hiseq_2500_1500.html

Nextera XT DNA Library Prep Kit,
www.illumina.com/products/nextera_xt_dna_library_prep_kit.html

References

1. Tasic B, Menon V, Nguyen TN, *et al*. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci*. 2016; 19(2): 335–346.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.