# Lesson 2: The Importance of Basic Research in Brain Science

Learning Objectives:

- Students will be able to articulate what transcriptomic data is and how it is gathered
- Students will be able to articulate how dendrograms and heatmaps can be used in conjunction with transcriptomic data in order to further differentiate cell types from one another
- Students will be able to apply basic principles of interpreting data visualization to complex transcriptome datasets
- Students will explore the nuanced differences between basic and applied research
- Students will be able to defend the importance of both basic and applied research within the field of biomedical science
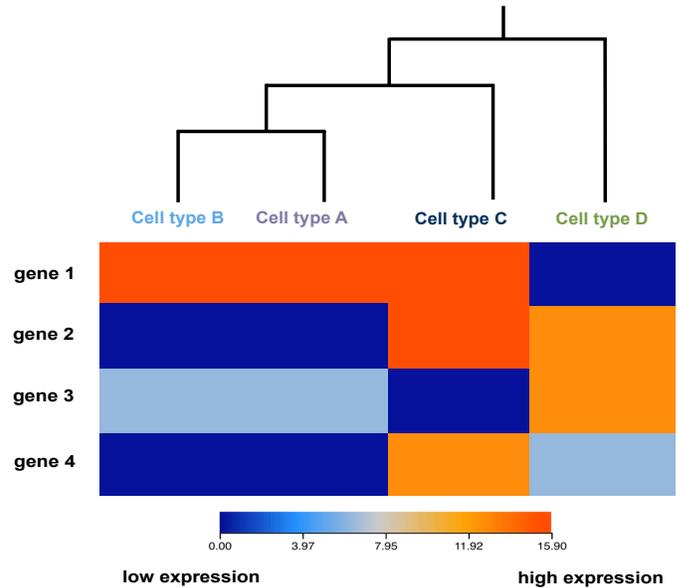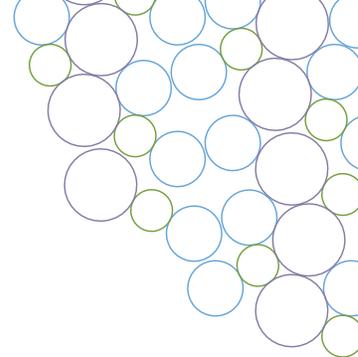
Introduction:

If you completed Lesson 1, you learned about the process through which an individual can donate their brain to science. Keeping what you learned about brain donation in mind, you will now explore what type of basic research underlies the field of neuroscience.

The Allen Institute, located in Seattle, Washington, is a nonprofit scientific research center that specifically prioritizes basic research. **Basic research**, also known as **foundational research**, is curiosity-driven research that aims to further our understanding of a particular phenomenon. For example, studying the healthy human brain is one type of basic research, where scientists further our understanding of the brain itself, rather than focusing on treating disease states or other concrete applications. A "healthy" brain is a term that refers strictly to a brain that is not known to have a neurological disease of some kind. "Healthy" is not synonymous with "normal." In fact, given the wide array of neurodiversity that exists within the population, "normal" is not a term used in science to describe brains. For more information about neurodiversity in brain science, see *Lesson 1: Brain Donation and Bioethics*.

While studying a healthy human brain is considered a type of basic research, the field of neuroscience also conducts a significant amount of applied research. Applied research specifically seeks to understand options for treatment and/or a cure for a particular disease. Both basic and applied research are necessary in order to further our scientific understanding and improve human health. Basic science helps provide a foundation of knowledge for later translational research and is crucial for the field of biomedical science as a whole. Studying a healthy, neurotypical brain allows us to understand what may change later during the onset of disease or the development of neuroatypical characteristics. In addition to these practical applications, basic research is also helpful for gaining a foundational understanding of the healthy human brain to expand our general knowledge of the field of neuroscience.

While both basic and applied research are integral to the field, science is often limited by time, money, and resources. **Given these restrictions, how do scientists decide whether to conduct basic or applied research?**
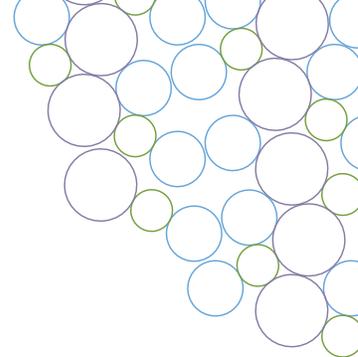
ALLEN
INSTITUTE

In this lesson, we will use current research from the Allen Institute for Brain Science as a specific case study in the process of conducting basic scientific research. Founded in 2003 by Paul G. Allen, the Allen Institute for Brain Science is a division of the larger Allen Institute.

The Allen Institute's mission is to conduct big, open, and team science. **Big science** means conducting large-scale, basic research projects that aim to uncover the mysteries of biological systems such as the brain, the immune system, and the cell. **Open science** means that much of the data generated by the Allen Institute is made openly available to the public. Lastly, **team science** means that each of the Allen Institute's projects relies heavily on scientific collaboration and large teams of people, versus a more traditional lab model where a Principal Investigator (or PI) leads a group of people who may not have the chance to collaborate with other labs.

In this lesson, you will have the chance to explore some open data available on the Allen Institute for Brain Science's website. In order to further understand the importance of basic scientific research, you will have the chance to interpret the transcriptomic data of the healthy human brain. By looking at dendrograms and heatmaps, you will strengthen your data analytic skills while also exploring why it is so important to further our understanding of the healthy human brain and its composite parts!

Activity 1: What would you fund?

Before we dive into looking at the actual Allen Institute for Brain Science's data, we want to do a quick exercise that prompts you to consider what type of research, you, yourself, would choose to fund. The type of research scientists can pursue is often restricted by limited resources and opportunities for funding. Given the limited availability of funding, scientists must communicate why their research is important and the impact it would have on society. This can pose a challenge for scientists pursuing basic research projects, since funding agencies typically are looking for research that will have specific implications for disease. This activity will provide you with the opportunity to explore how scientists in both applied and basic research can advocate for the importance of their research.

Scenario: You are given 10 million dollars to invest. Two teams of scientists approach you with a grant proposal. The grant proposals are written documents that outline these groups of scientists' plan for what they will study, how they will study it, and WHY you should choose to invest your money to help them study it.
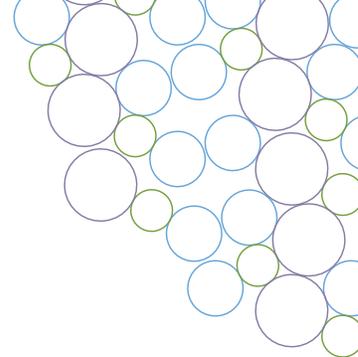
Proposal 1: The scientists in group 1 plan to study a possible treatment for Alzheimer's disease (AD). AD is a neurological disease that results in a decline in memory skills and cognitive functioning and is estimated to impact 24 million people globally. They are interested in testing whether a drug that is thought to break down phosphorylated tau protein works to treat AD. Scientists hypothesize that phosphorylated tau protein form neurofibrillary "tangles" that accumulate with neurons and disrupt normal functioning.

Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3405821/#:~:text=In%20the%20US%2C%20approximately%205.5,as%20high%20as%2024%20million

Proposal 2: The scientists in group 2 plan to study the varying cell types of the human brain. These scientists define cells as different "types" when they have different structure/forms (morphologies), gene expression, and electrophysiological activity. These scientists will study brain tissue donated from healthy, neurotypical individuals who had no known neurological diseases or conditions. The output of this research will be descriptions and classifications of different cell types within the human brain. This research hopes to further science's foundational knowledge of the structure and function of the human brain.

Activity: Given these descriptions of the two proposed projects, fill out the table below.

	Potential benefits	Potential limitations/drawbacks
Proposal 1: Researching a treatment for Alzheimer's Disease		
Proposal 2: Researching the cell types of the human brain		

ALLEN
INSTITUTE

Now it's time for you to make your investment decision. How do you plan on allocating your money? Is one group getting all 10 million dollars, are you splitting it evenly, or are you distributing it in a particular manner between the two groups? Be as specific as you can in explaining how you will allocate your money and WHY you made the decision you did: (1 paragraph)

This activity was meant to encourage you to think about the benefits and/or drawbacks of both basic and applied research. There is no one right answer as to how money should be allocated between the two, as they both have their own unique role in the field of science.
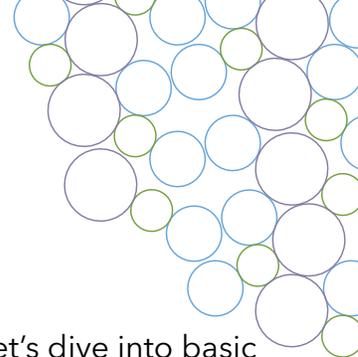
Basic Research in Practice: Studying a Healthy Human Brain

Now that you have had the chance to explore the importance of basic research, let's dive into basic research in practice by looking at some of the open data made available from the Allen Institute for Brain Science. In particular, we will be looking at the data the Allen Institute for Brain Science has collected from the **Middle Temporal Gyrus (MTG)** of five healthy, neurotypical donor brains. While the Institute has collected a diverse range of data from each of these donor brains, this lesson will specifically be looking at the **transcriptomic data** collected from the brain cells in each donor's MTG.

In order to understand what transcriptomic data is, let's first establish the difference between a **genome** and a **transcriptome**.

Genome vs. Transcriptome:

When you hear the term "genome," what do you think of?

A genome is a complete library of the genes/genetic material present within somatic cells. Somatic cells are the cells within the body that are not sperm or egg cells. The National Human Genome Research Institute defines a genome as "a fancy word for all your DNA."

Think about your own genome and the somatic cells within your body. Somatic cells could be anything from the cells that make up your skin to the cells that make up your brain.

Knowledge Check

- **Do your skin cells carry a different genome (set of genes) than your brain cells? In other words, do skin cells have different genes than brain cells?**

- **But if skin cells and brain cells contain the same genes, what makes one a skin cell and the other a brain cell?**

ALLEN INSTITUTE

Even though your skin cells and brain cells contain the same genes, this does not mean that they **express** all of the same genes.

How can we measure gene expression?

In order to answer this question, scientists gather **transcriptomic data**.

What is transcriptomic data?

One method scientists use to differentiate cell types from one another is to collect transcriptomic data. This transcriptomic data identifies **which genes a cell is transcribing into RNA transcripts** and in **what quantities**. If a cell, and more specifically, that cell's nucleus, contains a specific RNA transcript, this indicates that the cell is expressing the specific gene associated with that RNA. By (1) isolating nuclei, (2) sequencing the mRNA transcripts found within the nuclei, and (3) counting those transcripts, we can tell **which genes** the cell is expressing and **how much** these cells are expressing these genes. The figure below provides a quick reminder of what transcription is within the context of the central dogma of biology:
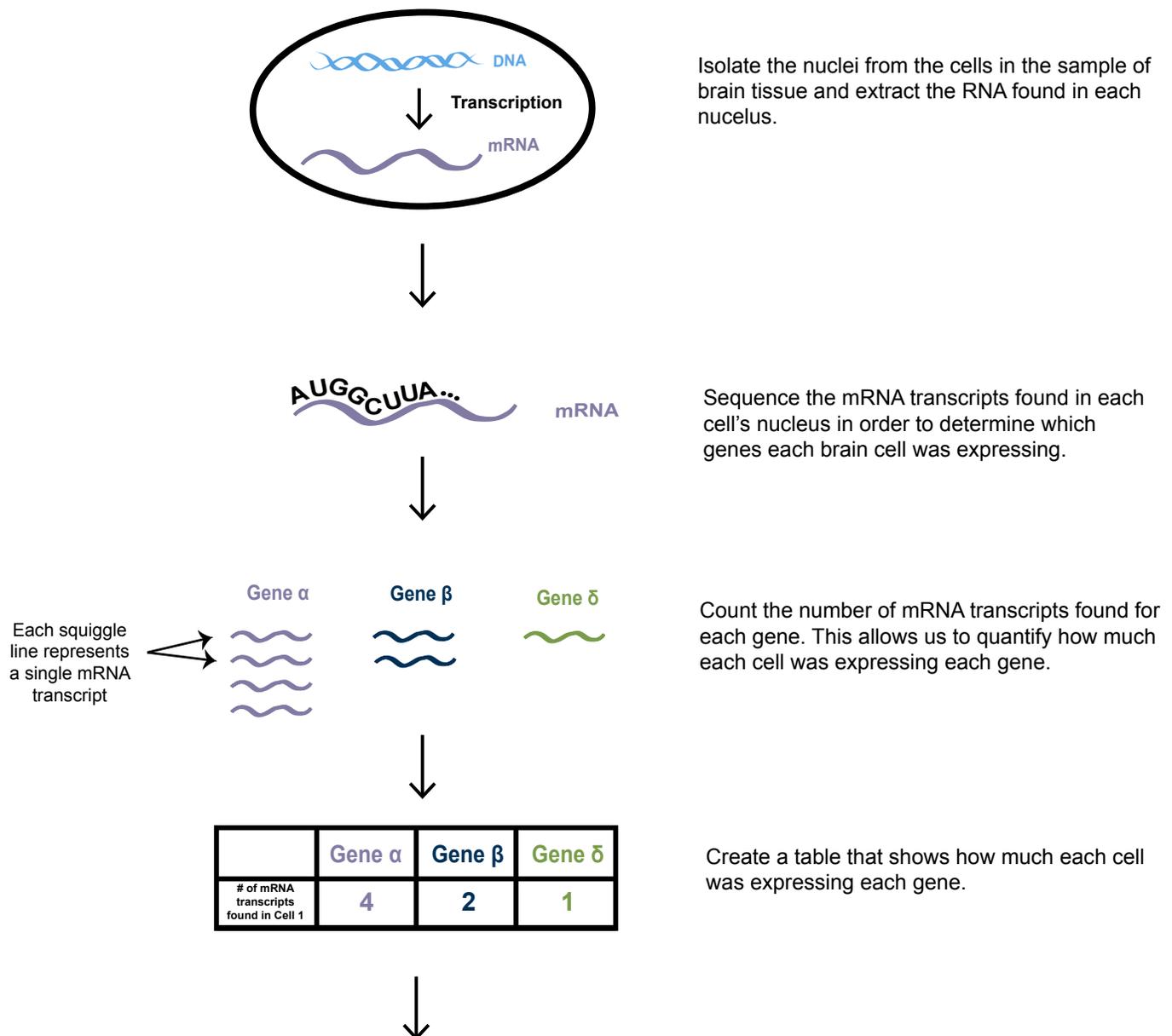
Central dogma
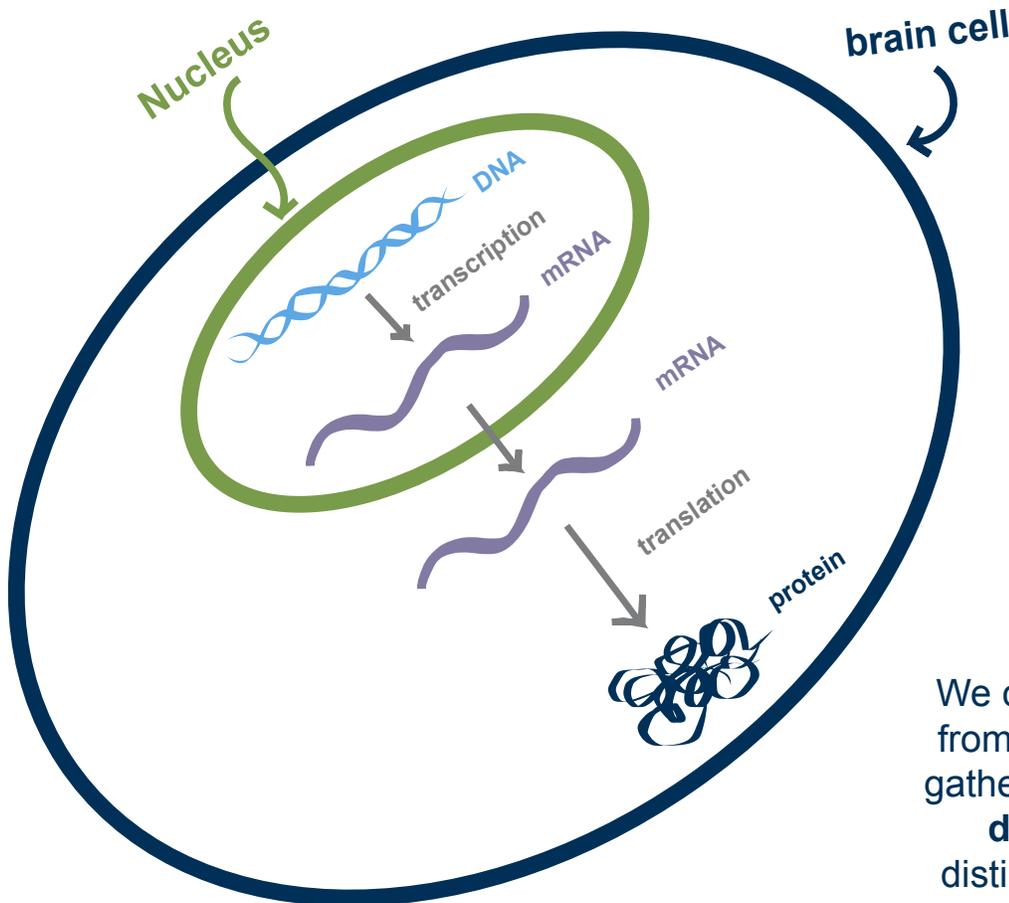
We can isolate nuclei from brain tissue and gather **transcriptomic data** to help us distinguish cell types from one another
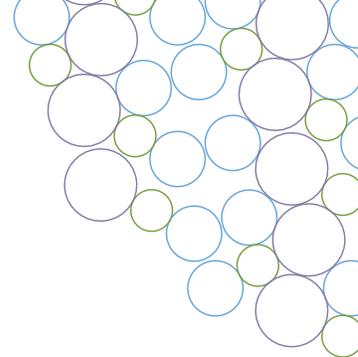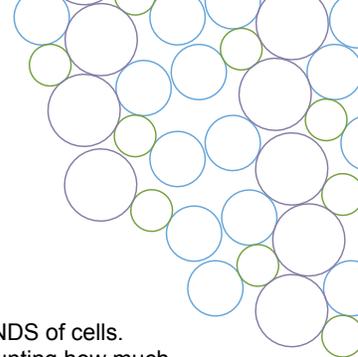
Transcriptomic data is growing increasingly popular within the field of neuroscience. Before you can analyze transcriptomic data, it is important for you to understand both what it is and how it is gathered. The figure below details how scientists gather transcriptomic data and use it to study different types of cells, also called "cell types," within the brain.

Transcriptomic data is a key tool used to construct a **transcriptome** for a particular cell. What is the difference between a genome and a transcriptome? While a genome is a complete catalog of all genes available for every cell in the body, a transcriptome is a catalog of the RNA transcripts found within a cell at a specific point in time. A transcriptome is often used as a method of measuring gene expression within a cell.

Transcriptomic data is gathered by following the steps outlined in the figure below:

Isolate the nuclei from the cells in the sample of brain tissue and extract the RNA found in each nucelus.

Sequence the mRNA transcripts found in each cell's nucleus in order to determine which genes each brain cell was expressing.

Count the number of mRNA transcripts found for each gene. This allows us to quantify how much each cell was expressing each gene.

Each squiggle line represents a single mRNA transcript

	Gene α	Gene β	Gene δ
# of mRNA transcripts found in Cell 1	4	2	1

Create a table that shows how much each cell was expressing each gene.

|	Gene α	Gene β	Gene δ
# of mRNA transcripts found in Cell 1	2	4	1
# of mRNA transcripts found in Cell 2	1	5	5
# of mRNA transcripts found in Cell 3	5	1	3
repeat count for thousands of cells...	...	...	...

Repeat this process for THOUSANDS of cells. Remember, this means we are counting how much EACH cell was expressing EACH gene. If we wanted to create a table that listed the data in full, this data table would have thousands of rows.

x = gene α value
y = gene β value
z = gene δ value

(1,5,5)
(2,4,1)
(5,1,3)
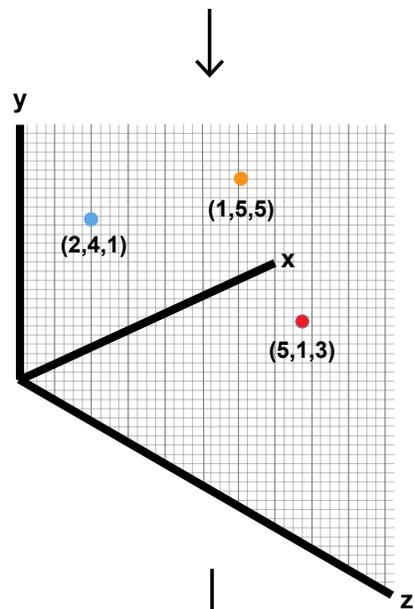
If we wanted to create a graph that plotted the inital data for cell 1, cell 2, and cell 3 and their relative amount of expression of gene alpha, gene beta, and gene delta, we would need a 3D graph like the one on the left.

x = gene α value
y = gene β value
z = gene δ value
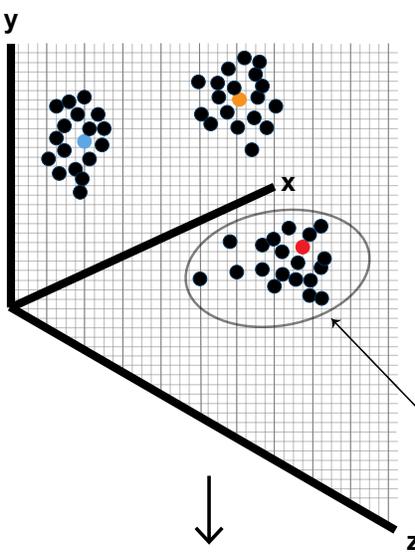
We can repeat this process for the thousands of cells that were collected from the brain tissue sample. Notice that the cells begin to cluster based on how similar their gene expression for gene alpha, gene beta, and gene delta is to one another. These clusters help us identify which cells may be more similar and/or dissimilar to one another!

when we plot the gene expression data for more cells, we notice that cell 3 (red dot) clusters next to these other cells from the sample

ALLEN INSTITUTE

9
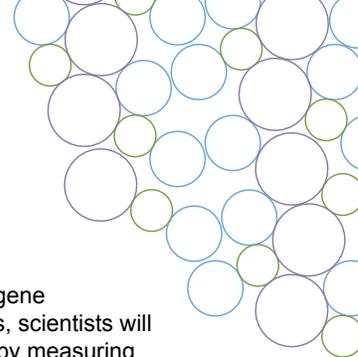
	Gene α	Gene β	Gene δ	repeat for thousands of genes...
# of mRNA transcripts found in Cell 1	2	4	1	...
# of mRNA transcripts found in Cell 2	1	5	5	...
# of mRNA transcripts found in Cell 3	5	1	3	...
repeat count for thousands of cells...	...	...	...	...

In addition to collecting data on gene expression for thousands of cells, scientists will add another layer of complexity by measuring the gene expression of these thousands of cells for THOUSANDS of genes. A table displaying this data would have thousands of rows and thousands of columns. Since the graph would now have much more than just 3 dimensions, we will need a special type of tool to graphically represent this data in a way that humans can visualize.

Each dot represents a single nucleus isolated from a single brain cell
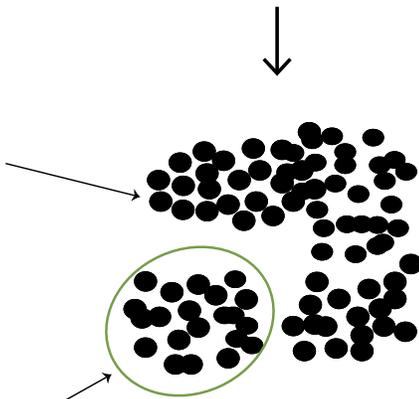
UMAP

Identify clusters in the data--these clusters represent cells that are more like each other than they are like any other cells

In order to plot this many-dimensional graph in a way humans can visualize, we use a dimensionality reduction tool, such as a UMAP, to plot it in a 2D space. Dimensionality reduction is a technique that helps represent many-dimensional data in just two or three dimensions.

Cell type B   Cell type A   Cell type C   Cell type D
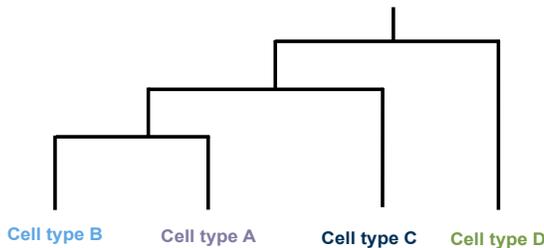
Organize the clusters identified in the UMAP to construct a dendrogram that displays hierarchical relationships between the clusters based on each cell type's similarity and dissimilarity of gene expression.

ALLEN INSTITUTE

10
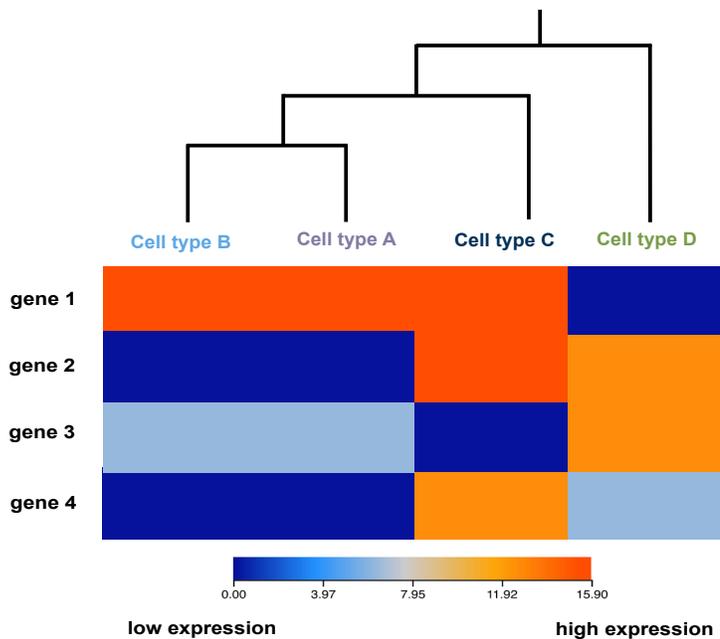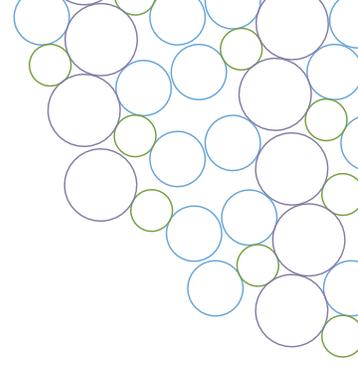
Use a heatmap below the dendrogram to compare the level of gene expression between each cell type for specific genes of interest.

As we established above, understanding **which genes a cell expresses** and in **what quantities** are two key pieces of information that help us distinguish different types of cells from one another. While transcriptomics can be used in a variety of scientific studies to understand organisms and/or cells, this lesson will focus on its applications within the field of brain science.

If two cells show a difference in which genes they express and/or if they express genes in different quantities, scientists can use this as evidence to support the hypothesis that these cells may be different **types** of cells. Organizing cells into different cell types is an essential part of understanding the brain. There are other ways to define cell types – historically, the most common was based on their shape – but in this lesson we will be working only with cell types defined based on the transcriptome. The driving motivation behind research to discover these different **cell types** is that in order to understand the whole (the brain), we have to first understand its parts.

In this lesson, you will have the opportunity to analyze transcriptomic data from healthy human brains. In this part of the lesson, we will use the Human Brain Atlas from the Allen Institute for Brain Science to explore how scientists can use gene expression to create a "map" of cell types in the healthy, neurotypical brain by identifying different "types" of cells in the brain.

Before we can look at this data, we will need to know how to read both a dendrogram and a heatmap, which are two ways of visualizing data. We will start by learning about dendrograms.

Activity 2: Dendrograms

What is a dendrogram?

A dendrogram is one type of visual aid used to show relatedness between different items of interest. It is often used to cluster items in a hierarchal manner, grouping items that are more similar to one another closer together on the "tree" and items that are more dissimilar further apart. One type of dendrogram is a "phylogeny," which is an evolutionary tree that shows relatedness between species based on common ancestry within the context of evolutionary biology.

Dendrograms can be used to categorize almost anything. For example, imagine that we wanted to organize the following foods based on how similar/dissimilar they are to one another:

- Crème brûlée
- Ice cream
- Pasta
- Popsicle

- Salad
- Cookies
- Sandwich
- Soup

Step 1: Look at the dendrogram below. This dendrogram compares these foods based on their level of similarity to one another.
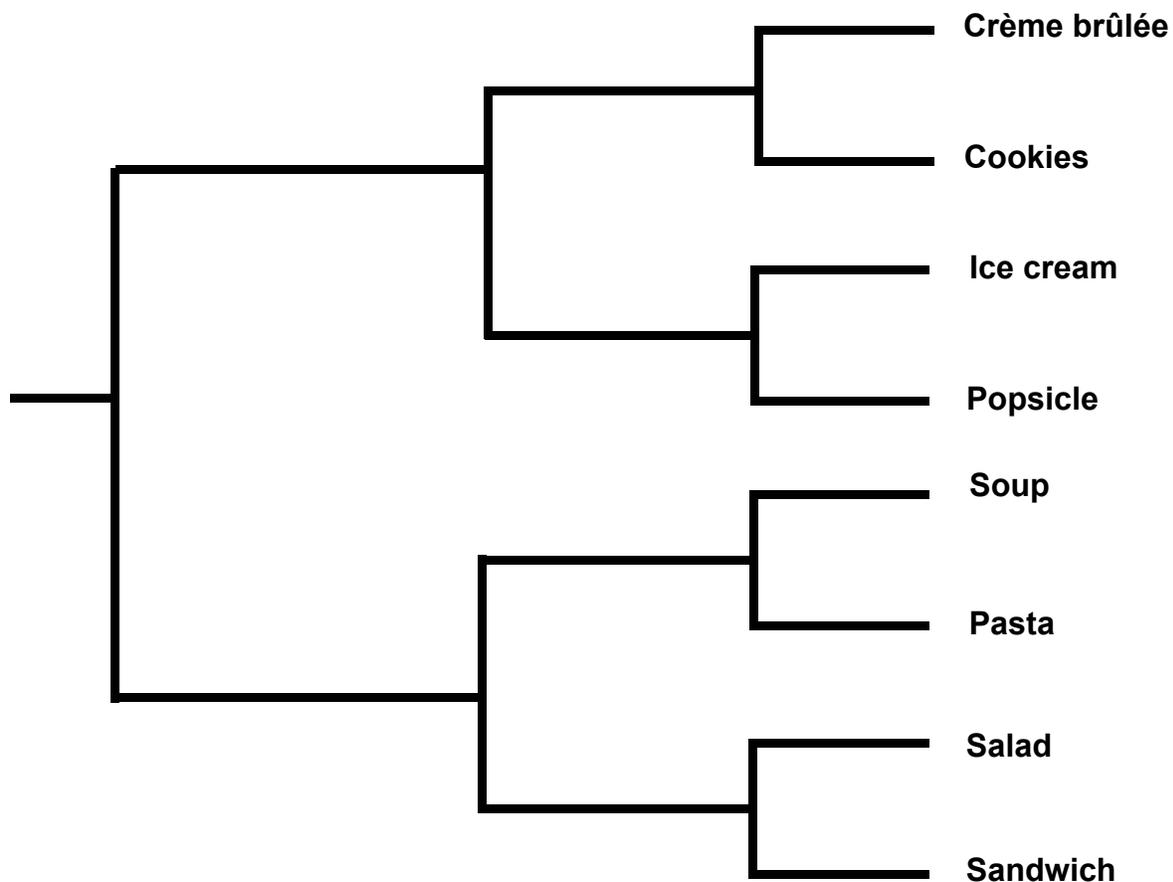
Food items that are grouped closer together on the tree are more similar to one another, while types of food that are grouped further apart are more dissimilar. This dendrogram can be enhanced by adding labels that describe what characteristics of these food items distinguish them from one another:
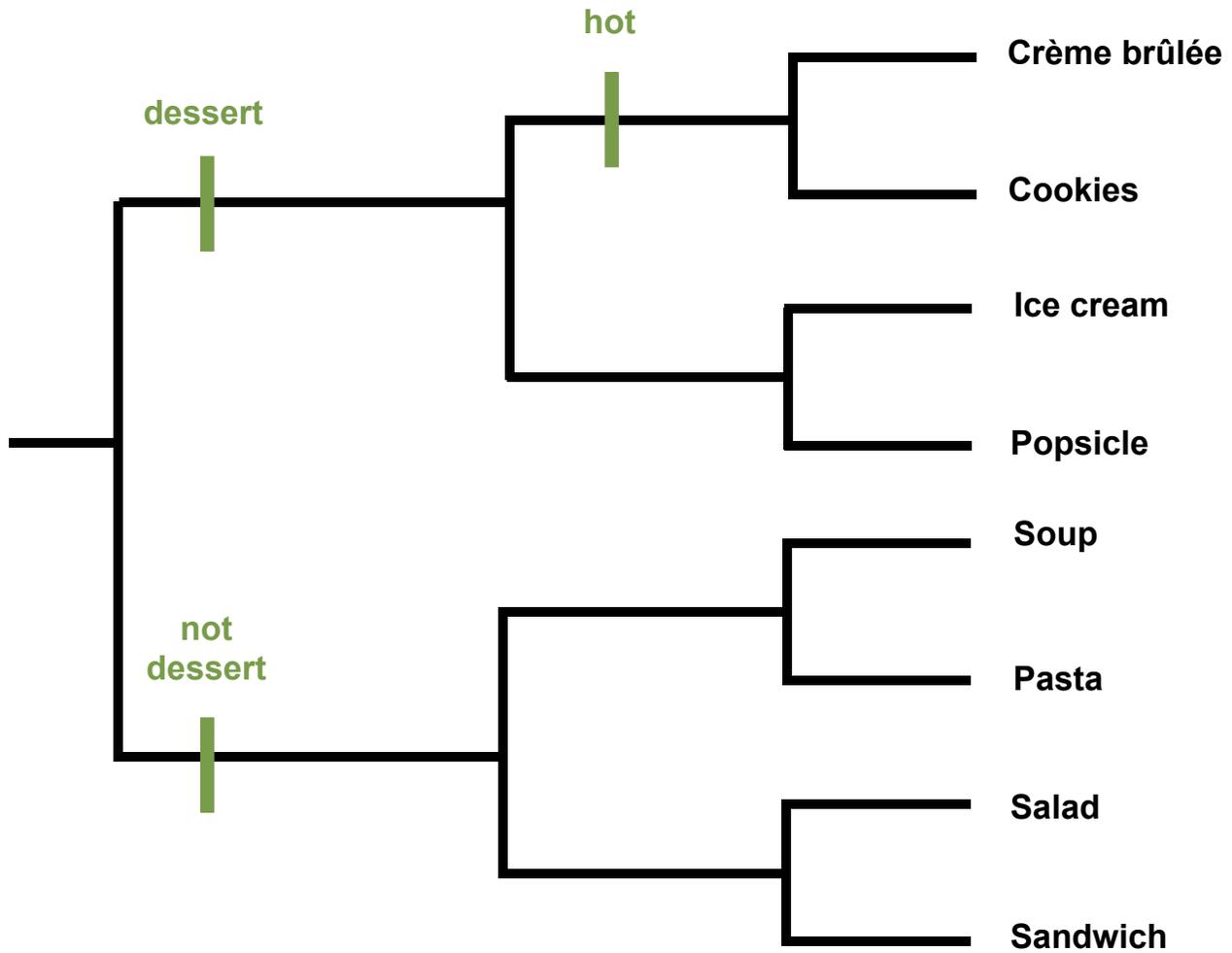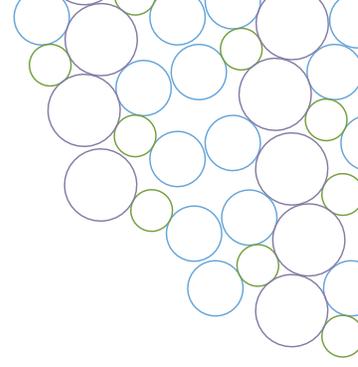

Step 2: Notice the green dashes on the dendrogram. These green dashes represent specific characteristics of the food. For example, the first green dash shows branching based on whether or not a food item is considered a dessert or not a dessert. If we follow the "dessert" branch, we see another branching event where the crème brûlée and cookies are on the branch labeled "hot," referencing the temperature at which they are usually served/eaten.

ALLEN
INSTITUTE
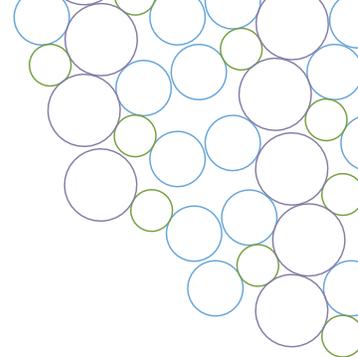
Step 3: We have already mapped the first few characteristics of the food items on the dendrogram. Now it is your turn to choose what other characteristics of the food items you can add to further distinguish them from one another.

Mark other characteristics that you could use to distinguish these foods from one another on the dendrogram. Three green dashes have been left blank for you to fill in on the dendrogram below:
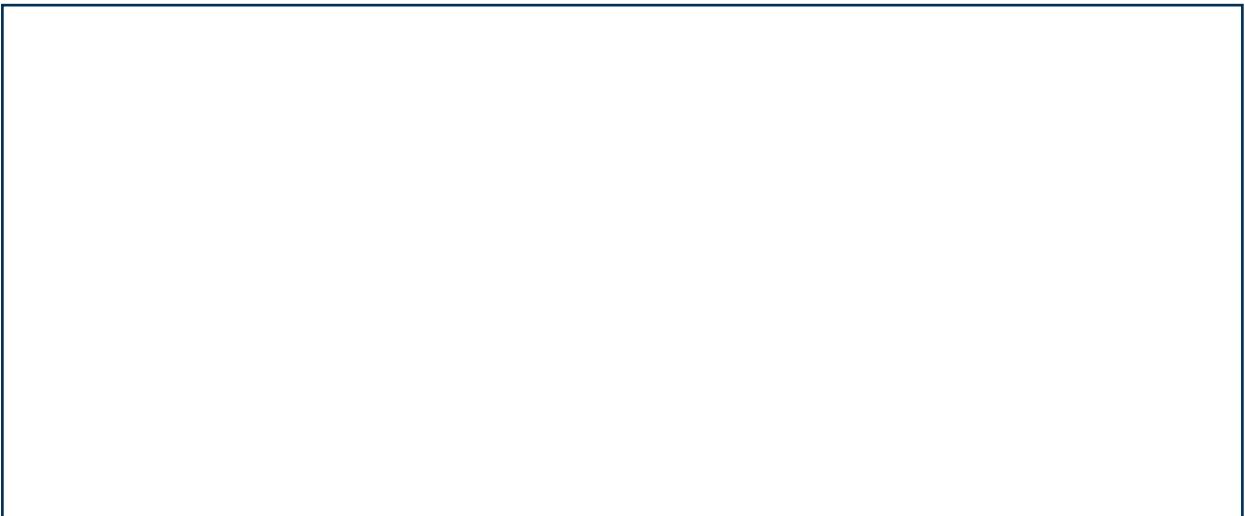
hot

dessert

Crème brûlée

Cookies

Ice cream

Popsicle

not dessert

Soup

Pasta

Salad

Sandwich
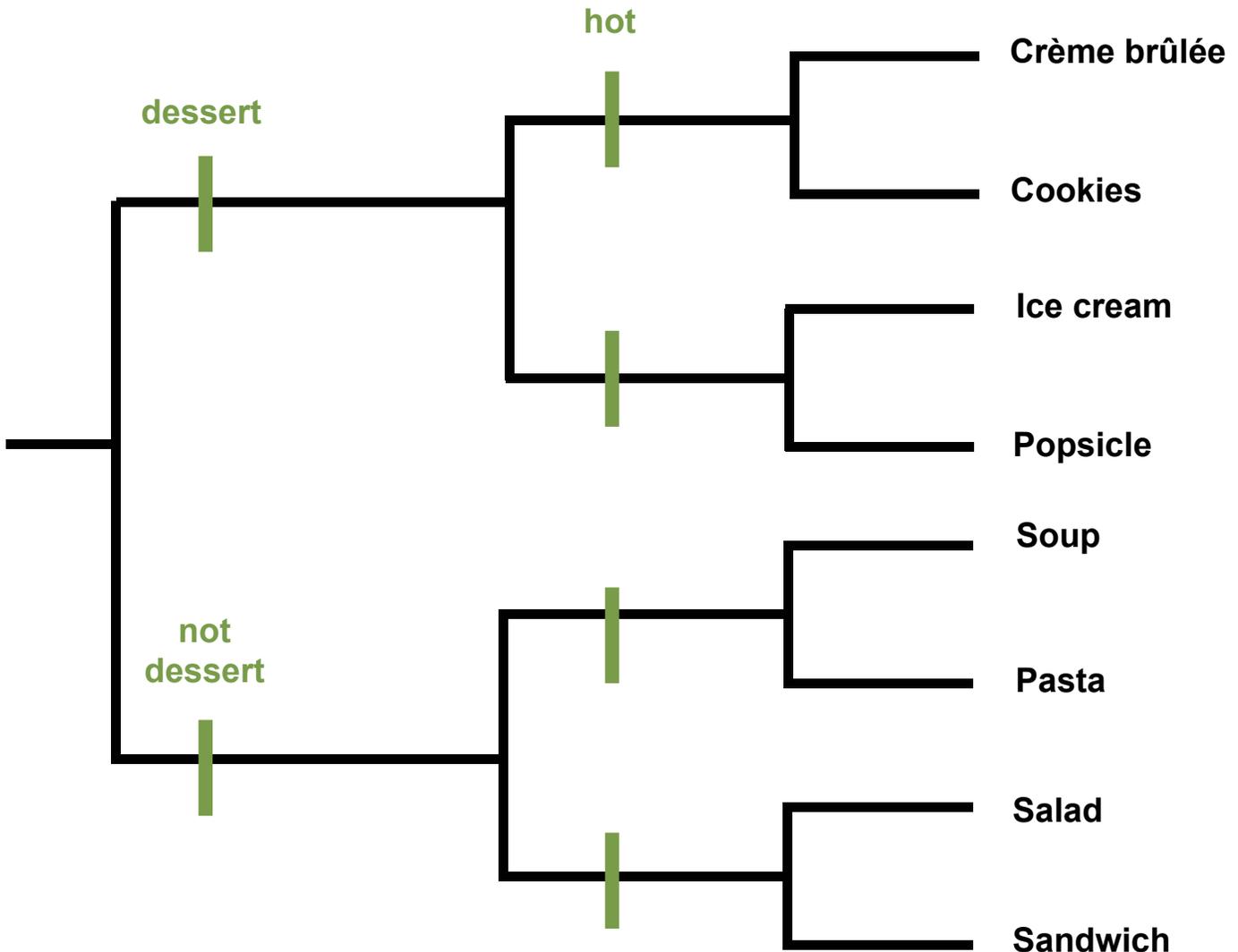
Knowledge Check

- Based on the characteristics you marked on the tree, why do you think soup and pasta are grouped closest together on the dendrogram compared to the other food items?
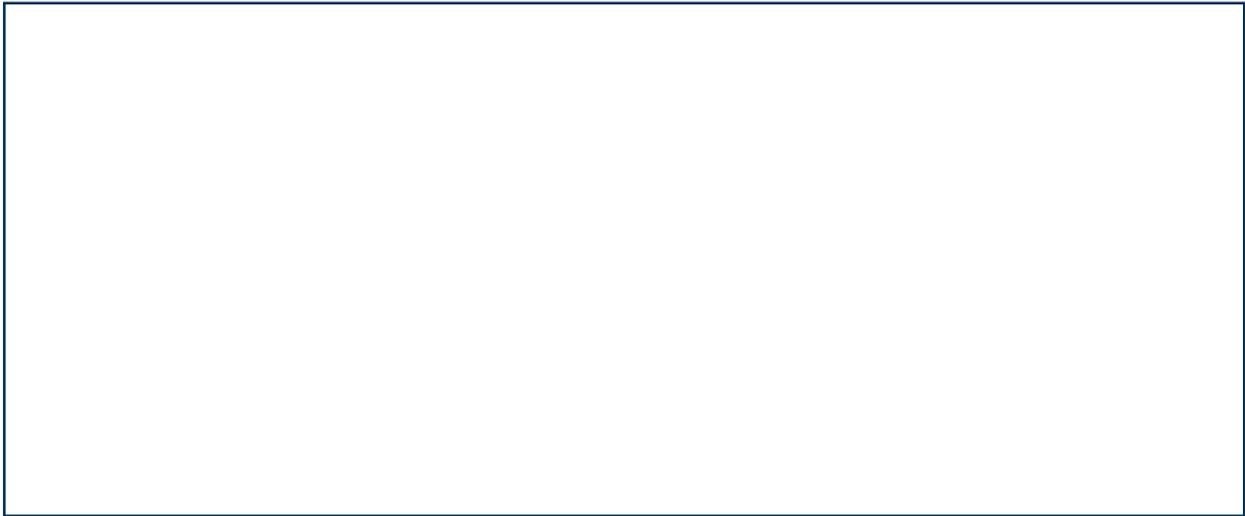
- Can you think of other characteristics other than temperature/type of food that you could use to distinguish these foods from one another?

ALLEN INSTITUTE

Pictured below is the dendrogram we created together and the characteristics we used to label the dendrogram. You may have chosen other food characteristics that could also be correct!

Activity 3: Heatmaps

Now that you are familiar with what dendrograms are and how they organize data in a hierarchal manner, we will build on the idea of dendrograms by adding another type of data visualization to our analysis. In this section, we will focus on interpreting heatmaps. Heatmaps are helpful ways to represent data as colors. Because heatmaps use color instead of numbers, this often makes heatmaps easier to interpret at a glance than simply a table of data.

Example heatmap:
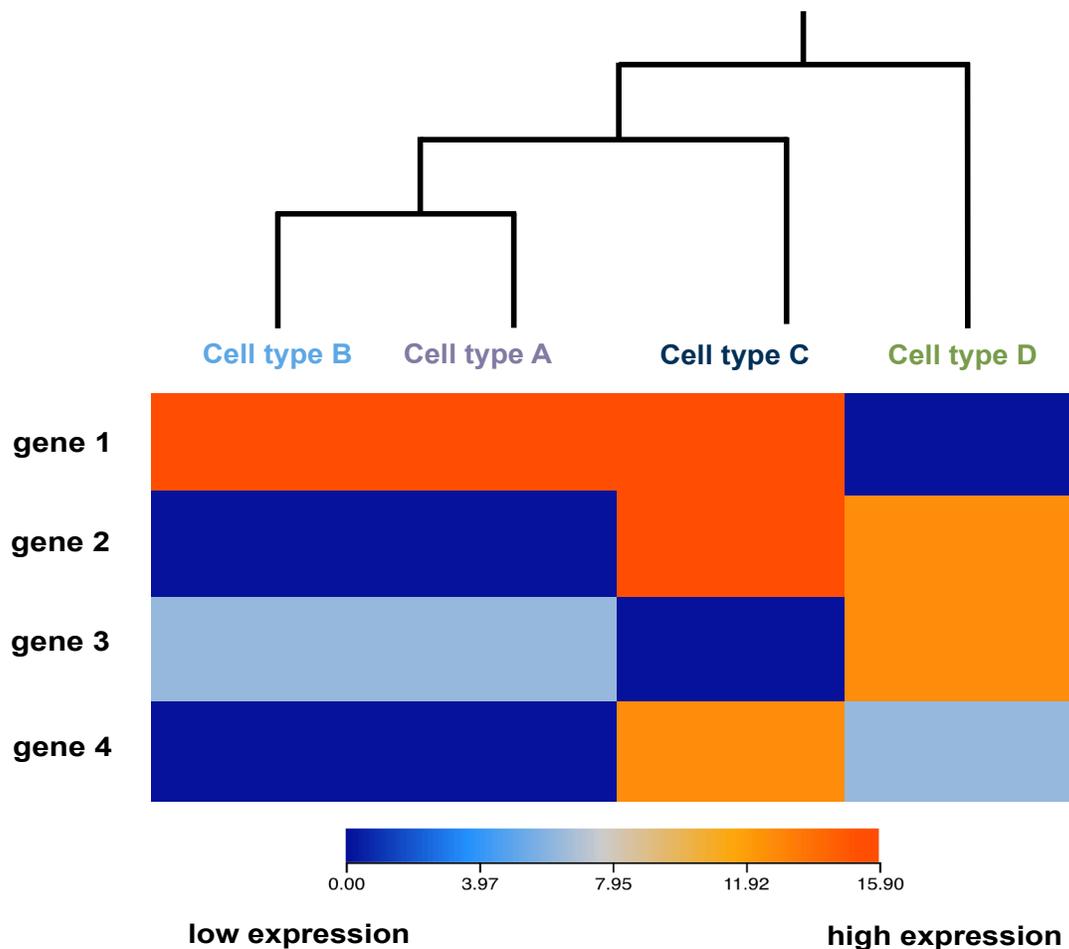
Notice that this figure contains both a dendrogram (shown above) AND a heatmap. The dendrogram above shows 4 cell types. The dendrogram shows which cell types, based on a transcriptomic analysis of their entire genome, are most similar to one another. The heatmap below uses colors to show the relative level of expression for genes 1, 2, 3, and 4 within each cell type. This is the same gene expression data that was used to define the cell types using the clustering method mentioned above, which we will explore more in lesson 4. The gene expression data is now organized and visualized by cell type classification.

Knowledge Check

To practice interpretating a dendrogram and a heatmap in combination with one another, answer the practice questions below using the example heatmap provided on the previous page:

- **Is gene 1 highly expressed across all four cell types? How can you use the heatmap to answer this question?**

- **Based on the data about relative gene expression provided within the heatmap, why do you think cell type A and cell type B are grouped so closely together on the dendrogram?**

ALLEN
INSTITUTE

Activity 4: Analyzing Basic Research Data

The Allen Institute for Brain Science has constructed an extremely detailed dendrogram of the different cell types of the human brain. The dendrogram featured below was constructed using transcriptomic data taken from donor brain tissue. As you learned in lesson 1, brain donation plays a significant role within biomedical research. Although this dendrogram looks significantly more complicated, remember what you learned in activity 1 and 2 about how they are constructed! The dendrogram featured below uses the same concept as you saw before, except now, a significantly higher number of cells and a higher number of genes are being compared.

To view the dendrogram, click here.

The full link is also included here: https://celltypes.brain-map.org/rnaseq/human_m1_10x?selectedVisualization=Heatmap&colorByFeature=Cell+Type&colorByFeatureValue=GAD1

The dendrogram can also be accessed by going to:

1) SEA-AD.org

2) On the Seattle Alzheimer's Disease Brain Cell Atlas page, scroll until you find the "cell types" section. Under "Cell Types" click on the "Transcriptomics Explorer (Reference MTG) link.

Your screen should look like this:
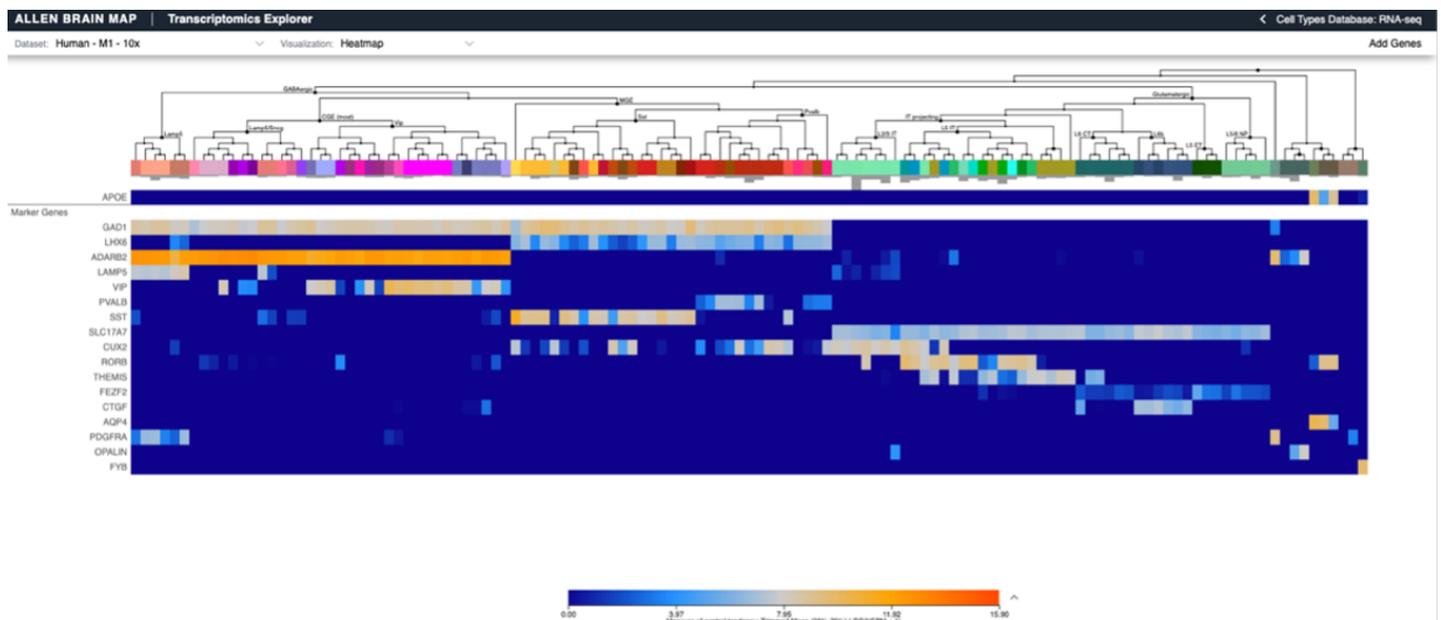
This dendrogram has several interactive elements that you can use to better analyze the data.

Step 1: Hover your mouse over the colors at the end of the **dendrogram**. Notice that hovering over these colors allows you to tell which **cell type** the color represents. For example, when you hover your mouse over the first color on the far left of the dendrogram, we are told this peach color represents the cell type labeled "Neuron 001."
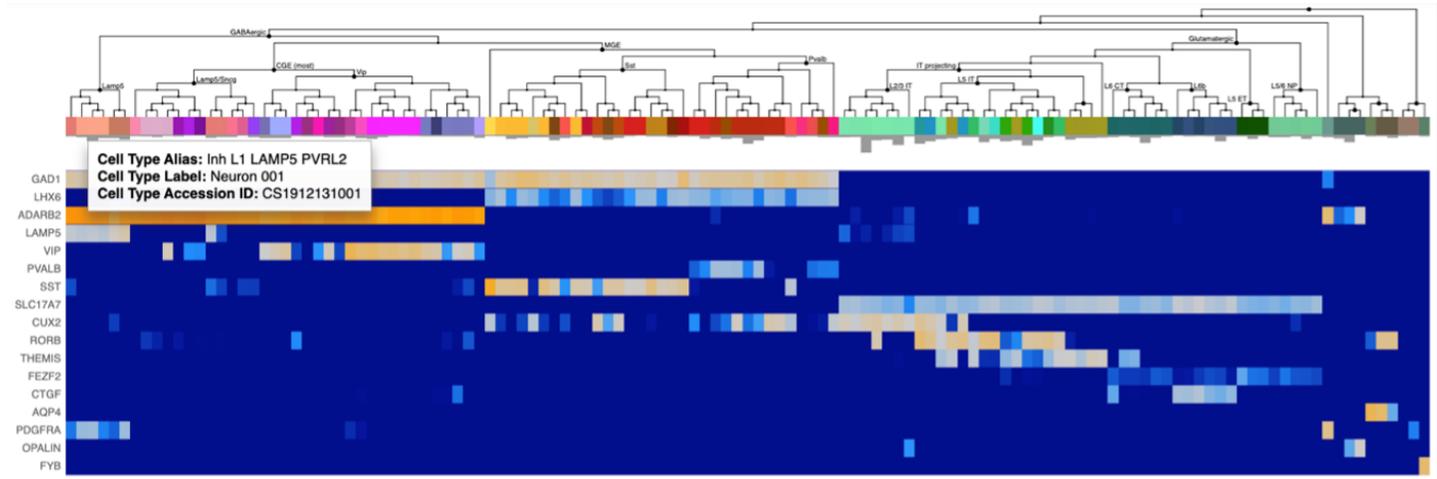

Step 2: Look at the list of genes on the far left. Next to the list of genes is a heatmap. This heatmap uses the exact same scale and color gradient as we saw in Activity 2!

Step 3: Move your cursor to hover over the bottom right hand corner of the heatmap (the orange square). Notice that hovering over this square allows you to see which cell type the box in the square map represents, which gene it is displaying level of expression for, and the "mean expression" of that gene for that specific cell type.
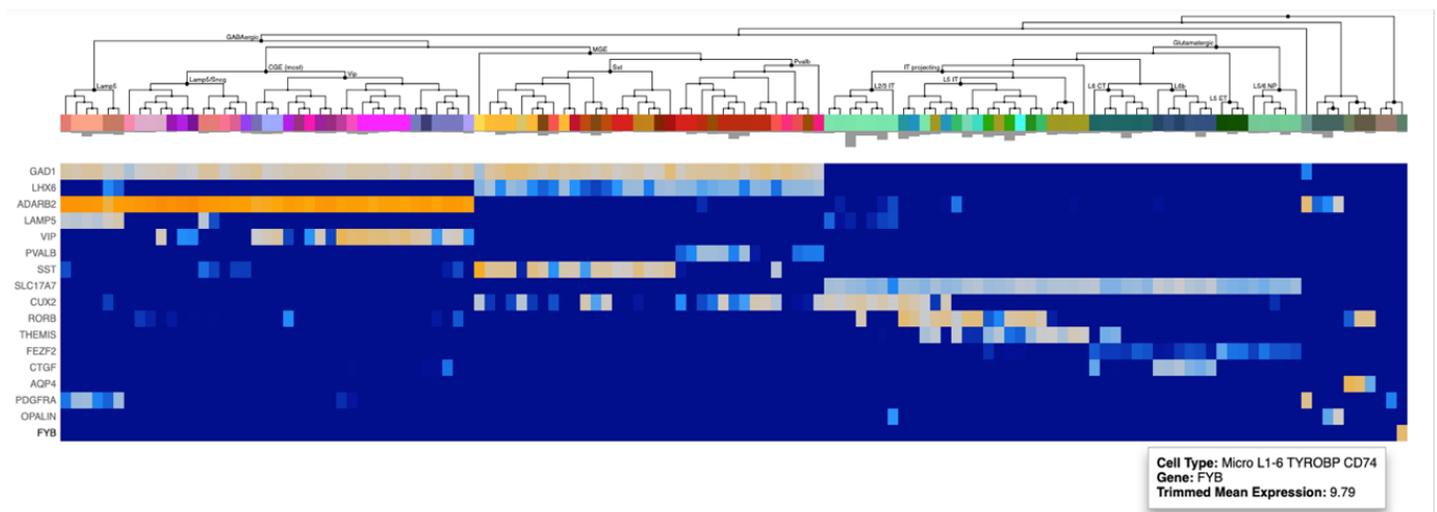
Step 4: Look at the branching of the dendrogram
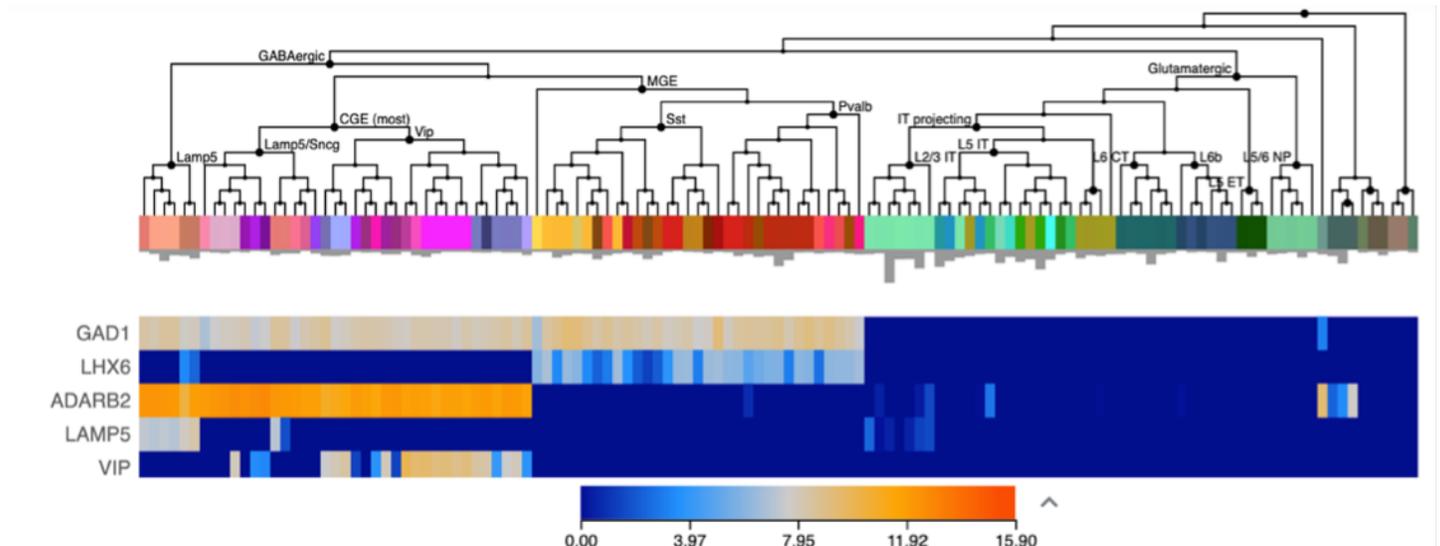
Notice on this dendrogram we have two main branches. One branch is labeled glutamatergic (excitatory) neurons and one is labeled GABAergic (inhibitory) neurons.

In the heatmap, you can see the level to which each **cell type** expresses the **specific genes listed**.

For example, if we look at the very top row of the table, we can see the extent to which each of the **cell types express GAD1**. Looking at the GAD1 gene expression, it appears that only the GABAergic neurons appear to express GAD1. We can see the level of that expression by using the key at the bottom of the page that shows the gradient of color and the relative level of expression associated with each color.

Step 5: Click on the node on the dendrogram labeled "**L6 CT**." Notice that clicking on this node allows you to look more clearly at only the gene expression for the cells that fall under the "L6 CT" branch on the dendrogram.

In order to practice interpreting a dendrogram and a heatmap, answer the following practice questions.

Practice Questions:

- Looking at the dendrogram, it appears like GAD1 is expressed across all types of GABAergic brain cells. According to the heatmap below the dendrogram, which gene seems to be expressed only amongst Glutamatergic brain cells?
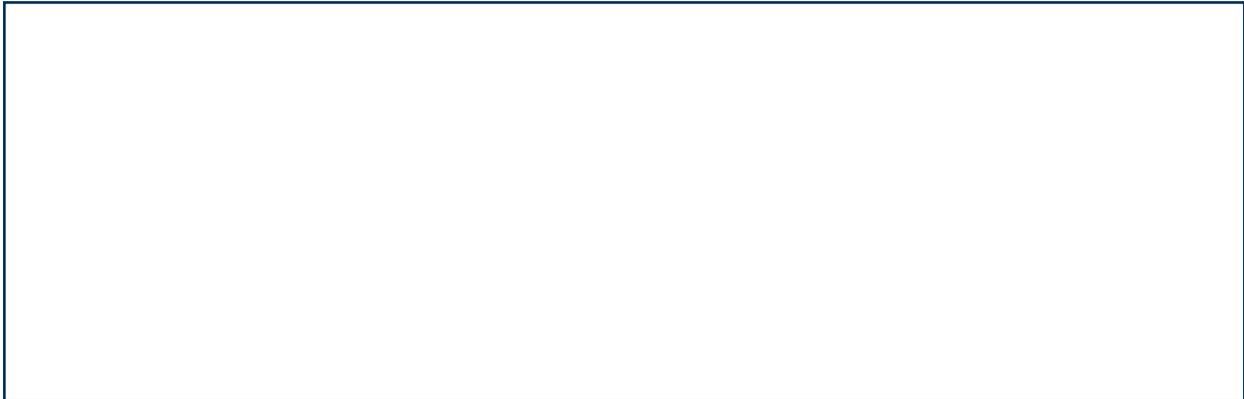
- Click on the node on the dendrogram labeled "L6 CT." Notice that clicking on this node allows you to look more clearly at only the gene expression for the cells that fall under the "L6 CT" branch on the dendrogram. You should see something like this:
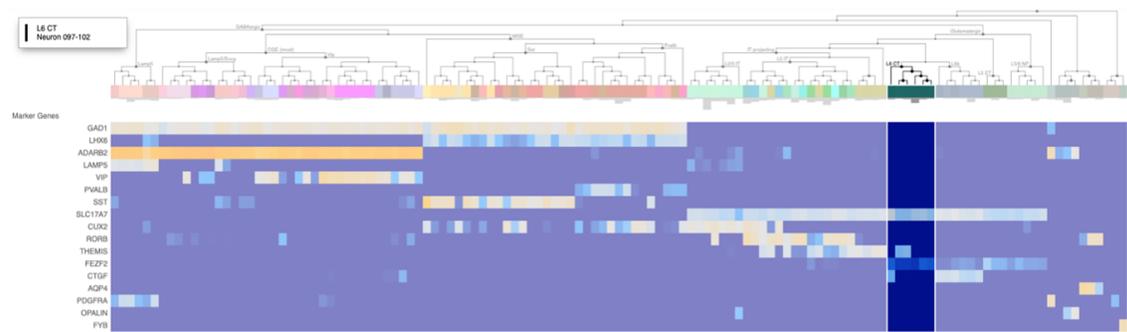
- Which two genes in this heatmap do ALL the L6 CT cells tend to express to at least a small extent?

- Which two cell types appear to express a small amount of the THEMIS gene? Instead of the long name of the cell type, feel free to answer using the neuron number listed below the cell type's name.

- Why is it helpful for scientists to find genes that are only expressed in a small number of cell types?

- This dendrogram uses only transcriptomic data to organize cells into different cell "types." Aside from transcriptomic data, what other type of data do you think would be helpful to compare cells to one another and classify different cell types?

- 4. The transcriptomic data used to create this dendrogram and heatmap was gathered from brain tissue donated by individuals who were considered healthy, neurotypical donors. A healthy donor is someone considered to have no known neurological diseases. In other words, these donors were not known to have neurological conditions such as dementia, Alzheimer's disease, etc. A neurotypical donor is someone who is considered to have the "standard" brain functioning and processing. Why do you think it is important for scientists to study these types of healthy, neurotypical brains in addition to studying brains with neurological conditions?

Conclusion:

Throughout the course of this lesson, you have explored why it is important to study both healthy and diseased brains, and why basic research, like that conducted at the Allen Institute, is so integral to the field of science as a whole. In order to explore the power of basic research, you had access to open transcriptomic data available to the public from the Allen Institute for Brain Science.

If you continue on to complete lessons 3 and 4, you will have the opportunity to dive into the basics of disease research by looking at the data from the Seattle Alzheimer's Disease Brain Cell Atlas from the Allen Institute for Brain Science and its research partners. While this lesson challenged you to practice your data analytic skills for dendrograms and heatmaps, lessons 3 and 4 will introduce you to different methods, such as neuropathology image analysis and UMAP interpretation.

If you develop your own lesson plan using Allen Institute resources, we invite you to share your experience with us at communications@alleninstitute.org. Teachers are also encouraged to publish original lessons using our open data, tools, and other re-sources, and to share those lessons with us.