PM566 Final Project

Christina Lin

Contents

Exploring Cytochrome P450 Enzymes Found in the Human Brain

Introduction

This project is based on my PhD thesis exploring the enzymes in the brain that can produce the neurosteroid pregnenolone from cholesterol. Neurosteroids have important functions in the brain including neuroprotection, anti-inflammation, and modulation of behavior. In classical steroid-producing organs such as the adrenals, pregnenolone is metabolized from cholesterol by the cytochrome P450 (CYP450) enzyme CYP11A1 (see Lin & Papadopoulos, 2021 for review). However, CYP11A1 protein is difficult to detect in the brain and preliminary experiments have revealed that a potential alternate pathway not involving CYP11A1 is used by human brain cells to produce pregnenolone. Our data also suggests that this alternative pathway involves activity of CYP450s. Therefore, this project will analyze known CYP450s in the UniProt database as well as microarray data from the Allen Brain Atlas to answer 2 main questions: 1) Which CYP450 enzymes are expressed in the brain and what are their expression levels? 2) Which of those CYP450 enzymes are involved in cholesterol/steroid metabolism?

Methods

List of CYP450s were obtained from the UniProt database by searching "cytochrome P450". Additional filters were applied: "Homo sapiens(human)" for species and "Reviewed" results to extract information only from manually annotated records from literature and curator-evaluated computational analysis. The columns of interest are protein name, gene name, length (of protein), mass, tissue specificity, cofactor, function, subcellular location, pathway, and sequence. The results were downloaded as a CSV file.

Data Wrangling for UniProt data

Some enzymes have multiple gene names. For this analysis, only the first gene name containing "CYP" will be used. Rows that do not have a gene name starting with "CYP" are removed.

The initial data table started with 82 proteins. After simplifying the gene names and removing entries that do not have "CYP" in the gene name, there are 62 proteins left.

Next, the “Mass” column will be converted to a numeric variable by removing the “,” character and converting the values to integers.

When checking for the masses of the proteins, we see that the lowest mass protein is 12413 Dalton and the shortest length protein is 118 amino acids. Since CYP450s are enzymes involved in complex metabolic pathways and typically have multiple functional domains, these small proteins are likely not CYP450 with cholesterol-metabolizing potential. Therefore, proteins that are less than 35 000 Dalton in mass will be removed.

To look for CYP450s that are expressed in the brains, terms such as “brain”, “cerebellum”, “cerebral”,and “hippocampus” (i.e. common references to different parts of the brain) will be used to filter the observations. For easier search for key terms in later analysis, data within Tissue_expression, Function, Subcellular_location, and Pathway columns will all be converted to lower case.

Data Wrangling for Microarray data

The limitation of using UniProt is that it only shows information about proteins and it does not quantify expression levels. Some CYP450s that are expressed in the brain at low levels, such as CYP11A1, were also not recorded by UniProt. Therefore, we will use another database, Allen Brain Atlas, to look quantitatively at expression of CYP450s in the human brain. The Allen Brain Atlas has microarray data of brain tissue from multiple donors that can be used to analyze RNA expression for genes of interest. The term “CYP” was used to search for cytochrome P450s and the results were downloaded for further analysis. Microarray data were downloaded as 3 separate CSV files: 1) RNA expression level of each sample, 2) details of probes used to measure expression, and 3) details of the tissues used and their donors. Data from the RNA expression and probes CSV files will be merged together. Data about the tissues will be used at the end to identify brain structures of interest. First, the expression data will be pivoted to allow easier merging and analysis later on. Each sample is assigned an arbitrary tissue_id that will be matched with the data from the tissues data set, which will also have the corresponding tissue_id matched. Similar to the UniProt data, probes not associated with a gene name that starts with “CYP” will be removed.

The “expression” column contains log2 intensity levels for each probe, which is indicative of RNA expression level.

Preliminary Results

Analysis of Protein Data

The average mass of cyptochrome P450 enzymes found in humans is 57816.35 Dalton and the average length is 507.2 amino acids. The smallest CYP450 is CYP4Z2P, which is 40159 Dalton in mass and 340 amino acids in length. The largest CYP450 is CYPOR, which is 76690 Dalton in mass and 677 amino acids in length.

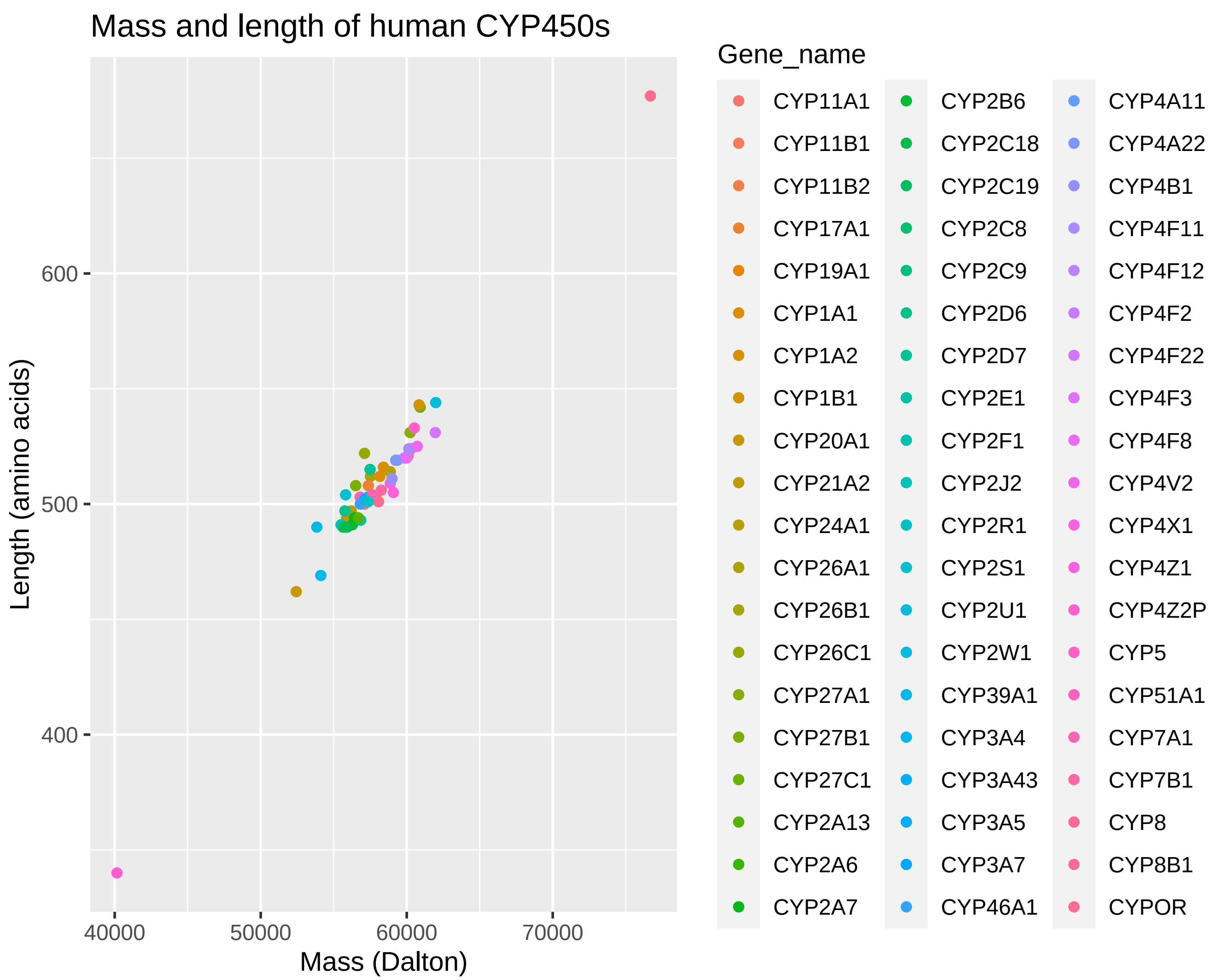

Figure 1: Correlation between mass and length of human CYP450s. As expected, there is a positive linear correlation between mass and length of CYP450 enzymes.

Out of the 60 CYP450 enzymes in humans, there are 13 CYP450s that have protein expression in the brain.

Gene_name	Length	Mass
CYP3A4	503	57343
CYP26B1	512	57513
CYP26A1	497	56199

Gene_name	Length	Mass
CYP19A1	503	57883
CYP27A1	531	60235
CYP2D7	515	57489
CYP46A1	500	56821
CYP2A13	494	56688
CYP1B1	543	60846
CYP4X1	509	58875
CYP2U1	544	61987
CYP7B1	506	58256
CYP4V2	525	60724

Table 1: List of CYP450s expressed in the brain.

Next, the brain CYP450s proteins involved in cholesterol- or steroid-related functions or pathways will be examined.

CYP450s with cholesterol- and/or steroid-related functions are CYP3A4, CYP19A1, CYP27A1, CYP46A1, CYP1B1, CYP4X1, CYP7B1. CYP450s involved in cholesterol- and/or steroid-related pathways are CYP3A4, CYP19A1, CYP27A1, CYP46A1, CYP1B1, CYP7B1. We are able to capture more enzymes when searching for CYP450s involved in cholesterol or steroid related functions. Of the 13 CYP450 enzymes expressed in the brain, 7 of them have functions relating to cholesterol or steroid hormones.

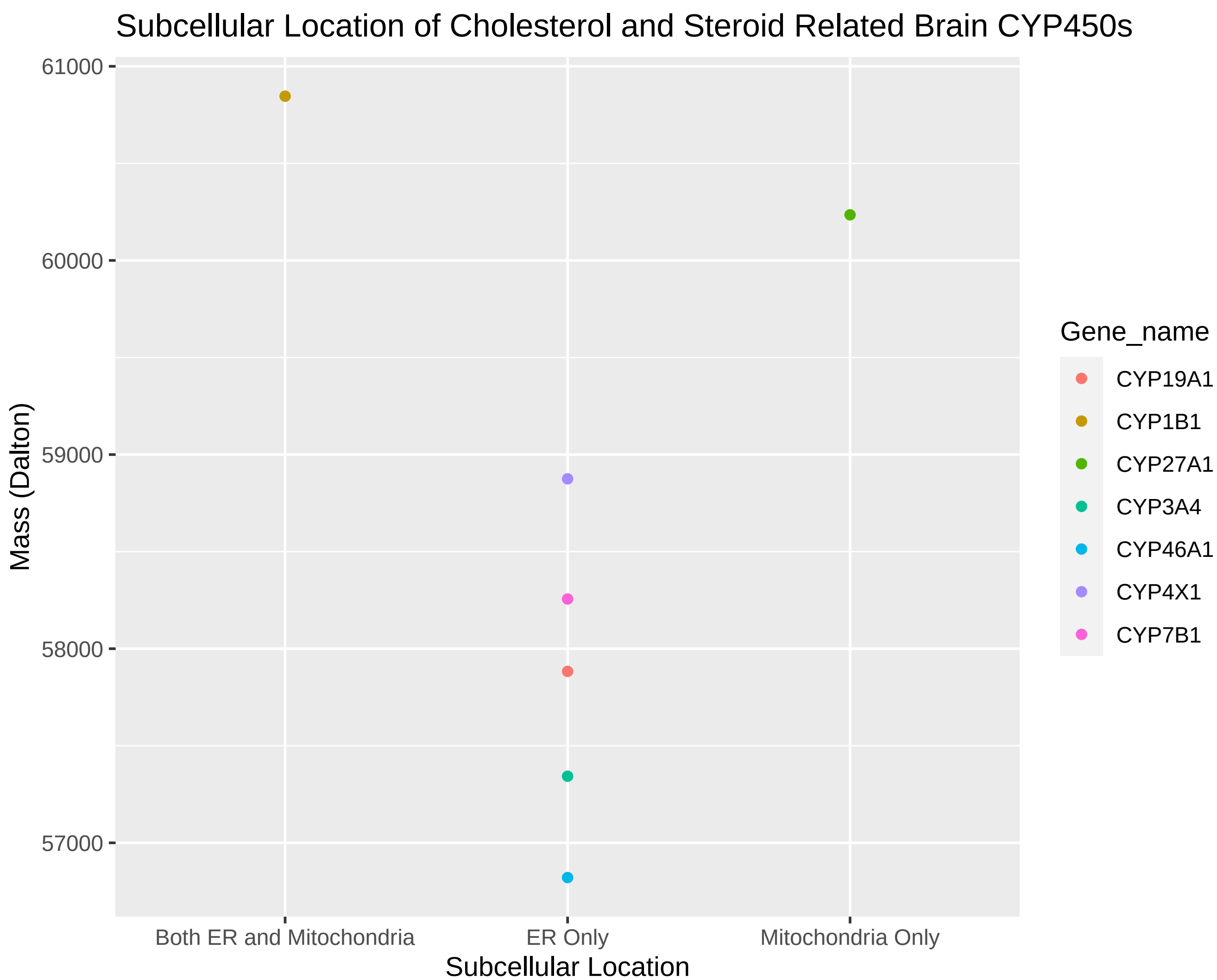

Figure 2: Subcellular location and masses of human CYP450s expressed in the brain that are involved in cholesterol and steroid functions. The majority of these enzymes are located in the endoplasmic reticulum. CYP27A1 is the only one of these enzymes that is only located in the mitochondria while CYP1B1 is found in both the endoplasmic reticulum and mitochondria. The enzymes that localize to the mitochondria also appear to have higher masses than those that are only found in the ER.

Therefore, the CYP450 enzymes of interest found using the UniProt database are listed in the table below.

Gene_name	Mass	Length	Organelle
CYP3A4	57343	503	ER Only
CYP19A1	57883	503	ER Only
CYP27A1	60235	531	Mitochondria Only
CYP46A1	56821	500	ER Only
CYP1B1	60846	543	Both ER and Mitochondria
CYP4X1	58875	509	ER Only
CYP7B1	58256	506	ER Only

Table 2: CYP450 Proteins Expressed in the Human Brain with Cholesterol and/or Steroid-related Functions.

However, the data from UniProt are not quantitative and the assays used to detect protein may not be sensitive enough to measure the expression of CYP450s in the brain, which can be bound to organelles or membranes. Therefore, the RNA expression for CYP450s were also examined and compared to CYP11A1, the enzyme whose activity we are trying to find an alternative pathway for.

Analysis of Microarray (RNA) Data

The average expression for each gene was calculated for each donor and brains structure, and compared to the average overall CYP11A1 expression. The genes with equal or higher expression than CYP11A1 were filtered out. The filtered data set was combined with the UniProt data to view more information about these genes.

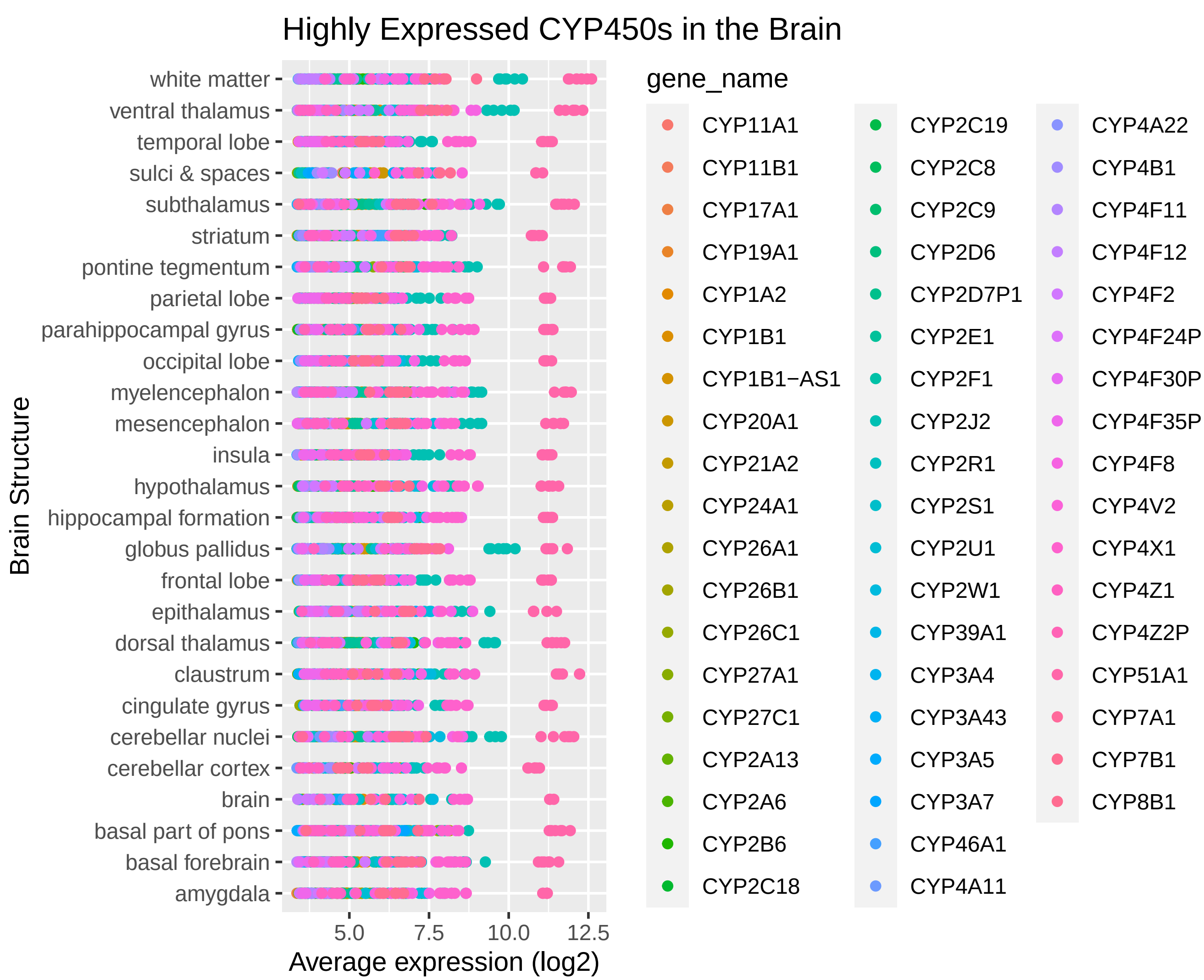

Figure 3: RNA Expression of Highly Expressed CYP450s at different ages. There are 55 CYP450 enzymes with equal or higher expression than CYP11A1 in the brain. As seen from the graph, CYP51A1 appears to have the highest expression out of all the CYP450s, followed by CYP4X1 and CYP2J2.

Now seeing which of these highly expressed CYP450s are involved in cholesterol or steroid function.

gene_name	Expression_log2	Mass	Organelle
CYP4X1	7.9	58875	ER Only
CYP7B1	6.4	58256	ER Only
CYP2B6	6.3	56278	ER Only
CYP1A2	5.3	58407	ER Only
CYP1B1	4.9	60846	Both ER and Mitochondria
CYP21A2	4.9	55887	ER Only
CYP27A1	4.8	60235	Mitochondria Only
CYP3A7	4.8	57470	ER Only
CYP46A1	4.4	56821	ER Only
CYP2D6	4.3	55769	ER Only
CYP3A5	4.3	57109	ER Only
CYP3A4	4.1	57343	ER Only
CYP2C9	4.1	55628	ER Only
CYP39A1	4.0	54116	ER Only
CYP4B1	4.0	58991	ER Only
CYP11A1	4.0	60102	Mitochondria Only
CYP2C8	3.8	55825	ER Only
CYP17A1	3.7	57371	ER Only
CYP24A1	3.7	58875	Mitochondria Only
CYP19A1	3.6	57883	ER Only
CYP7A1	3.5	57661	ER Only

Table 3: CYP450s with cholesterol related functions with RNA expression equal to or higher than CYP11A1 in the brain. We see that there are 21 CYP450s with cholesterol related function when the two datasets are combined together, which are listed above.

Finally, it would be interesting to see if the expression of these CYP450s change with age. Genes that do not have enough data points to show age-related trends were removed from the plots.

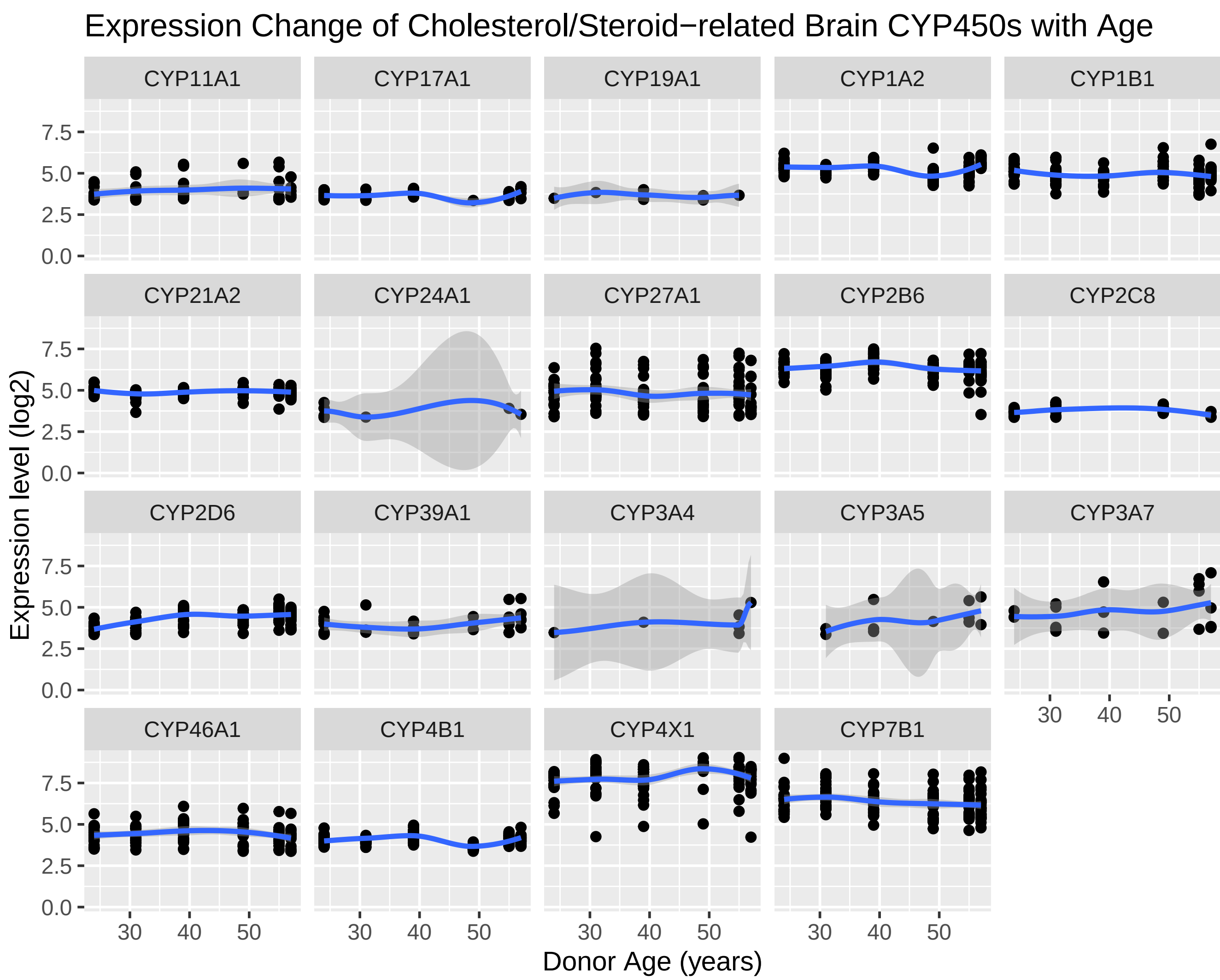

Figure 4: Expression change of brain CYP450s involved in cholesterol/steroid function with age. There appears to be no changes in the expressions of these CYP450s with age.

Conclusion

In this analysis, CYP450 enzymes that are expressed in the human brain with functions related to cholesterol and steroid were extracted from a data set containing all human CYP450s. There are 7 of such proteins, which are CYP3A4, CYP19A1, CYP27A1, CYP46A1, CYP1B1, CYP4X1, CYP7B1. Analysis of microarray data revealed 21 of such genes: CYP4X1, CYP7B1, CYP2B6, CYP1A2, CYP1B1, CYP21A2, CYP27A1, CYP3A7, CYP46A1, CYP2D6, CYP3A5, CYP3A4, CYP2C9, CYP39A1, CYP4B1, CYP11A1, CYP2C8, CYP17A1, CYP24A1, CYP19A1, CYP7A1.

The expression for these genes in the brain can then be validated in future experiments using qRT-PCR and protein analysis on human brain tissue RNA.