Understanding Fat Cadherins from an Evolutionary and Contextual Point of View

Citation

Permanent link

Terms of Use

Share Your Story

Understanding Fat Cadherins from an Evolutionary and Contextual Point of View

Sadie Schlabach

A Thesis in the Field of Biotechnology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

March, 2022

Abstract

Fat cadherins are an essential planar cell polarity molecule that is conserved throughout evolution. Yet little is known about what regions of these molecules has been conserved over time. This thesis describes the importance of these molecules in *C. elegans, Drosophila melanogaster*, *Mus musculus,* and *Homo Sapiens.* Furthermore, a Python pipeline has been established to examine gene expression of different molecules of interest in Mus musculus.

Dedication

This thesis is dedicated to my partner, Sean, who has been understanding of the late nights writing code, my parents who have supported me and believed in me since the beginning, and my siblings who are my biggest cheerleaders.

Acknowledgments

Thank you, Michael Agostino for your constant support throughout this process. Thank you for your quick responses and guidance when I was feeling lost. Thank you, Ian Woods for first showing me how to write code. Thank you, Evelyn, for first introducing me to the coolest planar cell polarization molecule – Fat3!

Table of Contents

List of Tables

List of Figures

Chapter I. Introduction

Planar Cell Polarization molecules autonomously allow a cell to have a sense of direction and distinction within a cell. This is important for development, coordination between cells, and localization of inner cell mechanisms. In development, cell orientation and movement permit proper placement of tissue and organs, which is critical to an organism's maturation. This cellular orientation and movement arise from cell polarization. Planar Cell Polarization molecules lead to differences within a cell through a defining signal which identifies "up" from "down." Without this signal a cell will not know which way to migrate during development and the developing organism will not grow with "normal" patterning. Moreover, organs and tissues will be dislocated, and efficacious organism maturation cannot occur.

Planar cell polarity is also important for non-autonomous cell communication when gene expression within one cell impacts the fate of another. Cell non-autonomous communication is critical for many key signaling pathways such as, mTOR, Shh, Notch, etc. Many inner cell mechanisms also rely on planar cell polarity molecules to provide a cue for residency. Without these molecules there is no sense of direction within a cell.

Cadherins are examples of planar cell polarity molecules. Cadherins play an integral role during development, particularly in the nervous system, where they help to organize cells into coherent networks through modulation of cell morphology, management of cell-cell interaction, communication, cell recognition, and protein

processing. Over 100 cadherin families have been identified in humans. These families can be organized into three main groups: major cadherins, protocadherins, and cadherin-related proteins (Sahin Gul, 2017).

The focus of this thesis is the Fat family of cadherins. The Fat family has only 2-4 members and is found throughout evolution, from fish to mice to humans (Katoh, 2006). It is suspected that in all of these organisms, Fats have a similar role, impacting developmental migration and cell-to-cell communication. While Fats are evolutionarily conserved and involved in pivotal processes there is still little known about the molecules. The Fat cadherins all have similar properties but Fat3, found in mammals, has been studied the most extensively (Aviles, 2017). My thesis project aims to provide a comprehensive analysis into these cadherin molecules. I have used bioinformatics to analyze Fats through two different lenses. First, I have examined the sequence of Fat family members across different, key, evolutionary time points. Secondly, I have investigated gene expression data within the Fat3 network. Together this creates a comprehensive analysis of an important planar cell polarization molecule. Furthermore, this provides an outline for how to analyze other molecules of interest.

To better understand the evolutionary context of Fats, I have examined the sequences of Fat across different organisms. This sequence analysis shows how Fat cadherins are conserved across different evolutionary time points. Importantly, it is valuable to compare Fat sequences at these different time points because this may determine what regions of each of the Fat cadherins have been deemed "important", and therefore conserved between primitive organisms and human.

The second part of my project examines gene expression data of Fat3, other Fat family members (Katoh, 2012), and genes known to interact with Fat3, so-called gene partners (Dang, 2013). I have utilized the Allen Brain Atlas to retrieve Fat1, Fat2, Fat3 and Fat4 expression data, showing differential expression patterning for each Fat expressed in mammals. I have also examined expression data for Fat3 partners, showing what brain regions and at what levels these partners are expressed throughout the brain. This analysis showed that there are specific Fat3 interactors that show similar patterning to Fat3 while some interactors show unique patterning. This differentiation in gene expression might indicate how interactions between Fat3 and Fat3 partners might occur. The Allen Brain Atlas (ABA) provides *in situ* data of specific genes through a searchable tool on their site. This resource has been especially significant for integrating and understanding extensive expression data simultaneously (Liscio, 2014). Utilizing Python, I have web-scraped the ABA data for each of the genes of interest and generated graphs showing gene expression data across different brain regions for each gene of interest. This data showcases the way that genes in the Fat3 network are expressed similarly and differently. This clarifies what Fat3 interactors are co-expressed in brain regions and at levels comparable to Fat3. Alternatively, this data highlights what interactors are not coordinately expressed with Fat3.

At the end of the project, I have a better understanding of the Fat family members from an evolutionary context through sequence analysis as well as gene expression data. I have examined how genes in the Fat3 network vary their expression across different brain regions. I have also examined the sequence structure of Fat3 across key evolutionary time points. Altogether, this project examines the Fat family in a multifaceted manner,

utilizing the many skills I have learned pursuing my Masters degree in Biotechnology. This project creates a two-dimensional outline for myself and others to use to better understand the evolutionary context through comparison of Fats over time, and expression data of Fat3 gene partners or specific genes of interest.

Background of the Research Problem

In general, Fats are extremely large, atypical cadherins, involved in cell adhesion, tissue growth, and planar cell polarity (Sopko, 2009). Fats are conserved across vertebrates and mammals, being found in humans as well. Despite this conservation, Fats play different roles in different organisms, though they are involved in the same processes. Table 1.1 summarizes what Fat cadherins are present in *Drosophila melanogaster, Danio rerio, Homo sapiens*, *C. elegans*, and *Mus Musculus*. The fat superfamily of cadherins consists of Fat1, Fat2, Fat3, and Fat4 in mammalian organisms. Fat-like and Fat2 were found in *Drosophila melanogaster* in 1923 (Mohr: Flybase, 1994). *cdh-4* was found in *C. elegans*, and Fat3a has been investigated in *Danio rerio.*

Drosophila melanogaster Fat, the establishing member of the Fat superfamily, was first identified by Mohr in 1923. Early studies showed that loss of Fat led to tissue overgrowth because without Fat, there was a lack of apical-basal polarity that Fat establishes in the *Drosophila melanogaster* egg chamber (Sopko, 2009). The egg chamber is an organ-like structure in the fly ovary that gives rise to a single egg and is essential to proper *Drosophila melanogaster* development. During development the egg chamber harnesses many signaling molecules in a coordinated effort to promote an ellipsoidal shape (Horne-Badovinac, 2017). During this period, another Fat family

4

member, Fat-like (or Fat2, or Kugelei), acts as a planar cell polarity molecule within the egg chamber, signaling from each cell's trailing edge to implement formation of protrusions in the cell behind. The Fat signal is an input to other cells to adjust their positioning. In this case another molecule, Lar, retracts at the trailing edge of the cell behind. This movement of Fat-like, and in turn Lar, helps the egg chamber to change its shape into the ellipsoidal shape that is necessary for proper egg formation and further development of *Drosophila melanogaster.*

It has also been shown that in *Drosophila melanogaster* the loss of Fat leads to a loss of organization of the cells within the epithelium (Sopko, 2009), suggesting that Fat is necessary for proper organization of *Drosophila melanogaster* epithelia. Interestingly and importantly, the extracellular domain of Fat does not seem to be necessary for proper egg chamber formation. This finding is unique to Fat cadherins as typically, extracellular domains are a critical region in cadherin structure. This region is often imperative to the role of cadherins in cellular polarity modulation. However, Matakatus, et al., showed that much of Fat function in *Drosophila melanogaster* could be rescued by a truncated version of the molecule that is limited to its cytoplasmic domain (Matakatsu, 2006). This finding illustrates the importance of the cytoplasmic domain to Fat function, at least in *Drosophila*, and may be in other versions of Fat cadherins as well.

In all organisms, precise and accurate neural migration is imperative for proper tissue patterning throughout nervous system development. In *Caenorhabditis elegans* this patterning happens through the migration of the Q neuroblast. *cdh-4* in *C. elegans* is a cadherin repeat transmembrane protein and part of the Fat family of cadherins. Like Fat cadherins, *cdh-4* has been shown to be involved with cell migration, axon development,

and migration of the Q cell neuroblast descendants. The guidance of *cdh-4* happens through an interaction with Lar (similar to *Drosophila melanogaster* Fat/Lar interactions), in a cell autonomous manner. Furthermore, *cdh-4* mutants in *C. elegans* have variable morphogenetic defects in which Q cells do not migrate to the proper destinations. It has also been shown that *cdh-4* mutants show fasciculation defects or improper axon and dendrite formation. Finally, there is larval lethality in *cdh-4* null animals (Schmitz, 2009). Altogether, this shows that in *C. elegans*, Fat cadherins or Fat cadherin ancestors again play an important role in cellular positioning and interaction during development.

Four mammalian Fat genes have been identified (Badouel, 2015). Fat1, Fat2, and Fat3 are most similar to D*rosophila melanogaster* Fat-like, while Fat4 is similar to *Drosophila melanogaster* Fat. Fat1, Fat2, and Fat3 are all found in early mammalian development, and Fat4 is only found in the adult. In particular, Fat3 has been studied extensively in mouse retinal amacrine cells. Here, it seems that Fat3 is necessary to ensure that amacrine cells develop a specific unipolar morphology that is critical for proper amacrine cell migration. In wild-type mice, amacrine cells migrate through the neuroblast layer to the inner nuclear and ganglion cell layers, gradually losing one neurite, becoming unipolar over this transition. However, in Fat3 null animals, amacrine cells retain the extraneous neurite and do not travel through the inner nuclear layer in the same coordinated manner as amacrine cells in wild-type mice do (Deans, 2011).

Mammalian Fat1 mutants die at birth. This has not been examined extensively but mutants seem to have deformed eyes and poor tissue formation, which likely leads to improper nervous system development and lethality. Fat1 also has a role in actin

cytoskeletal dynamics as well as muscle formation. Targeted ablation of Fat1 results in muscle shape defects in limbs (Caruso, 2013).

Meanwhile, Fat4 is involved in chick neural tube development (Van Hateran, 2011). Targeted RNAi knockdown of Fat4 in chicken results in overgrowth of neural progenitors in the developing chick neural tube. Again, suggesting that this fat cadherin has a role in cell-cell communication and signaling, further influencing tissue and organ formation.

While there has been extensive research on Fat cadherins or predicted Fat cadherins *in C. elegans, Drosophila melanogaster, Mus Musculus*, and *Homo sapiens,* there is limited analysis of Fat3a which is found in *Danio rerio.* However, unsurprisingly, it has been shown that in *Danio rerio*, Fat3a again is involved in skeletal and morphological processes during Zebrafish development (Pabic, 2014). In Zebrafish embryos that are Fat3 null, cartilage chondrocytes irregularly stack and lose polarization of cell-cell intercalation. This leads to structural malformations which can result in failure of properly organized tissue and organs.

Altogether, Fat family members and Fat predicted family members in mammals, *C. elegans*, *Drosophila,* and *Danio rerio* have some role in planar cell polarity, migration during development, and cell adhesion. In each of these organisms Fats have similar overarching roles and involvement in planar cell polarity pathways. However, mechanistically, Fats seem to be functioning in a diverse manner. While there have been many studies into the role of Fats for each of these different organisms (mainly *Mus Musculus* and *Drosophila*) there has yet to be a comprehensive examination of Fat phylogeny across variant species.

Though this molecule is conserved throughout many organisms and has a specific involvement in specific properties, it is clear that in these different organisms the mechanisms that Fat family members enact are unique to each organism. This uniqueness leads to an interesting question – how has the structure and properties of Fats changed overtime?

I have examined how Fats have evolved through sequence analysis of the different variants of Fats or predicted Fats in these organisms: *Danio rerio*, *C. elegans*, *Drosophila*, *Mus Musculus*, and *Homo sapiens*. This has allowed me to speculate as to what parts of the protein family have been conserved and changed over time.

Through the specific investigation of Fats in different species there has been some examination into potential and known Fat3 gene partners. Specifically, it has been shown with GST pull down techniques in mammals that Fat3 binds with the WAVE regulatory complex (WRC) (Chen, 2014). Like Fat3, the WRC is evolutionarily conserved and has some control over actin cytoskeletal dynamics which might play a role in planar cell polarity pathways. The WRC complex has five components: SRa1/Cyflip1, Nap1/Hem2/Kette, Abi2, and WAVE1/SCAR, and HSPC300/Brick1. Wave proteins have been shown to have some role in the Actin Related Protein 2/3 Complex (Arp2/3-complex) mediated actin assembly in many systems. The Arp 2/3 complex is a seven-subunit protein complex that is involved in actin skeletal dynamics. The WRC is traditionally clustered densely at its site of action in cells. This placement seems to be necessary for restricting actin assembly during processes like polarized cell movement (Chen, 2010). Because of the WRC being clustered near its site of action it would be interesting to examine the expression pattern of WRC members compared to Fat

members in mammals. This investigation could provide insight into how the WRC might bind to Fat. At the very least this analysis could lead to focused experimentation to validate how the WRC might interact with Fat members.

In *Drosophila melanogaster*, it's been shown that WAVE is activated by the Arp2/3 complex and its partner Ena/Vasodilator-stimulated phosophoprotein (Ena/VASP) to facilitate proper egg chamber elongation. The loss of the WRC or the loss of its interaction to Ena/VASP results in improperly shaped, round eggs. The specific localization of the WRC at tricellular junctions is similar to Fat-like and Dlar localization that has also been reported at those same junctions. The overlap of this localization between Fat-like and the WRC suggests that Fat-like could be involved in mediating WRC-dependent actions to influence proper egg chamber formation. Furthermore, it's been shown that the C-terminal intracellular tail of Fat-like has three highly conserved WIRS motifs that might mediate a direct interaction with the WRC (Squarr, 2016).

Because it seems like the WRC could directly interact with Fat3, Fat3 homologs and Fat3 family members (Fat2 in *Drosophila melanogaster*), it is worthwhile to compare mouse gene expression of Fat3 family members and the WRC. It is informative to examine whether they are expressed in similar areas, with similar levels. It is possible that one or more of the WRC members are expressed in the same areas of Fat3 while other WRC members are not. If WRC members are expressed in similar amounts and similar regions to Fat3 then it is possible that the proteins are interacting. Alternatively, if Fat3 is expressed in one brain region and a WRC member is not, those two proteins cannot be interacting. Altogether, this could provide insight into how the WRC might bind to Fat3 by identifying which members are interacting with Fat3.

This deeper analysis of Fat3, and other Fat family members, phylogeny, and the expression of the WRC allows us to have a more inclusive understanding of the role of Fats in mammals.

Chapter II.

Materials and Methods

My thesis project follows a multi-faceted approach aimed at answering two questions; How has Fat3, and other Fat family members, been conserved evolutionarily over time? And how does the protein expression of Fat3 and other Fat family members compare to the expression of Fat3 interactors?

Phylogenetic Analysis

To answer the first question, I have created phylogenetic trees of Fat3, Fat3 homologs and other Fat family members, by utilizing UniProt (Uniprot.org, retrieved September 2021) and the Clustal Omega software program (https://www.ebi.ac.uk/Tools/msa/clustalo/, utilized October 2021) . I retrieved protein sequences from UniProt (Table 1). The sequences were then uploaded to the Clustal Omega alignment program. ClustalO works so that it aligns the protein sequences and creates a phylogenetic tree based on similarity between the sequences (Figure 1).

Sequence Analysis

After establishing what organisms have some version of the Fat3 protein (or other Fat family members) I have analyzed how those protein sequences compare. I have gathered FASTA protein sequences from each organism of interest and performed a multiple sequence alignment through BLAST to see what regions of the gene have been conserved and what regions have been lost over time. This analysis gave me a starting point to do more sequence analysis aimed at understanding exactly what are the

11

conserved and divergent regions. Perhaps conserved regions relate to an important

protein domain. This information helps us to understand how sequence might relate to the

function of Fat3, and other Fat family members.

Expression Analysis

The second part of the project examines protein expression in mouse. The Allen

Brain Atlas is a public website (Brain Map, retrieved 2020) that provides *in situ* protein

expression data for all of known proteins expressed in mouse. Because we know that Fat3

has a binding motif for the Wave Regulatory Complex, I have examined where each of

the proteins in the WRC are expressed and compared this expression data to where Fat3

is expressed. I am curious whether there is overlap in the expression data which could

provide insight into how Fat3 is interacting with the WRC. To get at these questions I

downloaded the protein expression data via Python so that I could analyze the

location/quantification of the gene expression in a directed manner. This established

Python pipeline can be utilized in the future for analysis of other proteins of interest.

Both methods together should provide an outline for examining a protein of

interest in a multi-pronged manner, in turn yielding a starting point for thorough protein

analysis and understanding.

Chapter III.

Results

Phylogenetic Tree

The cytoplasmic domain for various Fat protein sequences, identified via UniProt were

inputted into Clustal Omega to generate a phylogenetic tree based on the protein

sequence alignment score (Figure 1). The protein sequence alignment score is established

from a multiple alignment of the sequences analyzed. The multiple alignment calculates

pairwise identities, and the scores are based on evolutionary distance between the protein

sequences analyzed. A short branch on the phylogeny represents a small evolutionary

distance between the sequences while a longer branch represents a larger distance, or

higher score (Bioinformatics Tools FAQ, 2019).

There are two Fat proteins that do not have an annotated cytoplasmic domain, *Danio*

rerio, Fat3a and *Mus musclus*, Fat1. These protein sequences were excluded in the

phylogenetic analysis so that the tree could be limited to the protein region of interest.

However, their UniProt ID is still included in Table 1 for review. The compete FASTA

sequences are found in Figure 2, Appendix.

The phylogenetic tree shows several clusters. *Homo sapiens* Fat3 and Mus

musculus Fat3 are grouped together. *Homo sapiens* Fat2 and *Mus musculus* Fat2 are

grouped together. *Homo sapiens* Fat4 and *Mus musculus* Fat4 are also grouped together.

Drosophila melanogaster Fat2 is separated from *Homo sapiens* and *Mus musculus* Fat2.

This suggests that there was some divergence in Fat2 between mammals and *Drosophila*

melanogaster. Homo sapiens Fat4 and *Mus musculus* Fat4 are grouped separately from

13

the other Fat proteins in *Homo sapiens* and *Mus musculus*. Fat4 is only expressed in adult mammals while the other Fat proteins are expressed throughout development. *C. elegans cdh-4* and *Drosophila melanogaster* Fat are on their own branch. *Cdh-4* shows more similarity to Fat1, Fat2, and Fat3 than Fat4. *Drosophila melanogaster* Fat shows more similarity to *Mus musculus* and *Homo sapiens* Fat4.

Altogether, there are similarities for Fat proteins that share the same name, as well as similarities for Fat proteins that are expressed in the same organism. This tree provides reason to examine whether *cdh-4* shows functional similarities to Fat1, Fat2, and Fat3. It also suggests that *Drosophila melanogaster* could show some similar functionality to mammalian Fat4. Fat4 likely has some unique functionality distinct from Fat1, Fat2, and Fat3. Finally, it would be worthwhile to establish the cytoplasmic domain for *Danio rerio* Fat3a and *Mus musculus* Fat1 to see how they would cluster among the other Fat proteins.

Sequence Analysis of Fat Cadherins

To better understand the various mechanisms that Fat cadherins utilize in different organisms, I have examined the sequences of these cadherins across organisms in which they are conserved. Because Fat cadherins are such large proteins, and to make sequence analysis more feasible, I have again limited my comparisons to the cytoplasmic domain for each of the proteins. There are two Fat proteins that do not have an annotated cytoplasmic domain, *Danio rerio*, Fat3a and *Mus musclus*, Fat1. These proteins are excluded from the sequence analysis.

14

The cytoplasmic domain is particularly interesting because this is where the WRC

binding site has been predicted (Bayou, 2014) in *Homo sapiens* Fat3. Furthermore, this

domain (in Fat-like) has been experimentally shown to be the site of WRC binding in

Drosophila. Finally, Fat3 developmental phenotypic dysfunction has been rescued

through a truncated version of the protein, establishing that the cytoplasmic domain is

likely responsible for the Fat3 functionality during development (Matakatsu, 2006).

Table 1.2 summarizes the sequence similarity scores based on the UniProt sequence IDs

(Table 1.1) inputted into ClustalO.

      Mus musculus Fat2 and *Homo sapiens* Fat2 have similar scores, .08359 and

.08427, while *Drosophila melanogaster* varies. *Mus musclus* Fat3 and *Homo sapiens* Fat3

is also similar, .05261 and .055. This suggests that these Fat proteins are nearly identical

across mammals. However, not all Fat mammalian protein sequences have as similar

scores. Fat4 varies, .02428 and .03299. Both Fat proteins analyzed in *Drosophila*

melanogaster show similar scores, there is not much difference sequence-wise between

these two proteins, .36501 and .39144. So while Fat proteins among the same organism

may show some differences in sequence, there is also similarity among the proteins for

certain organisms. This again suggests that there was some divergence for the Fat protein

from *Drosophila melanogaster* to mammals.

      Next, I wanted to see whether there was a specific set of amino acids within the

cytoplasmic domain of the Fat3 protein that is conserved across all organisms. *Mus*

musculus, *Danio rerio*, and *Homo sapiens* each have a string of amino acids that is

mostly conserved across all three species. An asterisk (*) denotes a position where there

is a single, fully conserved residue. A colon (:) indicates conservation between groups of

strongly similar properties and a period (.) indicates conservation between groups of weakly similar properties. To better understand what is important about this particular set of amino acids, I looked in the literature for specific binding regions and annotations for that amino acid region. The conserved region (4353-4472 of the entire Fat3 protein in human), highlighted in Figure 2, has been annotated as being compositionally biased and proline rich. It is also annotated as a splicing variant. Proline-rich regions have been indicative of being important mediators for protein-protein interactions. EVH1 binding sites, which are found in Fat1, have been shown to be proline rich and critical for protein interaction of Fat1 and Ena/VASP (Moeller, 2004). Furthermore, Fat3 has been shown to have three EVH1 binding sites (Krol, 2016) located in this cytoplasmic domain. This data suggests that Ena/VASP binding to the Fat cadherins may be conserved across evolution. Because of this likely conservation, this interaction might be critical for Fat cadherin function.

I also compared the cytoplasmic domain of the *Homo sapiens* Fat3 protein to the cytoplasmic domain of both *Mus musculus* and *Rattus.* This allows for a deeper analysis of Fat3 in mammalian species (Figure 3, Table 1.3). The UniProt IDs found in Table 1.3 are the identifiers used for this multiple sequence alignment performed via Clustal Omega. Figure 3 outlines the similarity found for the Fat3 sequence across these species. Overall, these three sequences are highly similar, sharing 91.903% identity. Highlighted are regions where there are differences between the amino acids. These differences could provide basis for the functional differences that we see for Fat3 in each of these organisms. There is only one residue that varies between each of the 3 organisms and it maps to residue 4255 in human. At this position there is a glycine in human, a valine in

mouse, and an arginine in rat. Overall, amino acid conservation is strongest between mouse and rat, as expected.

Finally, I utilized the UCSC Genome Browser to examine Fat3 across a greater number or organisms (Figure 4). It appears that Fat3 remains relatively conserved in mammals. Elephant, dog, mouse, and monkey all share a large amount of conservation for the Fat3 gene. However, there is sequence divergence between mammals and chicken, mostly in the introns. Notice the left arrow referring to where this divergence begins. The second takeaway I have from Figure 4 is the difference in conservation between the LINE sequences. LINE stands for Long Interspersed Nuclear Element. The longest LINES in the human Fat3 sequence are basically aligned to where the gaps occur between the more conserved Fat3 sequences and human Fat3. This means that these LINE insertions arose in human. Long interspersed repeat elements are retrotransposons that allow movement of DNA along the genome. Zachary Burton suggests in his book "Evolution Since Coding" (Burton, 2018) that enhancer elements, which regulate gene transcription, are moved in evolution via transposable elements. This disrupted conservation might suggest different regulatory mechanisms for Fat3 in humans versus these other organisms. This difference in regulation would result in expression changes in Fat3 as well.

Allen Brain Atlas, Gene Expression Analysis

The Allen Brain Atlas is a genome-wide high-resolution atlas of gene expression throughout the adult mouse brain. A particular gene can be queried against this databases and expression data for that gene is shown upon that query. Using this system, I searched for each of the genes in the Wave Regulatory Complex (SRa1/Cyflip1, Nap1/Hem2/Kette, Abi2, and WAVE1/SCAR), scraped that data via Python (Figure S.1)

and compared that gene expression to Fat3 (Figure 5). The supplemental python code is written so that the same process can be applied for other genes of interest.

I first retrieve the xml version of the search query in the Allen Brain Atlas. I save that xml file locally and open it with Python. Then I input that file into a Python package called Beautiful Soup (Figure 5). This package allows you to webscrape a web page of interest and then pull information from that page with Python syntax. After uploading the xml query into Beautiful Soup I save all of the content from the query into a string which I call "content" (Figure 6). The Beautiful Soup package allows you to "prettify" the information from the xml file so that it is a readable, easy to understand format. From this "prettify-ing" the output I can see what information I want to pull, i.e. what lines will give me the expression data that I am interested in examining. I create two lists, one for the expression values and one for the brain structures (Figure 7). Each of these lists needs to be formatted so that they only contain values and acronyms so that I can utilize these lists to plot the information I am interested in. I remove the labelling "<acronym>" for each acronym in the brain structure list and I convert all the expression values to floats, or python identifiable numbers with decimal points, so that the integers can be easily plotted (Figure 8). The brain structures analyzed in the program are: Midbrain (MB), Isocortex, Olfactory (OLF), Hippocampal Formation (HPF), Cortical Subplate (CTXsp), Striatum (STR), Pallidum (PAL), Thalamus (TH), Hypothalamus (HY), Pons (P), Medulla (MY), and the Cerebellum (CB). At this point I also check that my acronym list and expression value list are in the same order, i.e. the first brain structure corresponds to the first expression data value. This is just to double check that the xml file did not mis-format during extraction. Finally, these two lists can be plotted in a bar graph so that the

18

data can be visualized (Figure 9). This same approach is used for each of the genes of interest, only uploading the xml file that corresponds to the query of the particular gene examined.

Figure 10 outlines each of the bar graphs generated. Firstly, it shows that Fat3 and Nap1 are the most similarly expressed genes from the comparisons. Through a literature search I have not seen a direct interaction between Nap1 and Fat3 described. Although, Fat3 interacts with the WRC it might be doing so through direct interaction with any one of the proteins in the complex, allowing indirect interaction with the other proteins of the complex. While this is not experimental data, we know that the WRC has a binding site for Fat3, but it has not yet been shown how the complex interacts with Fat3. This analysis suggests that there is reason to examine whether Fat3 and Nap1 directly interact as their expression data is so similar. Perhaps Fat3 binds to the WRC through Nap1.

Furthermore, through this analysis I have seen that Abi1, Brick1, and Wasf1 also have similar expression data. The confirmation of the WRC is still unknown, but this information might help to create model predictions for this complex. It might be that Abi1, Brik1, and Wasf1 are all complexed more closely together than the other WRC members. Finally, Fat1 and Fat4 show differential expression compared to Fat3. Fat4 has low expression throughout the brain while Fat1 shows similar expression to Fat3. The expression similarity between Fat3 and Fat1 follows their similar roles during development, both being involved in cytoskeletal and actin dynamics. Both also have EVH1 binding sites so it would make sense that these two genes are expressed in a similar fashion. The ABA is missing ISH data for Fat2, so it is not included in this analysis.

Chapter IV.

Discussion

While this two-dimensional approach is aimed at gathering information about this family of super cadherins, there are some obvious limitations to this type of analysis. The first part of the project has been informative to understand the evolutionarily conservation of Fats. The sequence analysis provides hints as to what gene regions have been deemed "important" and therefore conserved evolutionarily but this analysis is dependent on the open-source websites like NCBI and UnitProt. It is possible that these resources do not have thorough information about every single organism where Fats are actually found. These resources may not even have information about every single Fat family member.

The second part of the project has been informative in a different way, yielding information about gene expression of mammalian Fats and the Wave Regulatory Complex. It will be imperative to be careful about what conclusions are drawn from this analysis. While this provides information about Fat family members it does not answer questions about Fat function but instead yields a starting point for hypothesis.

This analytic approach to gaining information about a specific gene is informative but it is not an experimental approach which might yield more information about function. Instead, this is a useful dive that provides insight into what questions should be asked experimentally at the bench. Importantly, this provides an outline to examine other hypotheses about proteins of interest outside of the Fat family.
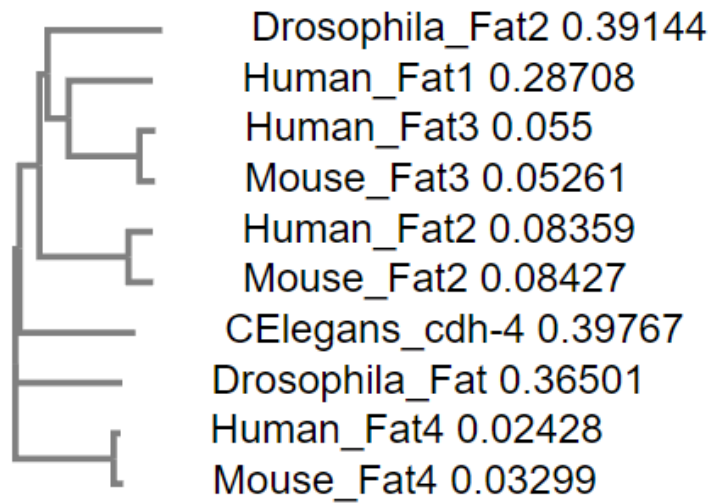
Figure 1. ClustalO Phylogentic Tree of Fat Protein sequences across Organisms.

Sequences used are in Table 1.

```
SP|Q8TDW7|FAT3_HUMAN    FQSDSGDDNAYHWDTSDWMPGARLSDIEEVPNYENQDGGSAHQGSTRELESDYYLGGYDI 4401

SP|Q8BNA6|FAT3_MOUSE    FQSDSGDDNAYHWDTSDWMPGARLSDIEEMPNYESQDGGAAHQGSTRELESDYYLGGYDI 4399

TR|F1QA09|F1QA09_DANRE  FQSDSCDDNAYHWDTSDWLPSTRLPDIEEVPTYETPEVGV-------ALESDYYLGGFDI 4423

                        ***** ************:*.:** ****:*.**. : *           *********:**



SP|Q8TDW7|FAT3_HUMAN    DSEYPPPHEEEFLSQDQLPPPLPEDFPDQYEALPPSQPVSLASTLSPDCRRRPQFHPSQY 4461

SP|Q8BNA6|FAT3_MOUSE    DSEYPPPHEEEFLSRDQLPPPLPEDFPDQYEALPPSQPTSLTSTMSPDCRRRPRFHPSQY 4459

TR|F1QA09|F1QA09_DANRE  DSDYPPPQEEEFMSQDQLPPPLPTE---EFDTSPANQPVSRESTLSGDRKPRPHFHPSQY 4480

                        **:****:****:*:********* :    :::: * .**.*  **:* * : **:******



SP|Q8TDW7|FAT3_HUMAN    LPPHPFPNETDLVGPPASCEFSTFAVSMNQGTEPTGPADSVSLSLHNSRGTSSSDVSANC 4521

SP|Q8BNA6|FAT3_MOUSE    LPPHPLPGETDLGGPSSSCDFSTFAVNMNQGTEVMAPTDSVSLSLHNSRGTSSSEMSARC 4519

TR|F1QA09|F1QA09_DANRE  LPPHQLPLGST------GEEFSTFTGRE----------ERLS----VNASSASSVSSAPC 4520

                        **** :* :       . :****:              : :*    . .::** ** *
```

Figure 2. Multiple Alignment of Human, Mouse, and Zebrafish Fat Sequences.

Highlighted is the conserved Ena/VASP interacting sequence.

```
SP|Q8TDW7|FAT3_Human   RKKVFRKNYSRNNITLVQDPATAAL 4200

SP|Q8BNA6|FAT3_MOUSE   RKKVFRKNYSRNNITLVQDPATAAL 4199

SP|Q8R508|FAT3_RAT     RKKVFRKNYSRNNITLVQDPATAAL 4199

                       *************************

SP|Q8TDW7|FAT3_HUMAN LNKSNGIPFRNLRGSGDGRNVYQEVGPPQVPVRPMAYTPCFQSDSRSNLDKIVDGLGGEH 4260

SP|Q8BNA6|FAT3_MOUSE LHKSNGIPFRSLR-AGDGRNVYQEVGPPQVPVRPMAYTPCFQSDSRSNLDKGLDVLGGEP 4258

SP|Q8R508|FAT3_RAT   LHKSNGIPFRSLR-AGDGRNVYQEVGPPQVPVRPMAYTPCFQSDSRSNLDKGLDALGGEP 4258

                     *:********.** :********************************** :* ****

SP|Q8TDW7|FAT3_HUMAN QEMTTFHPESPRILTARRGVVVCSVAPNLPAVSPCRSDCDSIRKNGWDAGTENKGVDDPG 4320

SP|Q8BNA6|FAT3_MOUSE QEMSTFHPESPRILTARRGVVVCSVAPNLPAVSPCRSDCDSIRKNGWDTGSENKGTEDTG 4318

SP|Q8R508|FAT3_RAT   QELSTFHPESPRILTARRGVVVCSVAPNLPAVSPCRSDCDSIRKNGWDTGSENKGAEDTG 4318

                     **::*********************************************:*:****.:* *

SP|Q8TDW7|FAT3_HUMAN EVTCFAGSNKGSNSEVQSLSSFQSDSGDDNAYHWDTSDWMPGARLSDIEEVPNYENQDGG 4380

SP|Q8BNA6|FAT3_MOUSE EVTCFTNSNKGSNSEVQSLSSFQSDSGDDNAYHWDTSDWMPGARLSDIEEMPNYESQDGG 4378

SP|Q8R508|FAT3_RAT   EVTCFANSNKGSNSEVQSLNSFQSDSGDDNAYHWDTSDWMPGARLSDIEEMPNYESQDGG 4378

                     *****:.*************.**************************:****.****

SP|Q8TDW7|FAT3_HUMAN SAHQGSTRELESDYYLGGYDIDSEYPPPHEEEFLSQDQLPPPLPEDFPDQYEALPPSQPV 4440

SP|Q8BNA6|FAT3_MOUSE AAHQGSTRELESDYYLGGYDIDSEYPPPHEEEFLSRDQLPPPLPEDFPDQYEALPPSQPT 4438

SP|Q8R508|FAT3_RAT   AVHQGSTRELESDYYLGGYDIDSEYPPPHEEEFLSQDQLPPPLPEDFPEQYEALPPSQPT 4438

                     :.********************************:***********:***********.

SP|Q8TDW7|FAT3_HUMAN SLASTLSPDCRRRPQFHPSQYLPPHPFPNETDLVGPPASCEFSTFAVSMNQGTEPTGPAD 4500

SP|Q8BNA6|FAT3_MOUSE SLTSTMSPDCRRRPRFHPSQYLPPHPLPGETDLGGPSSSCDFSTFAVNMNQGTEVMAPTD 4498

SP|Q8R508|FAT3_RAT   SLTGTMSPDCRRRPRFHPSQYLPPHPLPGETDLGGPSSSCDFSTFAVSMNQGTEVMAPTD 4498

                     **:.*:********:***********:*.**** ** :**:******.******  .*:*
```

23

```
SP|Q8TDW7|FAT3_HUMAN SVSLSLHNSRGTSSSDVSANCGFDDSEVAMSDYESVGELSLASLHIPFVETQHQTQV 4557


SP|Q8BNA6|FAT3_MOUSE SVSLSLHNSRGTSSSEMSARCGFDDSEVAMSDYESAGELSLTNLHIPFVETQQQTQV 4555


SP|Q8R508|FAT3_RAT   SVSLSLHNSRGTSSSDMSARCGFDDSEVAMSDYESAGELSLTNLHIPFVETQHQTQV 4555


                     ***************::**.***************.*****:.*********:****
```

Figure 3. Multiple Sequence Alignment of Fat3 between Mouse, Human, and Rat.

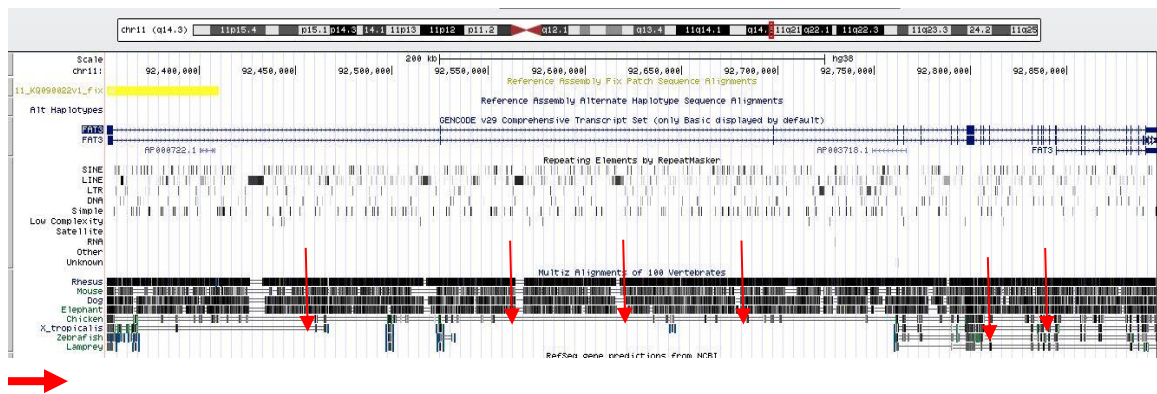*Highlighted are the residue differences between the sequences.*

Figure 4. UCSC Genome Browser Examination of Fat3.

Arrows are highlighting gaps between the sequences.

24

```
content = []

with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Fat3_Query.xml",

"r") a s file:

content = file.readlines()

#combine the lines in the list into a string content =

"".join(content)

bs_content = bs(content, "lxml")
```

Figure 5. Open xml File and Open it as a Single String, with Beautiful Soup Package

```
#content of webpage for Fat3 gene

#bs_content

#print(bs_content.prettify())
```

Figure 6. Make xml File Read-able

```
#create a list of the expression values

#create a list of the brain structures

expressions = bs_content.find_all("expression-energy") st =

bs_content.find_all("acronym")
```

Figure 7. Make a List of Expression Values and Brain Structures

```
#remove acronym bit from the list of structures so that I just have th

e structures itself.
```

```python
sts = []

for i in st:

    sts.append(i.text)

print(sts)
```

```
'''Now have a list of each of the brain structures'''

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P', 'P

AL', 'HPF', 'HY']
```

Out[8]: 'Now have a list of each of the brain structures'

In [9]: #check expression energy list expressions

Out[9]:

```
[<expression-energy>0.777972</expression-energy>,

 <expression-energy>3.77864</expression-energy>,

 <expression-energy>0.650517</expression-energy>,

 <expression-energy>1.51293</expression-energy>,

 <expression-energy>0.337382</expression-energy>,

 <expression-energy>0.782705</expression-energy>,

 <expression-energy>2.6014</expression-energy>,

 <expression-energy>1.74379</expression-energy>,

 <expression-energy>0.439608</expression-energy>,

 <expression-energy>0.665499</expression-energy>,
```

```
<expression-energy>1.62216</expression-energy>,

<expression-energy>0.485752</expression-energy>]


In [10]: expns = []

for i in expressions:

        expns.append(i.text)

    print(expns)

    ['0.777972', '3.77864', '0.650517', '1.51293', '0.337382',
'0.782705

    ', '2.6014', '1.74379', '0.439608', '0.665499', '1.62216',
'0.485752

    ']


In [11]: floatyexpn_Fat3 = [float(i) for i in expns]

print(floatyexpn_Fat3)

        [0.777972, 3.77864, 0.650517, 1.51293, 0.337382, 0.782705,
2.6014, 1

        .74379, 0.439608, 0.665499, 1.62216, 0.485752]
```

Figure 8. Format Expression Value List and Brain Structure List

```
#now plot this data

        '''THIS IS FOR FAT3'''
```

```
            # I want my x to be the structures and y to be the expn
values

            x_Fat3 = sts

            y_Fat3 = floatyexpn_Fat3

            plt.ylim(0,12)

            plt.bar(x_Fat3, y_Fat3)

Out[12]:  <BarContainer object of 12 artists>

In [12]:
```
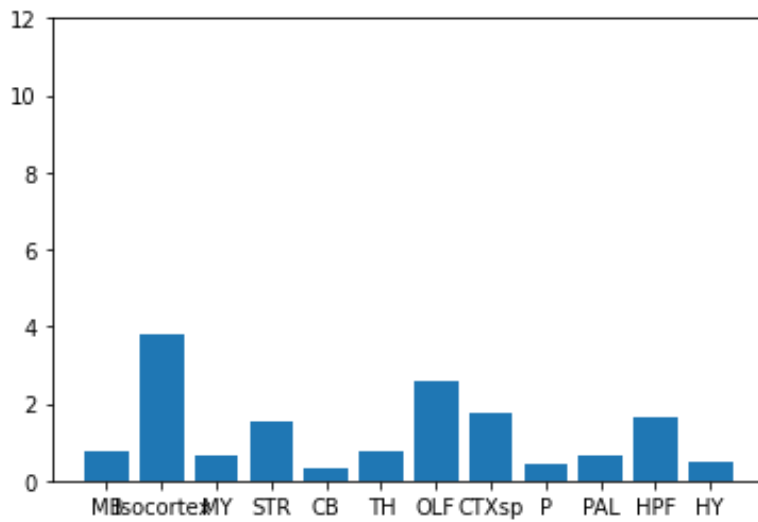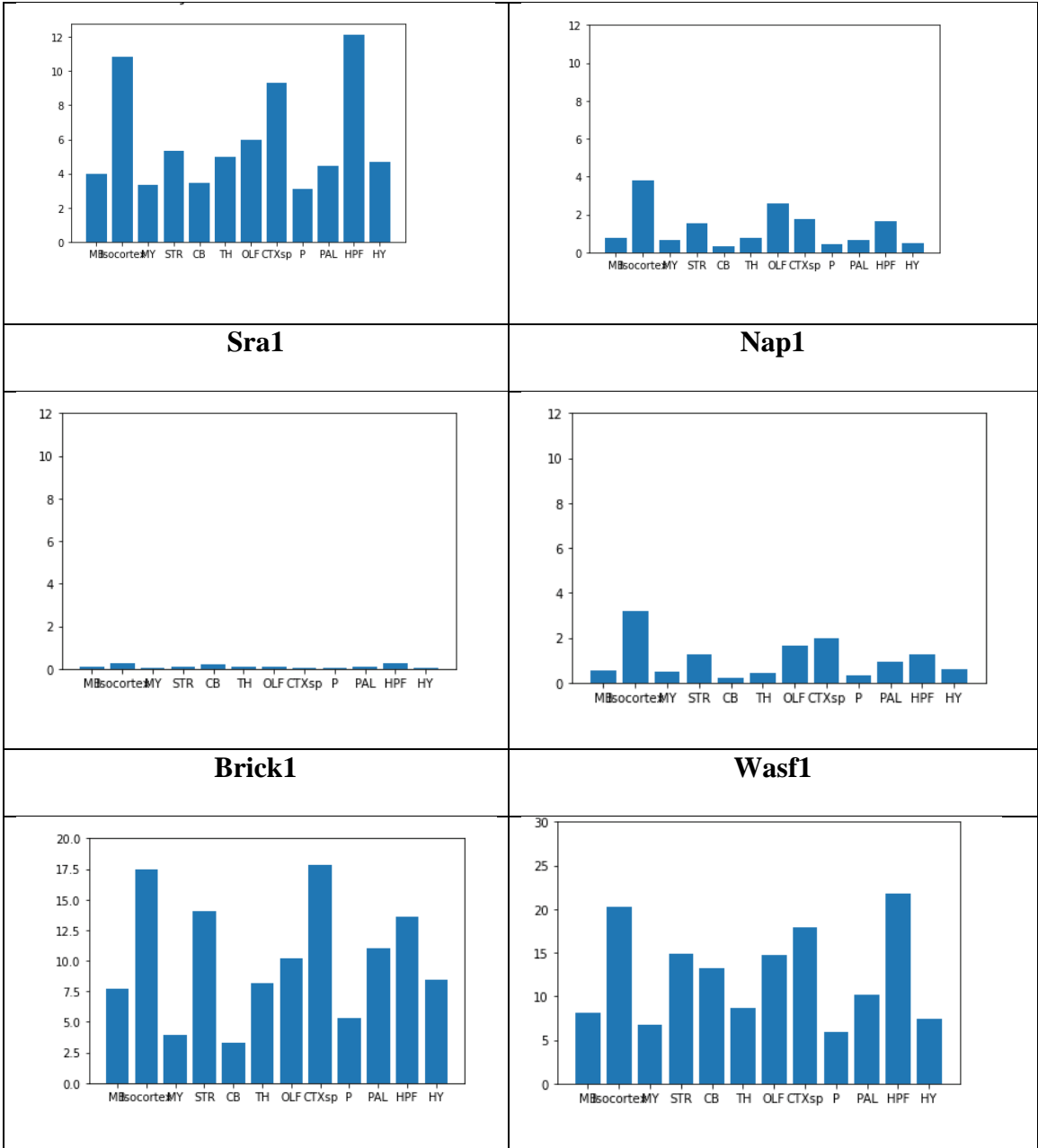

Figure 9. Plotting Expression Data and Brain Structures

Abi1	Fat3

Sra1

Nap1

Brick1

Wasf1

| Fat1 | Fat4 |

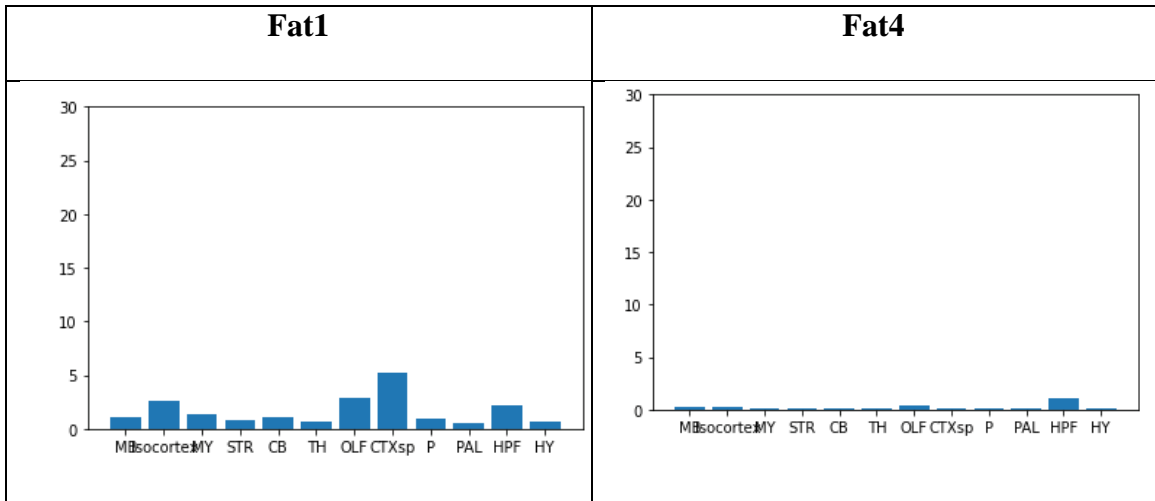Figure 10. Graphs of ABA Data for Various Proteins that are Part of the Fat Family or Part of the WRC.

Table 1.UniProt IDs for Fat Protein Sequences

Species	Fat cadherins	UniProt ID – Cytoplasmic Domain
Drosophila	Fat2 (kugelei or Fat-like)	Q9VW71: 4307-4699
	Fat (ft)	P33450: 4610-5147
Danio rerio	Fat3a	F1QA09 – full protein sequence, no annotated cytoplasmic domain
Homo sapiens	Fat1	Q14517: 4203-4588
	Fat2	Q9NYQ8: 4070-4349
	Fat3	Q8TDW7: 4176-4557
	Fat4	Q6V017: 4526-4981
C. elegans	*cdh-4*	Q19319: 4094-4328
Mous musculus	Fat1	Q80XT9, full protein sequence, no annotated cytoplasmic domain
	Fat2	Q5F226: 4072-4351
	Fat3	Q8BNA6: 4175-4555
	Fat4	Q2PZL6: 4527-4981

FASTA sequences are in Figure 2., Appendix V.

Table 2. Protein Sequence Alignment Score

Drosophila melanogaster		Homo Sapiens				Mus musculus			C. elegans
Fat	Fat2	Fat1	Fat2	Fat3	Fat4	Fat2	Fat3	Fat4	Cdh-4
.36501	.39144	.28708	.08359	.055	.02428	.08427	.05261	.03299	.39767

These scores are from ClustalO, using UniProt IDs from Table 1.

Table 3. UniProt IDs used for Protein Multiple Sequence Analysis

Organism	Gene	UniProt ID	Cytoplasmic Domain
Homo sapiens	Fat3	Q8TDW7	4176-4557
mouse	Fat3	Q8BNA6	4175-4555
Rattus norvegicus	Fat3	Q8R508	4175-4555
Danio rerio	Fat3a	F1QA09	Unannotated

Appendix III. Definition of Terms

1. Cadherin: cell adhesion molecule that contributes to cell-cell binding

2. Planar cell polarization: the intrinsic asymmetry of a cell

3. Fat cadherins: A particular family of cadherins

4. Cell autonomous communication: When a cell does not require any external signaling to act

5. Cell non-autonomous communication: Communication of a cell that is dependent on external signals

6. Wave regulatory complex: A five-subunit protein complex involved in the formation of the actin cytoskeleton

7. Python: A coding language

8. Amacrine cells: interneurons in the retina

9. GST pull down: An assay that uses a bait to capture a protein of interest and its interactors

10. Chrondocytes: Cartilage cells that produce and maintain the cartilage matrix.

11. Web Scraping: Python-mediated extraction of data from an html site.

Appendix IV. Python Code for ABA Web Scraping and Data Analysis

```
In [1]: '''Thesis scrap book - beautiful soup to get expression
data from ABA' ''

Out[1]: 'Thesis scrap book - beautiful soup to get expression
data from ABA'

In [2]:

#libraries that I need

from bs4 import BeautifulSoup as bs import requests
import lxml
import matplotlib.pyplot as plt import numpy as np

#webpage = requests.get("https://mouse.brain-
map.org/gene/show/13803")

#webtext = webpage.text

#webscraping the ABA page for Fat3

content = []
with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Fat3_Query.xml",
"r") a s file:

content = file.readlines()
#combine the lines in the list into a string content =
"".join(content)
bs_content = bs(content, "lxml")
```

```python
#content of webpage for Fat3 gene

#bs_content

#print(bs_content.prettify())

#create a list of the expression values
#create a list of the brain structures
expressions = bs_content.find_all("expression-energy") st =
bs_content.find_all("acronym")
```

In [7]: #check list st

      #remove fat3 from the list

```python
st.pop(0) st
```

Out[7]: [<acronym>MB</acronym>, <acronym>Isocortex</acronym>,

In [8]:

 <acronym>MY</acronym>,

 <acronym>STR</acronym>,

 <acronym>CB</acronym>,

 <acronym>TH</acronym>,

 <acronym>OLF</acronym>,

 <acronym>CTXsp</acronym>,

 <acronym>P</acronym>,

```
    <acronym>PAL</acronym>,

    <acronym>HPF</acronym>,

    <acronym>HY</acronym>]

#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts = []
for i in st:

    sts.append(i.text)

print(sts)



'''Now have a list of each of the brain structures'''

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']

Out[8]: 'Now have a list of each of the brain structures'

In [9]: #check expression energy list expressions

Out[9]:

[<expression-energy>0.777972</expression-energy>,

 <expression-energy>3.77864</expression-energy>,

 <expression-energy>0.650517</expression-energy>,
```

```
<expression-energy>1.51293</expression-energy>,

<expression-energy>0.337382</expression-energy>,

<expression-energy>0.782705</expression-energy>,

<expression-energy>2.6014</expression-energy>,

<expression-energy>1.74379</expression-energy>,

<expression-energy>0.439608</expression-energy>,

<expression-energy>0.665499</expression-energy>,

<expression-energy>1.62216</expression-energy>,

<expression-energy>0.485752</expression-energy>]

In [10]: expns = []
for i in expressions:

            expns.append(i.text)

        print(expns)

        ['0.777972', '3.77864', '0.650517', '1.51293',
'0.337382', '0.782705

        ', '2.6014', '1.74379', '0.439608', '0.665499',
'1.62216', '0.485752

        ']

In [11]: floatyexpn_Fat3 = [float(i) for i in expns]
print(floatyexpn_Fat3)
```

```
        [0.777972, 3.77864, 0.650517, 1.51293, 0.337382,
0.782705, 2.6014, 1

        .74379, 0.439608, 0.665499, 1.62216, 0.485752]


        #now plot this data

        '''THIS IS FOR FAT3'''

        # I want my x to be the structures and y to be the expn
values

        x_Fat3 = sts

        y_Fat3 = floatyexpn_Fat3

        plt.ylim(0,12)

        plt.bar(x_Fat3, y_Fat3)

Out[12]: <BarContainer object of 12 artists>

In [12]:
```
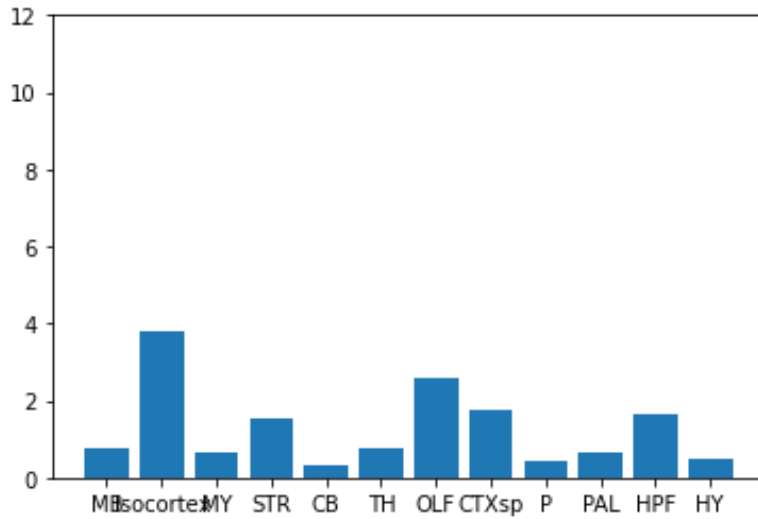
```
# Do the same for Abi1


content_Abi = []

with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Abi1_Query.xml",

"r") a s file_Abi:


content_Abi = file_Abi.readlines()

#combine hte lines in the list into a string content_Abi =

"".join(content_Abi) bs_content_Abi = bs(content_Abi, "lxml")


#content of webpage for Abi1 gene


#bs_content_Abi


#print(bs_content_Abi.prettify())


Out[19]:


#create a list of the expression values

#create a list of the brain structures
```

```python
expressions_Abi = bs_content_Abi.find_all("expression-energy")

st_Abi = bs_content_Abi.find_all("acronym")


#check list


st_Abi


#remove unwanted acronyms, that aren't structures


del st_Abi[0:6] st_Abi


[<acronym>MB</acronym>,

 <acronym>Isocortex</acronym>,

 <acronym>MY</acronym>,

 <acronym>STR</acronym>,

 <acronym>CB</acronym>,

 <acronym>TH</acronym>,

 <acronym>OLF</acronym>,

 <acronym>CTXsp</acronym>,

 <acronym>P</acronym>,

 <acronym>PAL</acronym>,

 <acronym>HPF</acronym>,

 <acronym>HY</acronym>]
```

```python
#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts_Abi1 = []
for i in st_Abi:

    sts_Abi1.append(i.text)

print(sts_Abi1)
```

```
['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']
```

```python
expns_Abi1 = []
for i in expressions_Abi:

    expns_Abi1.append(i.text)

print(expns_Abi1)
```

```
['3.96276', '10.8464', '3.32912', '5.33743', '3.43279', '4.9984',
'5

.94636', '9.32296', '3.12341', '4.43811', '12.1544', '4.65851']
```

```python
In [22]: floatyexpn = [float(i) for i in expns_Abi1]
print(floatyexpn)

In [23]:
```

```
[3.96276, 10.8464, 3.32912, 5.33743, 3.43279, 4.9984, 5.94636,
9.322

96, 3.12341, 4.43811, 12.1544, 4.65851]

#now plot this data

'''THIS IS FOR Abi1'''

# I want my x to be the structures and y to be the expn values

x_Abi1 = sts_Abi1

y_Abi1 = floatyexpn

plt.bar(x_Abi1, y_Abi1)
```
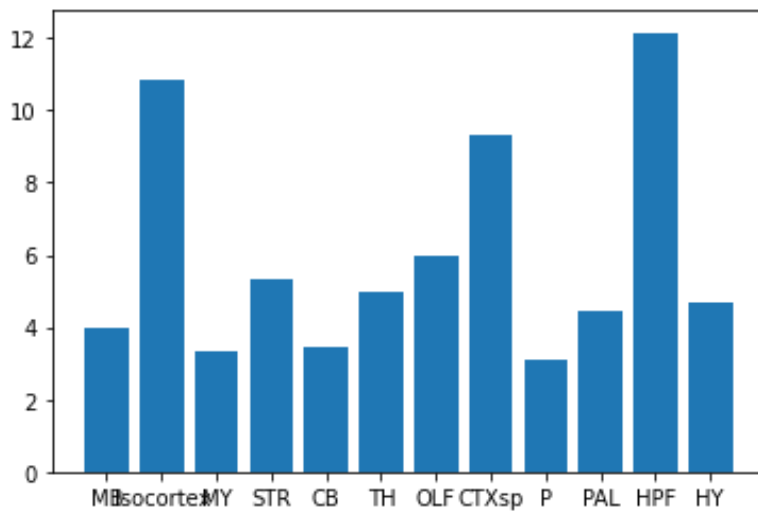
Out[23]: <BarContainer object of 12 artists>


```
#now do the same for Sra1

content_Sra1 = []
with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Sra1_Query.xml",
"r") as file_Sra1:
```

```python
content_Sra1 = file_Sra1.readlines() #combine hte lines in the
list into a string content_Sra1 = "".join(content_Sra1)
bs_content_Sra1 = bs(content_Sra1, "lxml")

#content of webpage for Sra1 gene

#bs_content_Sra1


#print(bs_content_Sra1.prettify())

#create a list of the expression values
#create a list of the brain structures
expressions_Sra1 = bs_content_Sra1.find_all("expression-energy")
st_Sra1 = bs_content_Sra1.find_all("acronym")

st_Sra1
```

Out[29]: [<acronym>Sra1</acronym>, <acronym>MB</acronym>,

         <acronym>Isocortex</acronym>,

         <acronym>MY</acronym>,

         <acronym>STR</acronym>,

         <acronym>CB</acronym>,

         <acronym>TH</acronym>,

         <acronym>OLF</acronym>,

         <acronym>CTXsp</acronym>,

```
        <acronym>P</acronym>,

        <acronym>PAL</acronym>,

        <acronym>HPF</acronym>,

        <acronym>HY</acronym>]

In [30]: st_Sra1.pop(0) st_Sra1

Out[30]: [<acronym>MB</acronym>, <acronym>Isocortex</acronym>,

        <acronym>MY</acronym>,

        <acronym>STR</acronym>,

        <acronym>CB</acronym>,

        <acronym>TH</acronym>,

        <acronym>OLF</acronym>,

        <acronym>CTXsp</acronym>,

        <acronym>P</acronym>,

        <acronym>PAL</acronym>,

        <acronym>HPF</acronym>,

        <acronym>HY</acronym>]

In [31]:

#remove acronym bit from the list of structures so that I just
have th e structures itself.
```

```python
sts_Sra1 = []

for i in st_Sra1:

    sts_Sra1.append(i.text)

print(sts_Sra1)

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']


In [32]: expns_Sra1 = []
for i in expressions_Sra1:

        expns_Sra1.append(i.text)

    print(expns_Sra1)

    ['0.0722852', '0.241825', '0.0662701', '0.104053',
'0.225495', '0.11

    9491', '0.0965946', '0.0698646', '0.0297136',
'0.117126', '0.285127'

    , '0.0429073']


In [33]: floatySra1=[float(i) for i in expns_Sra1]
print(floatySra1)

    [0.0722852, 0.241825, 0.0662701, 0.104053, 0.225495,
0.119491, 0.096
```
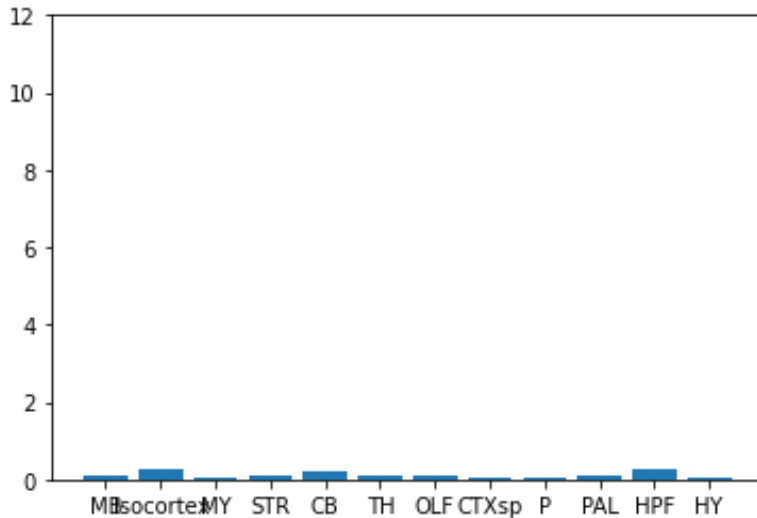
```
        5946, 0.0698646, 0.0297136, 0.117126, 0.285127,
0.0429073]
```

```python
#now plot this data

'''THIS IS FOR Sra1'''

# I want my x to be the structures and y to be the expn
values

x_Sra1 = sts_Sra1

y_Sra1 = floatySra1

plt.ylim(0,12)

plt.bar(x_Sra1, y_Sra1)
```

Out[34]: <BarContainer object of 12 artists>

In [34]:

```
#Now do the same for Nap1


content_Nap1 = []

with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Nap1_Query.xml",

"r") a s file_Nap1:


content_Nap1 = file_Nap1.readlines() #combine hte lines in the

list into a string content_Nap1 = "".join(content_Nap1)

bs_content_Nap1 = bs(content_Nap1, "lxml")


#content of webpage for Sra1 gene

#bs_content_Nap1

#print(bs_content_Nap1.prettify())


In [39]: #create a list of the expression values

#create a list of the brain structures

expressions_Nap1 = bs_content_Nap1.find_all("expression-energy")

st_Nap1 = bs_content_Nap1.find_all("acronym")
```

47

```
In [40]: st_Nap1

Out[40]: [<acronym>Nap1l1</acronym>, <acronym>MB</acronym>,

         <acronym>Isocortex</acronym>,

         <acronym>MY</acronym>,

         <acronym>STR</acronym>,

         <acronym>CB</acronym>,

         <acronym>TH</acronym>,

         <acronym>OLF</acronym>,

         <acronym>CTXsp</acronym>,

         <acronym>P</acronym>,

         <acronym>PAL</acronym>,

         <acronym>HPF</acronym>,

         <acronym>HY</acronym>]

In [41]: st_Nap1.pop(0)

Out[41]: <acronym>Nap1l1</acronym>

In [42]: st_Nap1

Out[42]: [<acronym>MB</acronym>, <acronym>Isocortex</acronym>,

         <acronym>MY</acronym>,

         <acronym>STR</acronym>,
```

```
        <acronym>CB</acronym>,

        <acronym>TH</acronym>,

        <acronym>OLF</acronym>,

        <acronym>CTXsp</acronym>,

        <acronym>P</acronym>,

        <acronym>PAL</acronym>,

        <acronym>HPF</acronym>,

        <acronym>HY</acronym>]
```

In [43]:

```python
#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts_Nap1 = []
for i in st_Nap1:

    sts_Nap1.append(i.text)

print(sts_Nap1)

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']
```

In [44]: expns_Nap1 = []
```python
for i in expressions_Nap1:
```

```
        expns_Nap1.append(i.text)

    print(expns_Nap1)

    ['0.531282', '3.19002', '0.461053', '1.27079',
'0.209521', '0.426722

    ', '1.65921', '1.96015', '0.296085', '0.930185',
'1.25497', '0.58129

    ']
```

```
In [45]: floatyNap1=[float(i) for i in expns_Nap1]
print(floatyNap1)

    [0.531282, 3.19002, 0.461053, 1.27079, 0.209521,
0.426722, 1.65921,

    1.96015, 0.296085, 0.930185, 1.25497, 0.58129]




    #now plot this data

    '''THIS IS FOR Nap1'''

    # I want my x to be the structures and y to be the expn
values

    x_Nap1 = sts_Nap1

    y_Nap1 = floatyNap1
```
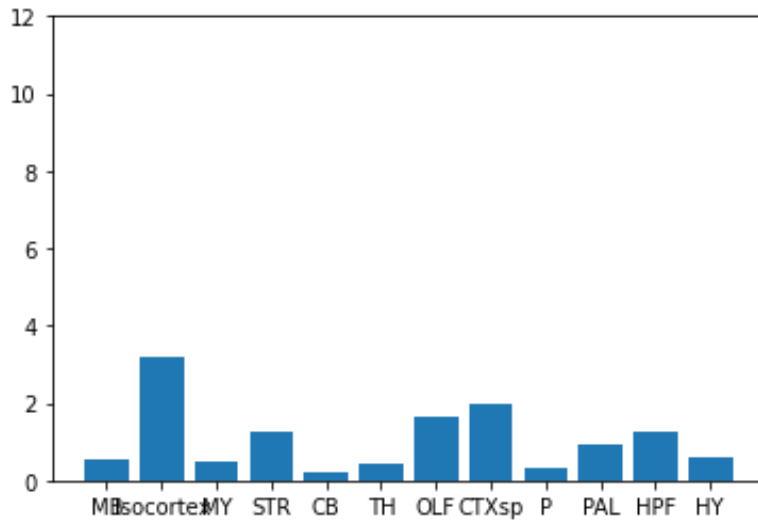
```
        plt.ylim(0,12)

        plt.bar(x_Nap1, y_Nap1)
```

Out[46]: <BarContainer object of 12 artists>

In [46]:


```
#Now do the same for Brick1

content_Brick1 = []
with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Brick1_Query.xml",
"r") as file_Brick1:

content_Brick1 = file_Brick1.readlines() #combine hte lines in
the list into a string content_Brick1 = "".join(content_Brick1)
bs_content_Brick1 = bs(content_Brick1, "lxml")

#content of webpage for Brick1 gene
```

```
#bs_content_Brick1

#print(bs_content_Brick1.prettify())
```

In [51]: *#create a list of the expression values*
#create a list of the brain structures
expressions_Brick1 = bs_content_Brick1.find_all("expression-energy") st_Brick1 = bs_content_Brick1.find_all("acronym")

In [52]: st_Brick1

Out[52]: [<acronym>Brk1</acronym>, <acronym>MB</acronym>,

<acronym>Isocortex</acronym>,

<acronym>MY</acronym>,

<acronym>STR</acronym>,

<acronym>CB</acronym>,

<acronym>TH</acronym>,

<acronym>OLF</acronym>,

<acronym>CTXsp</acronym>,

<acronym>P</acronym>,

<acronym>PAL</acronym>,

<acronym>HPF</acronym>,

<acronym>HY</acronym>]

```
In [53]: st_Brick1.pop(0) st_Brick1

Out[53]: [<acronym>MB</acronym>, <acronym>Isocortex</acronym>,

In [54]:

 <acronym>MY</acronym>,

 <acronym>STR</acronym>,

 <acronym>CB</acronym>,

 <acronym>TH</acronym>,

 <acronym>OLF</acronym>,

 <acronym>CTXsp</acronym>,

 <acronym>P</acronym>,

 <acronym>PAL</acronym>,

 <acronym>HPF</acronym>,

 <acronym>HY</acronym>]

#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts_Brick1 = []
for i in st_Brick1:

    sts_Brick1.append(i.text)

print(sts_Brick1)
```

```
['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']
```

```
In [55]: expns_Brick1 = []
for i in expressions_Brick1:

        expns_Brick1.append(i.text)

    print(expns_Brick1)

    ['7.7077', '17.4711', '3.92244', '14.0374', '3.31354',
'8.17503', '1

    0.1753', '17.844', '5.36064', '10.9981', '13.5612',
'8.43138']
```

```
In [56]: floatyBrick1=[float(i) for i in expns_Brick1]
print(floatyBrick1)

    [7.7077, 17.4711, 3.92244, 14.0374, 3.31354, 8.17503,
10.1753, 17.84

    4, 5.36064, 10.9981, 13.5612, 8.43138]
```

```
In [57]: #now plot this data '''THIS IS FOR Brick1'''

Out[57]:

# I want my x to be the structures and y to be the expn values
```

```
x_Brick1 = sts_Brick1

y_Brick1 = floatyBrick1

plt.ylim(0,20)

plt.bar(x_Brick1, y_Brick1)

<BarContainer object of 12 artists>
```
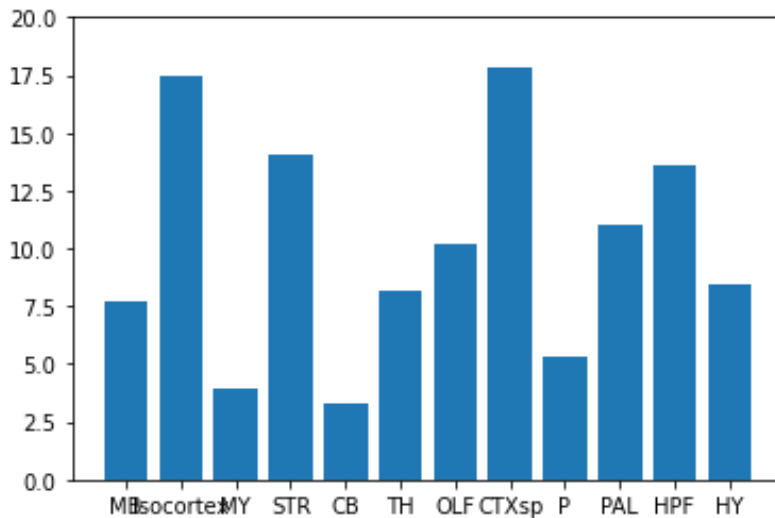


```
content_Wasf1 = []
```

with open("/Users/sadie/Desktop/ABR_DATA_LOCAL/Wasf1_Query.xml",

"r") **as** file_Wasf1:

content_Wasf1 = file_Wasf1.readlines() *#combine hte lines in the*

list into a string content_Wasf1 = "".join(content_Wasf1)

bs_content_Wasf1 = bs(content_Wasf1, "lxml")

#content of webpage for Brick1 gene

```python
#bs_content_Wasf1

#print(bs_content_Wasf1.prettify())

#create a list of the expression values
#create a list of the brain structures
expressions_Wasf1 = bs_content_Wasf1.find_all("expression-
energy") st_Wasf1 = bs_content_Wasf1.find_all("acronym")

st_Wasf1
```

Out[62]: [<acronym>Wasf1</acronym>, <acronym>AI838537</acronym>,

<acronym>Scar</acronym>,

<acronym>AI195380</acronym>,

<acronym>WAVE</acronym>,

<acronym>WAVE-1</acronym>,

<acronym>Wasf1</acronym>,

<acronym>MB</acronym>,

<acronym>Isocortex</acronym>,

<acronym>MY</acronym>,

<acronym>STR</acronym>,

<acronym>CB</acronym>,

<acronym>TH</acronym>,

<acronym>OLF</acronym>,

```
            <acronym>CTXsp</acronym>,

            <acronym>P</acronym>,

            <acronym>PAL</acronym>,

            <acronym>HPF</acronym>,

            <acronym>HY</acronym>]
```

Out[63]:

```python
del st_Wasf1[0:7] st_Wasf1
```

```
[<acronym>MB</acronym>,

 <acronym>Isocortex</acronym>,

 <acronym>MY</acronym>,

 <acronym>STR</acronym>,

 <acronym>CB</acronym>,

 <acronym>TH</acronym>,

 <acronym>OLF</acronym>,

 <acronym>CTXsp</acronym>,

 <acronym>P</acronym>,

 <acronym>PAL</acronym>,

 <acronym>HPF</acronym>,

 <acronym>HY</acronym>]
```

```python
#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts_Wasf1 = []
for i in st_Wasf1:

    sts_Wasf1.append(i.text)

print(sts_Wasf1)

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']

expns_Wasf1 = []
for i in expressions_Wasf1:

    expns_Wasf1.append(i.text)

print(expns_Wasf1)

['8.11857', '20.195', '6.79377', '14.8574', '13.1552', '8.61113',
'1

4.7324', '17.8426', '5.91061', '10.1982', '21.762', '7.40604']

floatyWasf1=[float(i) for i in expns_Wasf1] print(floatyWasf1)

[8.11857, 20.195, 6.79377, 14.8574, 13.1552, 8.61113, 14.7324,
17.84

26, 5.91061, 10.1982, 21.762, 7.40604]
```

```python
#now plot this data

'''THIS IS FOR Wasf1'''

# I want my x to be the structures and y to be the expn
values

x_Wasf1 = sts_Wasf1

y_Wasf1 = floatyWasf1

plt.ylim(0,30)

plt.bar(x_Wasf1, y_Wasf1)
```
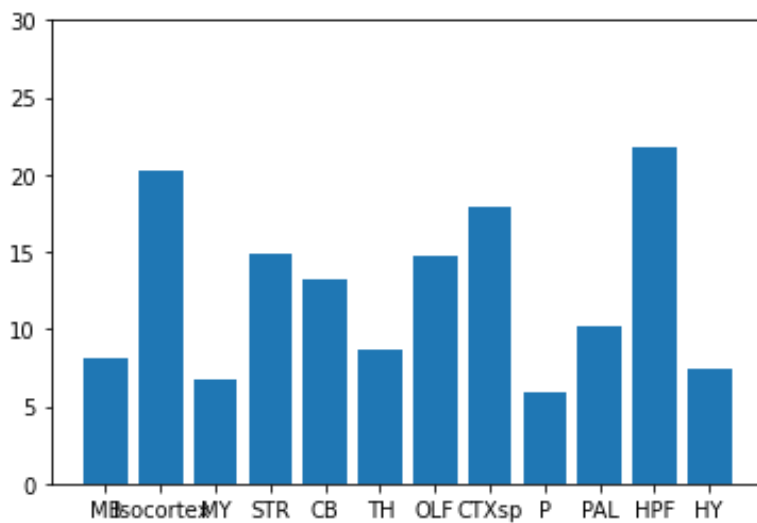
Out[67]: <BarContainer object of 12 artists>

In [67]:

In [68]:

```python
#This is for Fat1

content_Fat1 = []
with open("/Users/sadie/Desktop/Thesis/Fat1_Query.xml", "r") as
file_F at1:

content_Fat1 = file_Fat1.readlines() #combine hte lines in the
list into a string content_Fat1 = "".join(content_Fat1)
bs_content_Fat1 = bs(content_Fat1, "lxml")

#bs_content_Fat1

#print(bs_content_Fat1.prettify())

#create a list of the expression values
#create a list of the brain structures
expressions_Fat1 = bs_content_Fat1.find_all("expression-energy")
st_Fat1 = bs_content_Fat1.find_all("acronym")

In [72]: st_Fat1

Out[72]: [<acronym>Fat1</acronym>, <acronym>MB</acronym>,

          <acronym>Isocortex</acronym>,

          <acronym>MY</acronym>,

          <acronym>STR</acronym>,

          <acronym>CB</acronym>,

          <acronym>TH</acronym>,
```

60

```
                <acronym>OLF</acronym>,

                <acronym>CTXsp</acronym>,

                <acronym>P</acronym>,

                <acronym>PAL</acronym>,

                <acronym>HPF</acronym>,

                <acronym>HY</acronym>]

In [73]: st_Fat1.pop(0) st_Fat1

Out[73]: [<acronym>MB</acronym>, <acronym>Isocortex</acronym>,

In [74]:

 <acronym>MY</acronym>,

 <acronym>STR</acronym>,

 <acronym>CB</acronym>,

 <acronym>TH</acronym>,

 <acronym>OLF</acronym>,

 <acronym>CTXsp</acronym>,

 <acronym>P</acronym>,

 <acronym>PAL</acronym>,

 <acronym>HPF</acronym>,

 <acronym>HY</acronym>]
```

61

```python
#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts_Fat1 = []
for i in st_Fat1:

    sts_Fat1.append(i.text)

print(sts_Fat1)

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']

In [75]: expns_Fat1 = []
for i in expressions_Fat1:

        expns_Fat1.append(i.text)

    print(expns_Fat1)

    ['1.02698', '2.53271', '1.32583', '0.795676', '1.09813',
'0.661212',

    '2.80749', '5.26663', '1.00005', '0.474208', '2.2293',
'0.667607']

In [76]: floatyFat1=[float(i) for i in expns_Fat1]
print(floatyFat1)

    [1.02698, 2.53271, 1.32583, 0.795676, 1.09813, 0.661212,
2.80749, 5.
```

26663, 1.00005, 0.474208, 2.2293, 0.667607]

In [77]: *#now plot this data '''THIS IS FOR Fat1'''*

Out[77]:

 # I want my x to be the structures and y to be the expn values

x_Fat1 = sts_Fat1
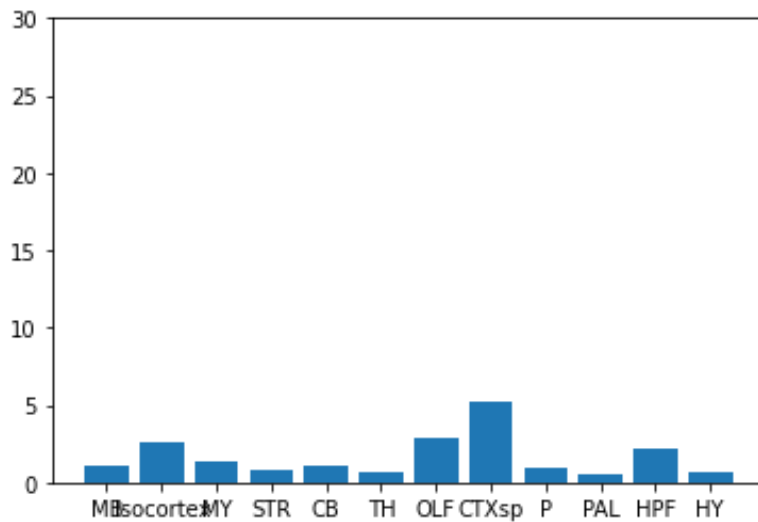
y_Fat1 = floatyFat1

plt.ylim(0,30)

plt.bar(x_Fat1, y_Fat1)

<BarContainer object of 12 artists>

In [78]:

#This is for Fat4

```python
content_Fat4 = []

with open("/Users/sadie/Desktop/Thesis/Fat4_Query.xml", "r") as
file_F at4:

content_Fat4 = file_Fat4.readlines() #combine hte lines in the
list into a string content_Fat4 = "".join(content_Fat4)
bs_content_Fat4 = bs(content_Fat4, "lxml")

#bs_content_Fat4

#print(bs_content_Fat4.prettify())

#create a list of the expression values
#create a list of the brain structures
expressions_Fat4 = bs_content_Fat4.find_all("expression-energy")
st_Fat4 = bs_content_Fat4.find_all("acronym")

st_Fat4

Out[83]: [<acronym>Fat4</acronym>, <acronym>MB</acronym>,

         <acronym>Isocortex</acronym>,

         <acronym>MY</acronym>,

         <acronym>STR</acronym>,

         <acronym>CB</acronym>,

         <acronym>TH</acronym>,

         <acronym>OLF</acronym>,
```

<acronym>CTXsp</acronym>,

<acronym>P</acronym>,

<acronym>PAL</acronym>,

<acronym>HPF</acronym>,

<acronym>HY</acronym>]

In [84]:

Out[84]:

```
st_Fat4.pop(0)

st_Fat4
```

[<acronym>MB</acronym>,

 <acronym>Isocortex</acronym>,

 <acronym>MY</acronym>,

 <acronym>STR</acronym>,

 <acronym>CB</acronym>,

 <acronym>TH</acronym>,

 <acronym>OLF</acronym>,

 <acronym>CTXsp</acronym>,

 <acronym>P</acronym>,

 <acronym>PAL</acronym>,

```
  <acronym>HPF</acronym>,

  <acronym>HY</acronym>]

#remove acronym bit from the list of structures so that I just
have th e structures itself.
sts_Fat4 = []
for i in st_Fat4:

    sts_Fat4.append(i.text)

print(sts_Fat4)

['MB', 'Isocortex', 'MY', 'STR', 'CB', 'TH', 'OLF', 'CTXsp', 'P',
'P

AL', 'HPF', 'HY']


expns_Fat4 = []
for i in expressions_Fat4:

    expns_Fat4.append(i.text)

print(expns_Fat4)

['0.215767', '0.276095', '0.0632197', '0.132598', '0.107561',
'0.156

018', '0.388578', '0.164246', '0.122053', '0.0559038', '1.10856',
'0

.113369']


floatyFat4=[float(i) for i in expns_Fat4] print(floatyFat4)
```

[0.215767, 0.276095, 0.0632197, 0.132598, 0.107561, 0.156018,

0.3885

78, 0.164246, 0.122053, 0.0559038, 1.10856, 0.113369]

In [88]: *#now plot this data '''THIS IS FOR Fat4'''*

 # I want my x to be the structures and y to be the expn
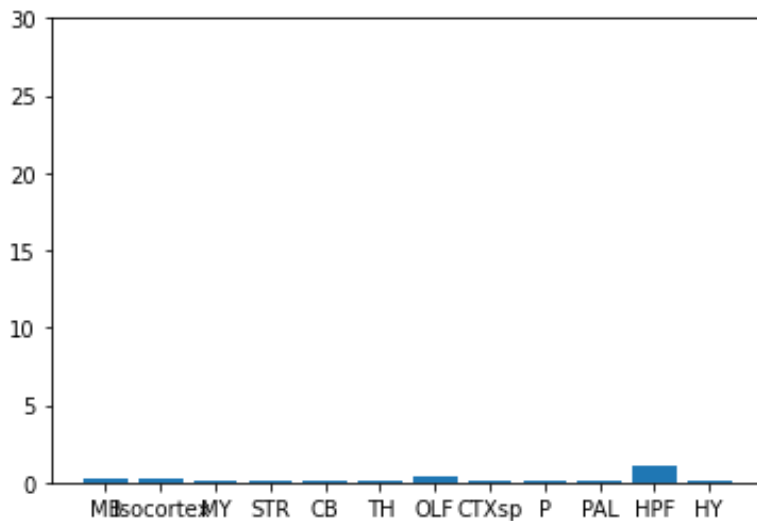
values

```
x_Fat4 = sts_Fat4

y_Fat4 = floatyFat4

plt.ylim(0,30)

plt.bar(x_Fat4, y_Fat4)
```

Out[88]: <BarContainer object of 12 artists>

In [ ]:

Appendix V. UniProt ID, FASTA Sequences for Phylogenetic Tree and Sequence

Analysis

>sp|Q9VW71|4307-4699: Fat2 Drosophila melanogaster

KRSSSSKRRNRLEKDKNKSSYKEANLNSLVDKDNYCKPNVKLSNLEVNQRPISYTAVPND

NLVLSNRNFVNNLDILRSYGSAGDELENVPFEYQKVNRNKQHVNINSCHSTDADNAYKQE

WCEQMHLRTFSENKLNNELKRDFGPSVSRFSTGKLIQVEMPNVCHSSSANFVDYSALANG

QYHWDCSDWVRKSHNPLPDITEVPGAEIADSSSLHSNDSNESKSKKAFFVHREDGDVDPT

RDIAALNEDIGSEYLDSEAESCLEPFMLPRSSNQPLSRLSSFNNIENEDYKSNTVPLPSK

VSHSCKVYLRHPDSYLPTMHFPSETDGESSMTEGPISRMEIKTRRTISENSEEAYLFPCT

VGEIGSNSNISVRLCEIEDSELEEFLPQQQTNN

>sp|P33450|4610-5147: Fat, Drosophila melanogaster

RFRGKQEKIGSLSCGVPGFKIKHPGGPVTQSQVDHVLVRNLHPSEAPSPPVGAGDHMRPP

VGSHHLVGPELLTKKFKEPTAEMPQPQQQQQRPQRPDIIERESPLIREDHHLPIPPLHPL

PLEHASSVDMGSEYPEHYDLENASSIAPSDIDIVYHYKGYREAAGLRKYKASVPPVSAYT

HHKHQNSGSQQQQQQHRHTAPFVTRNQGGQPPPPPTSASRTHQSTPLARLSPSSELSSQQ

PRILTLHDISGKPLQSALLATTSSSGGVGKDVHSNSERSLNSPVMSQLSGQSSSASRQKP

GVPQQQAQQTSMGLTAEEIERLNGRPRTCSLISTLDAVSSSSEAPRVSSSALHMSLGGDV

DAHSSTSTDESGNDSFTCSEIEYDNNSLSGDGKYSTSKSLLDGRSPVSRALSGGETSRNP

PTTVVKTPPIPPHAYDGFESSFRGSLSTLVASDDDIANHLSGIYRKANGAASPSATTLGW

EYLLNWGPSYENLMGVFKDIAELPDTNGPSQQQQQQTQVVSTLRMPSSNGPAAPEEYV

>sp|Q14517|4203-4588: Fat1, Homo Sapiens

RKMISRKKKHQAEPKDKHLGPATAFLQRPYFDSKLNKNIYSDIPPQVPVRPISYTPSIPS

DSRNNLDRNSFEGSAIPEHPEFSTFNPESVHGHRKAVAVCSVAPNLPPPPPSNSPSDSDS

IQKPSWDFDYDTKVVDLDPCLSKKPLEEKPSQPYSARESLSEVQSLSSFQSESCDDNGYH

WDTSDWMPSVPLPDIQEFPNYEVIDEQTPLYSADPNAIDTDYYPGGYDIESDFPPPPEDF

PAADELPPLPPEFSNQFESIHPPRDMPAAGSLGSSSRNRQRFNLNQYLPNFYPLDMSEPQ

TKGTGENSTCREPHAPYPPGYQRHFEAPAVESMPMSVYASTASCSDVSACCEVESEVMMS

DYESGDDGHFEEVTIPPLDSQQHTEV

>sp|Q9NYQ8|4070-4349: Fat2, Homo Sapiens

CRRCKSHKPVAMEDPDLLARSVGVDTQAMPAIELNPLSASSCNNLNQPEPSKASVPNELV

TFGPNSKQRPVVCSVPPRLPPAAVPSHSDNEPVIKRTWSSEEMVYPGGAMVWPPTYSRNE

RWEYPHSEVTQGPLPPSAHRHSTPVVMPEPNGLYGGFPFPLEMENKRAPLPPRYSNQNLE

DLMPSRPPSPRERLVAPCLNEYTAISYYHSQFRQGGGGGPCLADGGYKGVGMRLSRAGPSY

AVCEVEGAPLAGQGQPRVPPNYEGSDMVESDYGSCEEVMF

>sp|Q8TDW7|4176-4557: Fat3, Homo sapiens

RKKVFRKNYSRNNITLVQDPATAALLNKSNGIPFRNLRGSGDGRNVYQEVGPPQVPVRPM

AYTPCFQSDSRSNLDKIVDGLGGEHQEMTTFHPESPRILTARRGVVVCSVAPNLPAVSPC

RSDCDSIRKNGWDAGTENKGVDDPGEVTCFAGSNKGSNSEVQSLSSFQSDSGDDNAYHWD

TSDWMPGARLSDIEEVPNYENQDGGSAHQGSTRELESDYYLGGYDIDSEYPPPHEEEFLS

QDQLPPPLPEDFPDQYEALPPSQPVSLASTLSPDCRRRPQFHPSQYLPPHPFPNETDLVG

PPASCEFSTFAVSMNQGTEPTGPADSVSLSLHNSRGTSSSDVSANCGFDDSEVAMSDYES

VGELSLASLHIPFVETQHQTQV

>sp|Q6V0I7|4526-4981: Fat4, Homo Sapiens

CNQCRGKKAKNPKEEKKPKEKKKKGSENVAFDDPDNIPPYGDDMTVRKQPEGNPKPDIIE

RENPYLIYDETDIPHNSETIPSAPLASPEQEIEHYDIDNASSIAPSDADIIQHYKQFRSH

TPKFSIQRHSPLGFARQSPMPLGASSLTYQPSYGQGLRTSSLSHSACPTPNPLSRHSPAP

FSKSSTFYRNSPARELHLPIRDGNTLEMHGDTCQPGIFNYATRLGRRSKSPQAMASHGSR

PGSRLKQPIGQIPLESSPPVGLSIEEVERLNTPRPRNPSICSADHGRSSSEEDCRRPLSR

TRNPADGIPAPESSSDSDSHESFTCSEMEYDREKPMVYTSRMPKLSQVNESDADDEDNYG

ARLKPRRYHGRRAEGGPVGTQAAAPGTADNTLPMKLGQQAGTFNWDNLLNWGPGFGHYVD

VFKDLASLPEKAAANEEGKAGTTKPVPKDGEAEQYV

>sp|Q19319|4094-4328: cdh-4, C. elegans

WKCCKKNRDPKYGAHCDVPHMRNTRVLVPVVPPPLPPRGFRNDSSNFISTSSVTTSHRPM

VQVKPYSSDIRDSRSPSACGSSKGTRRDPLPSDKFRRVDETANRIRHSDRKDPRGDVLSS

LRDSSDEWMGIDDRIDSSLKYSRAAAGTVIVGDTELMPVINDNDYMTMKPRKDKNFEREK

PPAIPAHATPLESVLKLGSSSSGEEAPRNALYDDPISLDSQTFDDIDEEVNIHIS

>sp|Q5F226|4072-4351: Fat2, Mus musculus

CRRCKSHKPVAMEDPDLLARSIGVDTQASPAIELDPLNAGSCNDLNQLEPSKTSVPNELV

TFGPSSKQRPMVCSVPPRLPPAVVSSHPGHEPIIKRTWSGEELVYPSGAAVWPPTYSRKE

HWEYPHPEAMQGPLPPSPRRHVSPAVMPDPAGLYGGFPFPLELENKRAPLPPRYSNQNLE

DLIPPRPPSPREHLLAPCLNEYTAISYYHSQFRQGGGGGPCLAEGGYKGVSMRLSRAGPSY

ADCEVNGGPAPGRSQPRAPPNYEGSDMVESDYGSCEEVMF

>sp|Q8BNA6|4175-4555: Fat3, Mus musculus

RKKVFRKNYSRNNITLVQDPATAALLHKSNGIPFRSLRAGDGRNVYQEVGPPQVPVRPMA

YTPCFQSDSRSNLDKGLDVLGGEPQEMSTFHPESPRILTARRGVVVCSVAPNLPAVSPCR

SDCDSIRKNGWDTGSENKGTEDTGEVTCFTNSNKGSNSEVQSLSSFQSDSGDDNAYHWDT

SDWMPGARLSDIEEMPNYESQDGGAAHQGSTRELESDYYLGGYDIDSEYPPPHEEEFLSR

DQLPPPLPEDFPDQYEALPPSQPTSLTSTMSPDCRRRPRFHPSQYLPPHPLPGETDLGGP

SSSCDFSTFAVNMNQGTEVMAPTDSVSLSLHNSRGTSSSEMSARCGFDDSEVAMSDYESA

GELSLTNLHIPFVETQQQTQV

>sp|Q2PZL6|4527-4981: Fat4, Mus musculus

CNQCRGKMPKNPKEEKKPKEKKKKGSENVAFDDPDNIPPYGDDLAVRKQPEGNPKPDIIE

RENPYLIFDETDIPHNSETIPSAPLASPEQEIEHYDIDNASSIAPSDADIIQHYKQFRSH

TPKFSIQRHSPLGFARQSPMPLGASSLTYQPSSYGQGLRTSSLSHSACPTPNPLSRHSPA

PFSKPSAFYRNSPARELHLPLRDGGTLEMHGDPCQPGMFNYATRLGRRSKSPQAMASHGS

RPGSRLKQPIAQIPLESSPPVGLSIEEVERLNTPRPRNPSICSADHGRSSSEEDCRRPLS

RTRNPADGIPAPESSSDSDSHDSFTCSEMEYDREKPVVYTSRMPKLSQVNESDADDEDNY

GARLKPRRYHGRRAEGGPVGTPAAASGAADSTLKLGQQAGNFNWDNLLNWGPGFGHYVDV

FKDLASLPEKAAGNEEGKSGAAKPAAKDGEAEQYV

References

Aviles, Evelyn, et al. (2017). Configuring a Robust Nervous System with Fat Cadherins. *Seminars in Cell & Developmental Biology,* 69: 91-101.

Badouel, Caroline, et al. (2015). Fat1 Interacts with Fat4 to Regulate Neural Tube Closure, Neural Progenitor Proliferation and Apical Constriction During Mouse Brain Development. *Development,* 142(16): 2781-2791.

Bioinformatics Tools FAQ. 2019. https://www.ebi.ac.uk/seqdb/confluence/display/JDSAT/Bioinformatics+Tools+FAQ#BioinformaticsToolsFAQ-Whatarethemeaningofthescoresshowninaphylogenetictree.

Brain Map - brain-map.org. (n.d.). Retrieved November 11, 2020, from https://portal.brain-map.org/

Burton F. Zach. (2018). *Evolution Since Coding: Cradles, Halos, Barrels, and Wings.* London, UK: Elsevier.

Caruso, Nathalie, et al. (2013). Deregulation of the Protocadherin Gene Fat1 Alters Muscle Shapes: Implications for the Pathogenesis of Facioscapulohumeral Dystrophy. *PLOS Genetics*, 9(6): 1-29.

Chen, Baoyu, et al. (2014). The WAVE Regulatory Complex Links Diverse Receptors to the Actin Cytoskeleton. *Cell*, 156: 195-207.

Chen, Zhucheng, et al. (2010). Structure and control of the actin regulatory WAVE complex. *Nature*, 468: 533-540.

Dang, Chinh, et al. (2013). Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Research,* 41: D996-D1008.

Deans, Michael R, et al. (2011). Control of Neuronal Morphology by the Atypical Cadherin Fat3. *Neuron,* 71 5: 820-832.

Hitoshi Matakatsu and Seth S. Blair. (2006). Separating the adhesive and signaling functions of the Fat and Dachsous protocadherins. *Development*, 133(12): 2315-24.

Horne-Badovinac, Sally. (2017). Fat-like cadherins in cell migration – leading from both the front and the back. *Current opinions in Cell Biology,* 48: 26-32.

Ismaail Sahin Gul, et al. (2017). Evolution and diversity of cadherins and catenins. *Experimental Cell Research,* 358(1).

Katoh, Yuriko, et al. (2006). Comparative Integromics on Fat1, Fat2, Fat3, and Fat4. *The International Journal of Molecular Medicine,* 18 (3):523-528.

Katoh, Masaru. (2012). Function and cancer genomics of FAT family genes. *International Journal of Oncology*, 41(6): 1913-1918.

Krol, Alexandra, et al. (2016). Fat3 and Ena/VASP proteins influence the emergence of asymmetric cell morphology in the developing retina. *Development,* 143: 2172-2182.

Le Pabic, Pierre, et al. (2014). Fat-Dachsous signaling coordinates cartilage differentiation and polarity during craniofacial development. *PLoS genetics*, 10: 1-14.

Liscio, Nicole Theresa. (2014). Investigating Novel Interactors of Mammalian Fat and Dachsous cadherins. [Master's thesis, University of Toronto and Lunenfeld-Tanenbaum Research Institute].

Moeller, Marcus, J., et al. (2004). Protocadherin Fat1 binds Ena/VASP proteins and is necessary for atin dynamics and cell polarization. *The EMBO Journal*, 23: 3769-3779.

Mohr: Flybase. 1927 http://flybase.org.

Nick J. Van Hateren, et al. 2011. FatJ acts via the Hippo mediator Yap1 to restrict the size of neural progenitor cell pools. *Development*, 138(10): 1893-1902.

Pabic, Pierre Le, et al. (2014). Fat-Dachsous Signaling Coordinates Cartilage Differentiation and Polarity during Craniofacial Development. *PLOS Genetics,* 10: 1-14.

Schmitz, Caroline, et al. (2008). The Fat-like cadherin CDH-4 controls axon fasiculation, cell hypodermis, and pharynx development in Caenorhabditis elegans. *Developmental Biology,* 316: 249-259.

Sopko, Richelle, et al. (2009). The skinny on fat: an enormous cadherin that regulates cell adhesion, tissue growth, and planar cell polarity. *Current Opinion in Cell Biology,* 21: 717-723.

Squarr, Anna Julia, et al. (2016) Fat2 acts through the WAVE regulatory complex to drive collective cell migration during tissue rotation. *The Journal of Cell Biology,* 212 5: 591-603.