

How Cytoscouts Got Shot out of Dr. Ritz's New Interactome

---

A Thesis  
Presented to  
The Division of Mathematics and Natural Sciences  
Reed College

---

In Partial Fulfillment  
of the Requirements for the Degree  
Bachelor of Arts

---

Petra Wijngaard

May 2018



Approved for the Division  
(Biology)

---

Anna Ritz



# Acknowledgements

I would have gotten nowhere without the patience, acceptance and guidance of so many people. In particular:

The two professors whom I owe my progress to, Anna, whose great teaching opened up the world of computer programming to me, and Derek, my *de facto* second adviser, who made biology accessible and interesting to me again. I will treasure the memories I made with you both.

My Mom and my Dad who have been patient, loving and understanding. I am so lucky to have you as my family.

Alex Mitchell, my best friend in the whole wide world.

My friends, Alystair, Aviva, Gwen, Jesse, Rose, Roxy and Saxton, who made the year so much easier.

Thank you all so much.



# Table of Contents

<b>Introduction</b> . . . . .	<b>1</b>
<b>Chapter 1: Background</b> . . . . .	<b>3</b>
1.1 The Cytoskeleton . . . . .	3
1.1.1 Actin . . . . .	3
1.1.2 Microtubules . . . . .	4
1.2 Spectraplakins . . . . .	5
1.2.1 Regulation of Spectraplakins . . . . .	9
1.3 Cell Migration . . . . .	10
1.3.1 Spectraplakins and Migration . . . . .	12
1.4 Networks . . . . .	14
1.5 Protein-Protein Interactomes . . . . .	14
1.6 Python Concepts . . . . .	16
1.7 Bioinformatics Tools and Resources . . . . .	18
<b>Chapter 2: Materials and Methods</b> . . . . .	<b>21</b>
2.1 The Eclectus Fly Interactome . . . . .	21
2.1.1 The Concept of Collapsing the Interactome . . . . .	22
2.2 Cytoscouts . . . . .	24
2.2.1 The Uncollapsed Interactome . . . . .	25
2.2.2 Collapsing the Interactome . . . . .	29
<b>Chapter 3: Results and Discussion</b> . . . . .	<b>33</b>
3.1 The Uncollapsed Interactome . . . . .	33
3.1.1 Shot in the Uncollapsed Interactome . . . . .	34
3.2 The Collapsed Interactome . . . . .	40
3.2.1 Shot in the Collapsed Interactome . . . . .	40
3.3 The Gene Ontology . . . . .	43
3.4 Discussion . . . . .	50
<b>Conclusion</b> . . . . .	<b>53</b>
<b>Appendix A: RhoGTPases</b> . . . . .	<b>55</b>
<b>Appendix B: Common Name Protein Description</b> . . . . .	<b>57</b>

<b>Appendix C: Gene Ontology Definitions</b>	<b>69</b>
<b>Appendix D: PANTHER Results</b>	<b>73</b>
<b>References</b>	<b>89</b>

# List of Tables

1.1	Spectraplakin functional domain names and roles. . . . .	6
1.2	An alternative representation of the network in Figure 1.6. . . . .	16
1.3	An alternative representation of the network in Figure 1.6 as a list of lists. . . . .	17
1.4	An alternative representation of the network in Figure 1.6 as a set of tuples. . . . .	18
2.1	Interactome sources with edge counts. . . . .	21
2.2	A table representation of the collapsing method in Figure 2.1. . . . .	24
3.1	The 32 uncharacterized nodes that interact with Shot A. . . . .	35
3.2	Top ten UniProt IDs that interact with Shot A, A1Z9J1, by node degree.	36
3.3	Shot isoforms in the uncollapsed interactome. . . . .	38
3.4	Top ten common name neighbor nodes of Shot by node degree. . . . .	41
3.5	The common names of Shot's neighbors. . . . .	41
3.6	The common names of the 118 proteins neighboring Shot recognized by PANTHER. . . . .	44
3.7	The 97 UniProt IDs that were unable to be classified by PANTHER.	45
3.8	PANTHER overrepresentation test for biological process. . . . .	46
3.9	PANTHER overrepresentation test for molecular function. . . . .	47
3.10	PANTHER classified these Neighbors of Shot A as transcription factors.	47
B.1	Descriptions of proteins from the literature. . . . .	57
C.1	Gene Ontology definitons, protein class. . . . .	70
C.2	Gene Ontology definitons, molecular function. . . . .	71
C.3	Gene Ontology definitons, biological process. . . . .	72
D.1	Abridged PANTHER results. . . . .	73



# List of Figures

1.1	The cytoskeleton, taken from Currie <i>et al.</i>	4
1.2	Spectraplakin schematic.	6
1.3	Autoinhibition model taken from Applewhite <i>et al.</i> (2013)	8
1.4	A migrating cell.	11
1.5	Spectraplakins in cell migration.	13
1.6	An example undirected network with nine nodes.	15
1.7	A GraphSpace network.	20
2.1	Collapsing the network.	23
2.2	A screenshot of cytoscouts.	24
2.3	A flow chart of processing the uncollapsed interactome.	26
2.4	A flow chart of processing the collapsed interactome.	29
3.1	Interactome histograms.	34
3.2	Histograms of Shot's neighbors.	39
3.3	A PathLinker diagram of Shot.	42
3.4	PANTHER Gene Ontology results for protein class.	48
3.5	PANTHER Gene Ontology results for molecular function.	49



# List of Functions

2.1	<code>importCSV</code> creates a set of nodes and a list of edges from a user supplied .csv file.	27
2.2	<code>getDegree</code> creates a dictionary of node:degree pairs.	28
2.3	<code>neighbors</code> creates a set of nodes that share an edge with a specified node.	28
2.4	<code>collapseInteractome</code> creates a set of tuples from the list of list <code>edgeList</code> and the dictionary <code>reference</code> , skipping edges that are not defined in <code>reference</code> and excluding duplicates that are tuples in reverse order.	31
2.5	<code>collapseNodes</code> creates a set of common names from the set <code>nodeSet</code> and the dictionary <code>reference</code> , skipping names that do not have a definition in <code>reference</code> .	31



# Abstract

Protein-protein interaction networks, here referred to as interactomes, are an increasingly used tool in systems biology. Interactomes have applications in basic and applied science. The size and complexity of interactomes mean they are not human-readable in their unmodified form. I wrote a Python script called cytoscouts for reading interactomes and finding “neighbors” in the interactome, proteins that have a documented interaction with each other, and saving a list of those neighbors and a summary statistic of the neighbors’ neighbors to an output file for further analysis. The script also is capable of “collapsing” an interactome to remove redundant interactome entries, which are prevalent. I tested cytoscouts on a *Drosophila melanogaster* interactome and specifically on the protein Shot, a crosslinker between microtubules and Actin. My script is generalizable to any protein in any interactome stored in the common comma separated value format, meaning it may be useful to researchers for browsing interactome data.



# Dedication

Dedicated to my Mom.



# Introduction

Cell migration is critical to life, disease and death. From a humble amoeba to a deadly cancer cell, movement is an essential biological process. The protein Shot drastically impacts the way the fruit fly *Drosophila melanogaster* cells migrate. When Shot is removed from *D. melanogaster* cells, the cells move faster, flying in the face of what we have come to expect from experiments with similar proteins in other organisms. This thesis is my attempt to learn more about the ways Shot makes cells move.

Shot, also known as Short stop or Kakapo, is a protein in the fruit fly *D. melanogaster*. Shot is a member of the spectraplakin family, and it links the proteins of Actin and microtubules together. Actin and microtubules are cytoskeletal proteins, meaning they form the structure that gives the cell its form and rigidity. Actin, microtubules and spectraplakins all play roles in the way cells move.

In unpublished research, Applewhite *et al.* observed that removing Shot from *D. melanogaster* cells caused them to migrate faster, which is the opposite of the observations in other organisms that spectraplakins are needed for cell migration. While it is only one of many possible explanations, I wondered if Shot somehow acted as an inhibiting regulator on the process of cell migration or if it interacted with something that did. While not a particularly rigorous hypothesis, it was a question that could be investigated using computational biology with a protein-protein interactome.

A protein-protein interactome is a collection of all of the observed pairwise interactions between proteins in a given biological subject, as well as potential interactions that can be inferred from experiments in other organisms. It is like a map of documented proteins in an organism, except instead of geographic proximity the map shows which proteins have physical interactions. There is a lot of information in an interactome though, and reading through it all is a challenge. The interactome I worked with is a plain text file thirty-six megabytes in size with over half a million lines of text.

Because it was impossible for me to read through all of that manually, I made a tool cytoscouts to make it easier to interpret the information contained in a protein-protein interactome. Cytoscouts is a script written in the Python programming language which makes it easier to find and characterize the proteins that interact with a target protein. Finding which proteins that Shot interacted with would tell me if there were proteins that could possibly link Shot to cell migration.

Cytoscouts is a script that finds the proteins that interact with each other, the neighbors in the interactome. It can work with any interactome in the commonly used comma separated value, .csv, format, and it works with any user-supplied pro-

tein ID contained in the user-supplied interactome. Once supplied with an ID and an interactome cytoscouts produces a text file containing a list of all the other IDs that are neighbors with the user-specified ID in the interactome, along with a summary statistic of those other IDs' neighbors too. The text file can then be imported into Excel, R, or an online bioinformatics tool to be further analyzed. In addition, cytoscouts saves a visual representation of the results.

I used cytoscouts on a new *D. melanogaster* interactome, the Eclectus Fly Interactome. The Eclectus Fly Interactome is named after the Eclectus parrot and for its “eclectic” union of six separate *D. melanogaster* interactomes, each with a different set of protein interactions. The interactome stores each protein as a Universal Protein Resource, or UniProt, code. The UniProt database is an online resource of protein sequences and annotations. UniProt entries are highly specific, but this trait combined with the fact that UniProt automatically sources its entries means that UniProt can generate redundant entries for the same protein from different sources.

Because of this, cytoscouts also has an optional feature, “collapse,” meant to deal with redundant entries in interactomes. Collapsing the interactome takes every UniProt node that shares a common name and combines them. Doing this sacrifices potential information, but it eliminates redundant entries. Collapsing is an option that may improve the results’ accuracy if the uncollapsed results contain too many redundant entries.

My thesis is about how I built cytoscouts and how I used it on a protein-protein interactome to learn more about Shot. Once cytoscouts produced the list of proteins that interact with Shot, I used the publicly available bioinformatics toolset PANTHER to characterize the proteins in Shot’s list of neighbors. My hope is that the information gained through this thesis on the interactions of Shot and other proteins can be used to further investigate these interactions in the laboratory.

The thesis starts with a background section on Shot and what it is and does, along with background information on interactomes as well as some bioinformatics tools. After the background, I discuss the Eclectus interactome and the functions that make up cytoscouts. Then, I talk about my results and interpret them before providing concluding thoughts.

# Chapter 1

## Background

This section reviews topics relevant to the *D. melanogaster* spectraplakin Shot. It covers the two cytoskeleton proteins that spectraplakins link together, Actin and microtubules. Spectraplakins themselves and their structure and interaction are covered next. I give a brief overview of cellular migration and some findings from the literature about spectraplakins' role in migration. Since this thesis is about Shot in the interactome, I introduce networks and interactomes. Then I talk about the Python concepts needed to process interactomes. Finally, I cover some bioinformatics tools that helped me interpret the interactome.

### 1.1 The Cytoskeleton

The cytoskeleton is a structure within the cell that organizes the cell, gives it rigidity and provides a surface for internal mechanisms to work upon (Fletcher and Mullins, 2010). Alongside a host of other proteins, the cytoskeleton is comprised of three main protein families: actin, microtubules and intermediate filaments. Of these, Actin and microtubules are most relevant to Shot, because Shot links Actin and microtubules together. Figure 1.1 shows the cytoskeleton of a *D. melanogaster* cell and the Actin and microtubule structures that comprise it.

#### 1.1.1 Actin

The Actin cytoskeleton is a polymer structure used by nearly all eukaryotes to give the cell structure and the ability to exert force (Gunning et al., 2015). Actin does more than just organize the cell. Actin holds protein complexes together and it is important in cell division (Stoddard et al., 2017). Proteins travel inside the cell along the structures created by Actin (Pollard and Cooper, 2009). These are just a few of the many roles of Actin structures. I will discuss Actin structures later in this thesis, but for now I focus on Actin as a protein.

Actin as a protein is a molecule that works as a subunit in larger structures. On its own it is known as globular actin, or G-Actin, but it is called filamentous Actin, or F-Actin, when polymerized into the structures that make up the cytoskeleton. F-

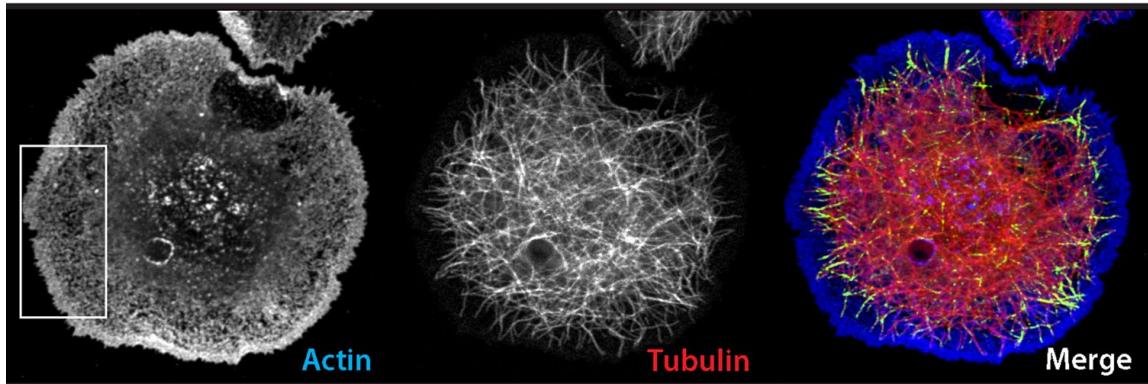


Figure 1.1: This figure, taken with permission from Currie *et al.*, depicts the cytoskeleton of a *D. melanogaster* S2 cell (2011). Immunostained Tubulin highlights the microtubule cytoskeleton, and immunostained Actin highlights the Actin cytoskeleton.

Actin is polarized polymer with a plus end and a minus end. While F-actin elongates on both plus and minus ends, Actin polymerization occurs ten times faster at the plus end than it does at the minus end (Blanchoin *et al.*, 2014). Regulatory proteins localize to the plus end of the Actin polymer.

### 1.1.2 Microtubules

Microtubules fill many of the same roles Actin does in the cytoskeleton. They organize the cell, facilitate transport of cell contents and enable movement (Alfaro-Aco and Petry, 2015). They also give cells polarity, a front and a back (Etienne-Manneville, 2013). Microtubules are a critical part of the cytoskeleton.

Microtubules are rigid hollow tube-like structures that have a diameter of 25 nm but can extend to hundreds of micrometers. Microtubules form at microtubules organizing centers, which in most animal cells are single anchor points called centrosomes (Kollman *et al.*, 2011). Exceptions to this, where microtubules do not grow from the centrosome and are called non-centrosomal, include muscle cells, neural cells, and epithelial cells (Bartolini and Gundersen, 2006). Certain *D. melanogaster* cell lines are non-centrosomal during interphase, including the cell lines used in Dr. Derek Applewhite's lab (Baumbach *et al.*, 2015; Debec *et al.*, 1995; Rogers *et al.*, 2008). Microtubules are constructed from heterodimer subunits, meaning that the subunits are made up of pairs of different proteins: one  $\alpha$ -tubulin molecule and one  $\beta$ -tubulin molecule. The tubulin dimers bind head-to-tail to create protofilaments that have a plus end and a minus end. Multiple protofilaments attach laterally to themselves to create a sheet that then rolls over onto itself into a cylinder. The microtubule is formed as more tubulin heterodimers are added the plus end of the protofilament cylinder (Etienne-Manneville, 2013).

The plus end of the microtubule is where most microtubule regulation takes place. (Etienne-Manneville, 2013; Fletcher and Mullins, 2010). The proteins that regulate microtubules at their heads are called plus end tracking protein (+TIPs),

which localize proteins from elsewhere in the cell to the microtubule head. These proteins microtubule associated proteins brought in by the +TIPs can extend, contract, protect, destroy, or otherwise modify the microtubule structure (Alfaro-Aco and Petry, 2015). Among these +TIPs is end-binding protein 1, known as EB1. EB1 localizes to the plus ends of growing microtubules and recruits other +TIPs to the growing microtubule (Fong et al., 2017). Shot is localized to the growing microtubule by EB1 (Applewhite et al., 2010).

Microtubules are the routes upon which vesicle transport operates (Etienne-Manneville, 2013; Fletcher and Mullins, 2010). Vesicles are membrane-bound packages containing proteins and chemicals to be sent around the cell. Thus, microtubules are required to regulate many cell processes, since regulating proteins cannot reach their targets without microtubules. In addition to delivering signaling proteins via vesicles, microtubules interact with signaling proteins directly (Etienne-Manneville, 2013).

## 1.2 Spectraplakins

Spectraplakins link Actin and microtubules together. In addition to that, spectraplakins are versatile proteins that both comprise and regulate the cytoskeleton. This thesis focuses on the two mammalian spectraplakins ACF7, as known as MACF1, and BPAG1, also known as Dystonin, and the sole *D. melanogaster* spectraplakin Shot. The focus is more on Shot and ACF7 because Shot is more similar to ACF7 than it is to BPAG1. At least 24 variants, called isoforms, of Shot are known to exist. This thesis focuses on the Shot A isoform because it is the most commonly referenced isoform in the literature (Voelzmann et al., 2017).

Spectraplakins are large, flexible, two-headed, rod-like proteins up to 9,000 amino acids long that can potentially extend for 0.4  $\mu\text{m}$ , longer than some bacteria, inside a cell (Roepert et al., 2002). These proteins are involved in cell migration, wound healing, eye development, epithelial development, and neuron development; spectraplakins are involved in muscles, tendons and oocytes among many other biological structures and processes (Lee and Kolodziej, 2002; Voelzmann et al., 2017; Zhang et al., 2017). Spectraplakins gain their size and versatility from their functional domains.

In this thesis, I describe a composite representative spectraplakin whose attributes are taken from the literature. Spectraplakin functional domains are evolutionarily conserved, and their roles are broadly applicable across organisms (Voelzmann et al., 2017). Please see Figure 1.2 for a schematic of a representative spectraplakin and Table 1.1 for a brief overview of its functional domains.

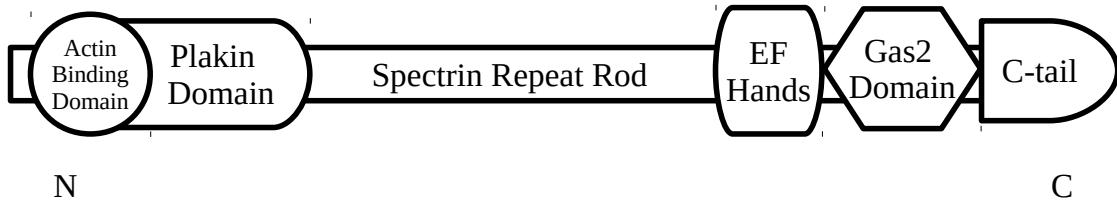


Figure 1.2: A schematic depiction, not to scale, of a generic spectraplakin depicting its major conserved functional domains.

Table 1.1: Spectraplakin functional domain names and roles.

Name	Role
Actin-binding domain	Binds actin, target of phosphorylation
Plakin domain	Thought to interact with transmembrane proteins
Spectrin repeat rod	Elastic spacer, bends for autoinhibition
EF hand motif	Inhibits Actin and microtubule binding
Gas2-related domain	Associates with microtubules
Ctail	Associates with EB1 to track microtuble plus ends, target of phosphorylation

## Actin Binding Domain

Starting from the N terminus, the “head,” of the protein, the first domain is the actin binding domain. The Actin binding domains of Shot and ACF7 are around 90-70% similar to each other (Voelzmann et al., 2017). The Actin binding domain is comprised of two calponin homology domains, CH1 and CH2. Calponin homology domains are found in other Actin binding proteins, and they enable the spectraplakin to associate directly with Actin (Zhang et al., 2017). There are isoforms where one or both of the calponin homology domains are missing, and Actin binding affinity is different depending on whether CH1 or CH2 was removed (Voelzmann et al., 2017).

## Plakin Domain

Next along the protein is the plakin domain. The plakin domains of Shot and ACF7 are around 30-75% similar to each other (Voelzmann et al., 2017). If the plakin domain is removed from Shot, motor neurons fail to localize the neural cell adhesion molecule Fascilin2 as if the cells had no Shot at all (Bottenberg et al., 2009). In BPAG1, the plakin domain interacts with the transmembrane protein integrin in a yeast two-hybrid system (Koster et al., 2003). The literature suggests that the plakin domain interacts with transmembrane proteins.

Certain spectraplakin isoforms contain an additional functional domain called the plakin repeat region. In isoforms that contain one, the plakin repeat region is after the plakin domain (Voelzmann et al., 2017). The plakin repeat region binds intermediate filaments, but *D. melanogaster* has no cytoplasmic intermediate filaments and little is currently known about its function in Shot isoforms beyond that it localizes to the aherens junction and may be involved in the nervous system (Voelzmann et al., 2017).

## Spectrin Repeat Rod

After the plakin domain is the spectrin repeat rod. The spectrin repeat rods of Shot and ACF7 are around 50-60% similar to each other (Voelzmann et al., 2017). About one-third of a spectraplakin's amino acids are made up by the spectrin repeat rod (Voelzmann et al., 2017; Zhang et al., 2017). The spectrin repeat rod comprises the midsection of the spectraplakin and puts space between the head of the protein and its tail. This section is elastic and made up of coiled subsegments that give the spectraplakin its stretchiness. The spectrin repeat rod is flexible enough that spectraplakins are able to double over themselves (Applewhite et al., 2013). This ability to double over will be discussed further when we talk about the regulation of spectraplakins.

ACF7 binds on the minus end of microtubules in a complex with the CAMSAP3 protein, via an interaction on its spectrin repeat rod (Ning et al., 2016). The homolog, the equivalent gene, of CAMSAP3 in *D. melanogaster* is Patronin. Patronin and Shot, in a complex with Katanin, are required for microtubule organization in epithelial follicle cells (Nashchekin et al., 2016).

## EF Hands

Traveling further along the length of the spectraplakin, we encounter next two EF hand domains. The EF hand domains of Shot and ACF7 are around 75-80% similar to each other (Voelzmann et al., 2017). These domains bind to the Actin binding domain and cause the spectraplakin to double over. This conformation is thought to prevent the Actin binding domain from binding to Actin and the Gas2 domain from binding to microtubules, as the autoinhibitory characteristic is absent when the EF hands are removed (Applewhite et al., 2013). Removing the EF hands from BPAG1 caused similar lack of inhibition (Kapur et al., 2012). In Shot, the EF hands are a binding site for the translational regulator Krasavietz (Lee et al., 2007).

EF hand domains are found in proteins that bind calcium in the form of  $\text{Ca}^{2+}$ .

Intracellular  $\text{Ca}^{2+}$  regulates BPAG1 by causing it to bind to microtubules (Kapur et al., 2012). However, in Shot, EF hands did not act as calcium sensors nor did calcium cause Shot to bind microtubules (Applewhite et al., 2013).

## Gas2 Domain

Further down the spectraplakin lies the Gas2 domain. The Gas2 domains of Shot, ACF7 and BPAG1 are around 80% similar to each other (Voelzmann et al., 2017). In ACF7 and in Shot, the Gas2 domain works with the C-tail to bind to the microtubule lattice, as removing either greatly reduces microtubule association (Sun et al., 2001; Applewhite et al., 2010).

## C-tail

The C-tail, as the name suggests, forms the tail of the spectraplakin at its C terminus. The C-tails of Shot and ACF7 are around 60% similar to each other (Voelzmann et al., 2017). In Shot, the C-tail is necessary and sufficient for Shot to track the plus ends of microtubules. Removing it from Shot removes plus end tracking and adding it to an unrelated protein causes the protein to track the plus end (Applewhite et al., 2010). Applewhite *et al.* demonstrated that plus end tracking is achieved through coordination with EB1 (2010).

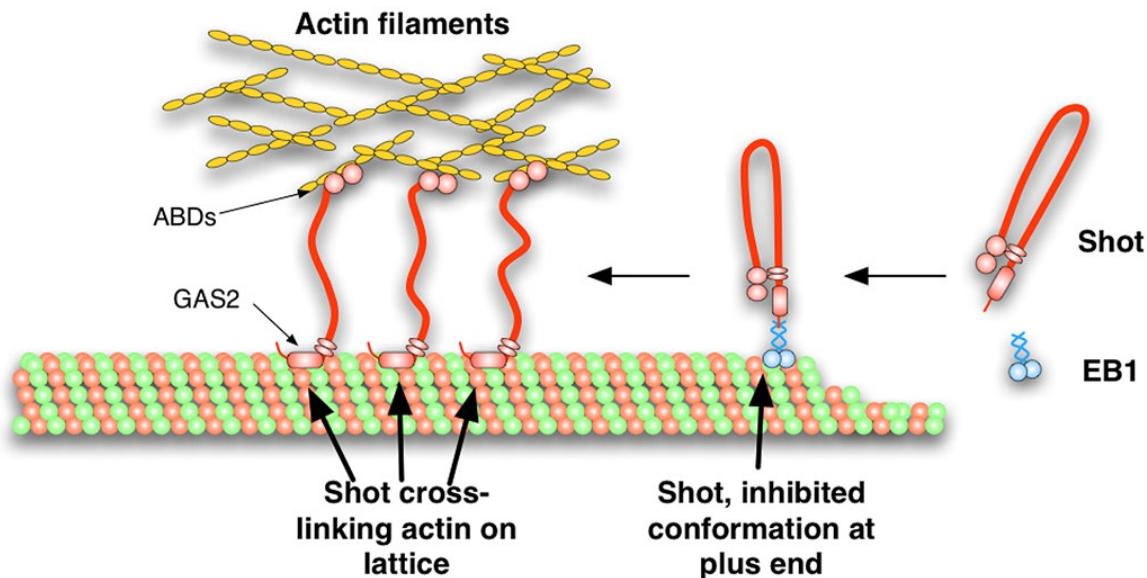


Figure 1.3: This figure, taken with permission from Applewhite *et al.*, depicts the autoinhibition model of Shot (2013). In this model, Shot is bent over onto itself and the Actin binding domain and EF hand/GAS2 domain interact with each other. The interaction prevents the Actin binding domain and EF hand/GAS2 domain from binding Actin or microtubules, respectively. The bent-over Shot is directed to the plus ends of microtubules by EB1, until, by a yet unknown means, the autoinhibition is released and Shot opens up to bind to both Actin and microtubules.

The conserved domains of the spectraplakin give it a wide range of functions. Certain combinations of functional domains are required only in certain contexts, so a spectraplakin can play a variety of roles in the cell (Bottenberg et al., 2009). Its many roles require regulation in order for it to play the right role in the right situation.

### 1.2.1 Regulation of Spectraplakins

Spectraplakins are regulated by other proteins to activate or deactivate their functions in the cell. As mentioned in the section on EF hands, intracellular calcium is a way of regulating BPAG1 (Kapur et al., 2012). Applewhite *et al.* did not observe this in Shot, however (2013). There are additional ways in which spectraplakins are regulated. In the literature, I have found evidence for autoinhibition, phosphorylation and targeted degradation.

One model for Shot regulation is autoinhibition. Autoinhibition is the negative regulation of a protein by intramolecular interactions between its domains (Pufall and Graves, 2002). In the autoinhibition model, inactive Shot is conformationally autoinhibited by the doubling over of its spectrin repeat rod which causes the interaction between its Actin binding domain and its EF hands. The autoinhibited Shot is localized to the plus end of the microtubule by EB1. When Shot is activated, it both binds to side of the microtubule and to Actin filaments, thus crosslinking Actin and microtubules. The regulation of this action is yet unknown (Applewhite et al., 2013). See Figure ?? for an illustration of the autoinhibition model.

ACF7 is regulated in part by phosphorylation. Phosphorylation is the process of introducing a phosphate group into a protein, which can activate, deactivate or change the protein. Proteins that phosphorylate molecules are called kinases. ACF7 requires phosphorylation by the dual kinase complex, Focal Adhesion Kinase (FAK)/Src, before it can bind F-actin *in vivo* (Yue et al., 2016). Phosphorylation of the C-tail by GSK $\beta$  is required for ACF7 to disassociate from microtubules in mice (Wu et al., 2011). Expression of Zw3, also known as Shaggy, the *Drosophila* homolog of GSK $\beta$ , did not affect Shot dynamics however (Applewhite et al., 2013). The Par1 kinase excludes Shot from associating with actin, but the mechanism of this is unknown (Nashchekin et al., 2016). The role of phosphorylation in Shot dynamics is less studied than it is in ACF7.

The last method I have found in the literature for regulation of ACF7 is ubiquitination. Ubiquitin is a protein that marks cells for destruction by the proteasome, a massive complex that breaks proteins down into their constituent amino acids (Kleiger and Mayor, 2014). The protein HectD1 targets ACF7 for degradation by the ubiquitin pathway by attaching a ubiquitin molecule to ACF7 (Duhamel et al., 2018). Ubiquitination can be itself regulated by Ubiquitin specific proteases, proteins that remove the attached Ubiquitin (Komander, 2010).

Intracellular calcium, autoinhibition, phosphorylation and ubiquitination are the ways I have found in the literature for regulation of spectraplakins. These methods of regulation are not universally shared among all spectraplakins. I want to now discuss

how spectraplakins affect cell migration, but first I introduce cell migration itself.

## 1.3 Cell Migration

When a cell migrates across a surface, the cell reaches forward with a part of itself toward its destination, then grips the surface and pushes off from that point to drag the rest of itself in that direction. Then the cell lets go so it can extend, grip, and move itself again. This method of locomotion is called mesenchymal cell migration (Figure 1.4). Mesenchymal cell migration is important during development, wound healing, immune responses to common diseases, and the invasion of healthy tissue by cancer (Franz et al., 2002; Etienne-Manneville, 2013; Case and Waterman, 2015).

In mesenchymal cell migration, the cell moves itself forward using structures called the lamella and lamellipodium that are made out of Actin (Gardel et al., 2010). Nascent adhesions form in the lamellipodium to grip the cell surface, then move back toward the lamella to grow into focal adhesions to provide more traction and later release the surface (Gupton et al., 2007). To insure that “forward” is toward its destination, the cell orients itself using microtubules (Etienne-Manneville, 2013). Many proteins work together as regulators to ensure that this process of extension, contraction, gripping and releasing occurs in a coordinated manner.

Mesenchymal migration takes place in four phases: protrusion, adhesion, contraction and retraction (Etienne-Manneville, 2013; Gardel et al., 2010). Protrusion is the cell reaching out across a surface, adhesion is the cell gripping onto the surface, contraction is the cell pulling itself forward along the surface, and retraction is the gathering of any left-behind parts of the cell in preparation for further movement. This section describes the first two of these phases, as there is more in the literature about spectraplakins in these phases.

### Protrusion: the Lamellipodium

The lamellipodium is a transient structure that appears at the leading edge of a migrating cell. It reaches out ahead of the main body of the cell and attaches to the surface. It is primarily made of tightly interconnected branched F-actin in a dense gel-like mesh (Gardel et al., 2010). While microtubules are largely excluded from the lamellipodium they do exert force on it, and in most cells a combination of F-actin elongation and microtubules generate the force for lamellipodial protrusion (Etienne-Manneville, 2013). Certain microtubules are present in the lamellipodium however, as we will see.

As the lamellipodium is a protrusive structure, branched F-actin is the kind of Actin that composes the lamellipodium (Gardel et al., 2010). The synthesis of Actin is what pushes the cell membrane forward. To make branched F-actin, Actin nucleator ARP2/3 is recruited and activated by the SCAR/WAVE complex of nucleation-promoting factors (Gautreau and Krause, 2014). Controlling the activation of the SCAR/WAVE complex’s activation of ARP2/3 controls the speed and direction of the cell’s migration. The SCAR/WAVE complex is thought to be regulated by the inte-

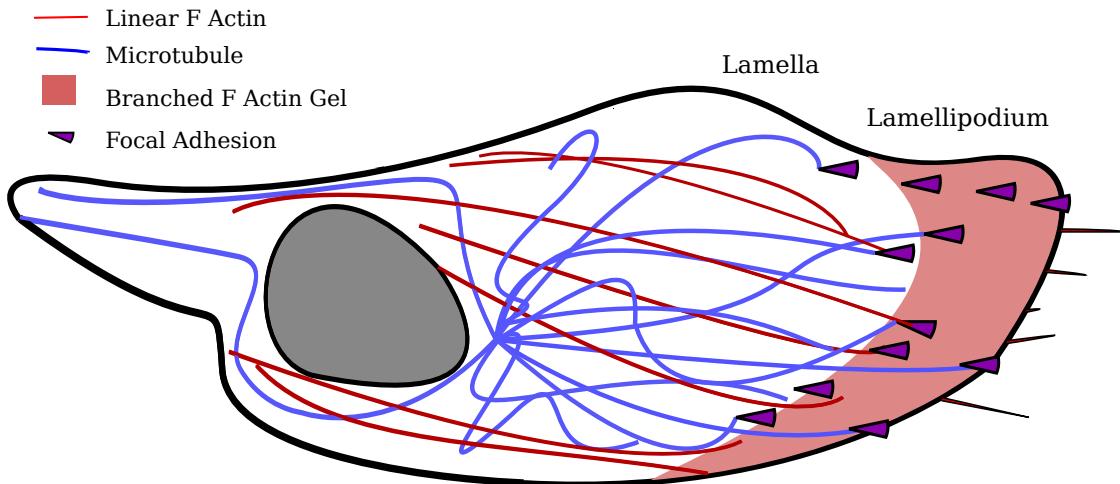


Figure 1.4: A schematic depiction of a eukaryotic cell (with centrosomal microtubules) migrating to the right of the page. The lamellipodium protrudes forward as focal adhesions provide traction with the surface. Branched Actin structures ahead of the focal adhesions provide motive force by pushing off from the focal adhesions, while linear Actin structures behind the focal adhesions pull the trailing contents of the cell forward via the adhesions. Focal adhesions form in the lamellipodium where microtubules are then attached to regulate their maturation in the lamella and their eventual disassembly as the main body of the cell overtakes their fixed position.

grative regulation of a combination of GTPases, phosphorylation and transmembrane receptors. The central GTPases for cell migration are Rho, Rac and Cdc42 (Gardel et al., 2010; Ridley, 2015; Lawson and Burridge, 2014). For more about GTPases, see Appendix A.

### Adhesion: Focal Adhesions

Mesenchymal cell migration requires the coordinated protrusion and adhesion of the lamellipodium to the outside environment (Case and Waterman, 2015). Adhesion is achieved through the formation of focal adhesions. Focal adhesions are the traction points the cell uses to grip onto the surface.

Embedded in the cell membrane around the lamellipodium are heterodimers of a protein called integrin. Integrins are a family of proteins that facilitate adhesion to the extracellular matrix, or ECM. *D. melanogaster* has five  $\alpha$  integrins and two  $\beta$  integrins (Humphries, 2000). Integrins signal bidirectionally based on the amount of tension they experience at the focal adhesion (Sun et al., 2016). Their signaling is propagated by RhoGTPases. (Lawson and Burridge, 2014).

When integrin binds to the ECM it creates a nascent focal adhesion in the lamellipodia containing integrin, focal adhesion kinase and paxillin (Case and Waterman, 2015). The nascent focal adhesion then binds the Actin cytoskeleton to the ECM-bound integrin acting as a “molecular clutch” (Sun et al., 2016). Tailin, vinculin, and  $\alpha$ -actinin bind Actin to the focal adhesion (Case and Waterman, 2015; Swami-

nathan et al., 2017). The majority of these newly created focal adhesions dissipate in the lamellipodium, but those that persist aggregate at the border of the lamella and lamellipodium to mature and lay down more traction (Case and Waterman, 2015; Shemesh et al., 2009). This molecular clutch is the traction point at which the protruding lamellipodium can push off of and the trailing lamella can pull itself toward. The Actin network being drawn back from the focal adhesion site to the cytosol is thought to be the mechanism by which motive force is generated to pull the rear end of the cell forward (Callan-Jones and Voituriez, 2016; Gardel et al., 2010).

Because the focal adhesion attaches the cell to the surface, the focal adhesions must be regulated so that they do not anchor the cell to somewhere it is trying to leave. Microtubules control focal adhesion formation, maturation and disassembly (Etienne-Manneville, 2013). Spectraplakins connect microtubules to the focal adhesion (Wu et al., 2008). Focal adhesion dynamics and cell migration are regulated by microtubules through the MAP4K4/IQSEC/Arf6 pathway (Ning et al., 2016; Yue et al., 2014). In addition to MAP4K4, disassembly is also regulated by FAK and Dynamin, which are activated by microtubules (Ezratty et al., 2005). Focal adhesions are bound to Actin and also need to interact with microtubules to function properly. Spectraplakins are involved in the meeting of Actin and microtubules at focal adhesions.

### 1.3.1 Spectraplakins and Migration

In both ACF7 and Shot, spectraplakins affect how cells migrate. Figure 1.5 is visual overview of spectraplakins in cell migration. More has been written about ACF7 than Shot regarding migration. I assume this is because cell migration is more studied in mammalian cells than in *D. melanogaster* because of cancer research. I begin with ACF7 because it is better understood.

In mouse epithelial cells, cells that cover the internal or external surfaces of an anatomical structure, cannot migrate to heal wounds when they are deprived of ACF7. Both *in vitro* and *in vivo*, mouse keratinocytes in skin do not migrate without ACF7 (Yue et al., 2016; Wu et al., 2008). In intestinal epithelial tissue, ACF7 affects intestinal wound healing and its deficiency is associated with ulcerative colitis in both mice and humans (Ma et al., 2017). Ulcerative colitis is symptomatic of epithelial cells being unable to bunch together tight enough to maintain a proper barrier in the gut.

Ubiquitination is a regulatory pathway for ACF7 in migration. HectD1 is a ubiquitin ligase that targets ACF7 for degradation (Duhamel et al., 2018). When Duhamel *et al.* depleted ACF7 in human breast cancer cells, those cells stopped migrating, but when Duhamel *et al.* depleted HectD1 from human breast cancer cells, the cells began to migrate more (2018). Additionally, in patients with breast cancer, low expression of HectD1 correlated with poorer outcomes (Duhamel et al., 2018).

Focal adhesions are an interface site for microtubules and the Actin cytoskeleton, in which spectraplakins are involved. Without ACF7 in mice keratinocytes, defects appear in focal adhesions and cause affected cells to move more slowly (Wu et al., 2008). Also in mice keratinocytes, ACF7 is localized at focal adhesions connecting the

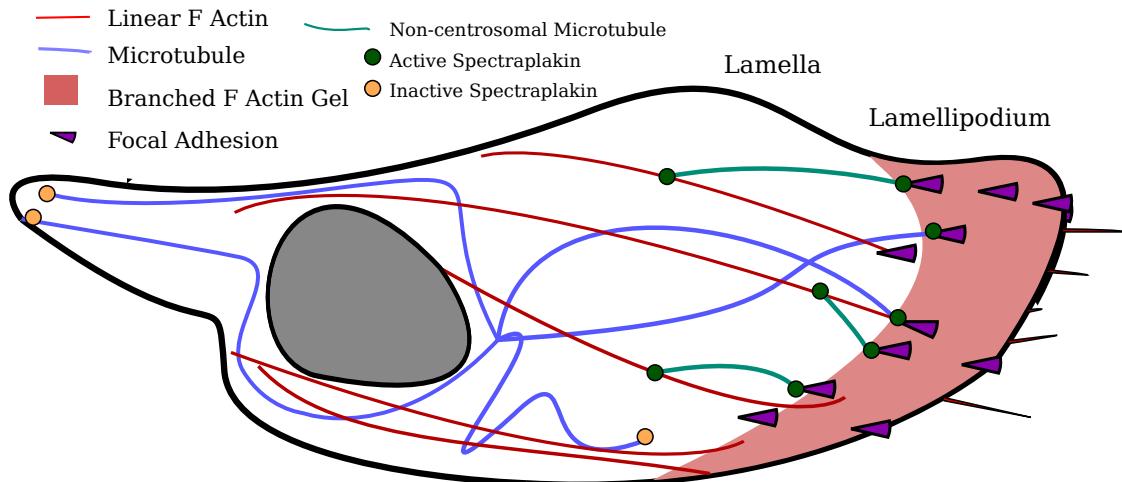


Figure 1.5: A schematic depiction of a eukaryotic cell (with centrosomal and non-centrosomal microtubules) migrating to the right of the page, with the locations of spectraplakins playing roles in cell migration. Inactive spectraplakins localize to the plus ends of extending microtubules via EB1. Active spectraplakins cross link Actin and microtubules at focal adhesions on the plus ends of microtubules. Active spectraplakins also link non-centrosomal microtubules to the actin cytoskeleton.

plus end of microtubules with the Actin network, thus guiding microtubules to focal adhesions (Yue et al., 2016). The signalling protein APC and the +TIP CLASP2 are required for ACF7 to guide microtubules to focal adhesions (Zaoui et al., 2010). APC and CLASP2 are regulated by the kinase ErbB2 (Zaoui et al., 2010). ACF7 also binds to the minus ends of non-centrosomal microtubules in conjunction with the CAMSAP 3 complex (Ning et al., 2016). This aligns the microtubules to be perpendicular with the leading edge of the cell. Without ACF7, the microtubules are misaligned, and they cannot traffic vesicles to the focal adhesions, as measured through MAPK4K4 (Ning et al., 2016).

Given the centrality and importance of the RhoGTPases Rho, Rac and Cdc42 to cell migration in general and focal adhesions in specific, these RhoGTPases are candidates for regulating spectraplakins (Ridley, 2015). However, in mouse epidermal cells when ACF7 was absent from the cell, the levels of Rho, Rac and Cdc42 did not change despite absence of ACF7 affecting focal adhesion dynamics (Wu et al., 2008). Additionally, Wu *et al* found that the absence of ACF7 did not affect the levels of FAK/Src activity (2008). While FAK/Src is a regulator of ACF7 and a regulator of migration, it does not appear to be a regulator of ACF7 in the context of migration.

The role of Shot in migration is more ambiguous, and less is known (Voelzmann et al., 2017). In unpublished research, Applewhite *et al.* found that knocking down Shot in motile cells caused the cells to migrate faster in a scratch wound assay, in contrast to what is seen with ACF7 in mice. Shot forms a complex with EB1 and APC1 (Subramanian et al., 2003). Without Shot, EB1 and APC1 disassociate from the muscle-tendon junction in developing flies. APC1 is the *D. melanogaster* equivalent of APC. This raises the possibility that ErbB2's *D. melanogaster* equivalent,

dEGFR, could regulate Shot through APC1, like ErbB2 regulates ACF7 through APC (Alvarado et al., 2009). Shot’s role in protrusion may be part of its role in migration. Shot links the microtubules that extend into the lamellipodium with the Actin cytoskeleton (Applewhite et al., 2010). Both the Actin and microtubule binding activities of Shot are required for epithelial cells to properly “zipper” together and close the hole along the back of a developing *D. melanogaster* embryo (Takcs et al., 2017). Takacs *et al.* hypothesize that lack of Shot function causes abnormal protrusion and thus causes the embryo to fail to zipper (2017). Shot’s role in migration has been studied less, but there are some results in the literature.

This concludes the background portion for biological topics. With the knowledge about spectraplakins and the role they play in migration, I now turn to interactomes as a method for answering questions about spectraplakins and migration. Before I can talk about how I used the interactome explore Shot and migration, I first explain some computational concepts.

## 1.4 Networks

A network is a description of a set of elements with connections between them. An ecological food web is an example of a network. Networks are represented by graphs, and a graph is defined as consisting of a set  $N$  of nodes and a set  $E$  of edges (Steuer and Lopez, 2008). A node is an object in a network. For example, the circle labeled “A” in the example network in Figure 1.6 is a node. An edge is a connection between two nodes. For example, the line between the circle labeled “A” and the circle labeled “B” in Figure 1.6 is an edge. A node’s degree is defined as the number of edges connected to it. For example, node “A” in Figure 1.6 has five edges and has degree five, whereas the node “G” only has one edge connected to it and has degree one. This thesis deals with undirected networks, where no distinction is made as to whether the connection that an edge makes goes from one node to the other. Networks can also be represented as a table. Table 1.2 contains the same information as Figure 1.6, where the edges are rows and the nodes are columns. The networks in this thesis are stored in tabular form.

## 1.5 Protein-Protein Interactomes

Protein-protein-interactomes are an application of networks to model the interactions between proteins, where proteins are nodes and interactions are edges. A protein-protein interactome is a large scale network catalog of all observed protein-protein interactions in a given cell. Protein-protein interactions are defined as physical contact on a molecular level between proteins that occurs in a cell or organism (De Las Rivas and Fontanillo, 2010). Networks of proteins are undirected and scale-free, which means the majority of nodes have low degrees of interconnectedness and are linked through a minority of high-degree hubs (Vidal et al., 2011; Sevimoglu and Arga, 2014).

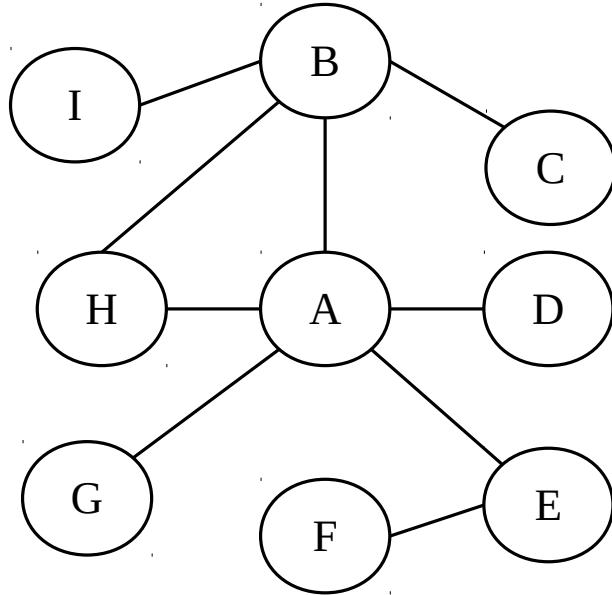


Figure 1.6: An example undirected network with nine nodes, which are the labeled circles, and nine edges connecting the nodes.

Interactomes can be built with high-throughput experimental data, such as in the case with co-affinity purification mass spectrometry and yeast two hybrid systems; computational data like interologs; or by hand using curated datasets. Co-affinity purification mass spectrometry (CoAP/MS) is technique which uses a bait protein to collect any protein or complex that interacts with the bait protein. When the bait protein is isolated, the interacting proteins are collected with it. (Yu et al., 2008). CoAP/MS interactomes collect protein interactions that interact indirectly as part of a complex. A yeast two hybrid system is a technique in which a reporter gene's transcription factor is split into two halves and a protein of interest is fused to each half. If the two proteins of interest interact, then the reporter gene is expressed. Yeast two hybrid systems produce interactomes that are enriched with transient or condition-specific interactions (Yu et al., 2008). An interolog is the assumption that if an interaction occurs between two proteins that are both homologous across organisms, then the interaction between the two homologous proteins is also homologous (Yu et al., 2004). Finally, there are curated interactomes which are simply interactomes built by hand from data in the literature. These interactomes have a high signal:noise ratio, but they are limited by the slow resource-intensive process it takes to create one and the fact that they can only contain information that is already available in the literature.

Interactomes are large-scale catalogs and as such are difficult to interpret unaided. I developed cytoscouts to assist me in understanding interactomes. Cytoscouts is a script written in Python, so I now introduce Python.

Table 1.2: An alternative representation of the network in Figure 1.6.

	Node <sub>1</sub>	Node <sub>2</sub>
	A	B
	A	D
	A	E
	A	G
	A	H
	B	C
	B	I
	B	H
	E	F

## 1.6 Python Concepts

Python is a commonly-used computer programming language. I wrote cytoscouts in Python 3.6.4. In this section, I explain the Python concepts needed to understand what cytoscouts does. The definitions used in this section were taken from the official Python documentation (docs.python.org, 2018).

### Lists

Lists are defined as mutable sequences, meaning that they can be changed after they are created. They are ordered collections of information and are denoted with square brackets. Lists have indexes that allow you to modify elements of lists freely as long as you know their index. Below is an example of a list with four elements:

```
[ 'A' , 'B' , 'C' , 'D' ]
```

A list can be contained within a list. Lists of lists can be used as tables with the outer list representing the rows of a table and the inner list representing the columns.

For example the list of lists:

```
[ [ 'A' , 'B' ] , [ 'C' , 'D' ] ]
```

Is equivalent to the table:

	Column <sub>1</sub>	Column <sub>2</sub>
Row <sub>1</sub>	A	B
Row <sub>2</sub>	C	D

I used this approach to store networks as a list of lists. For an example, see Table 1.3.

Table 1.3: An alternative representation of the network in Figure 1.6, also showing how the table can be stored as a list of lists.

Node <sub>1</sub>	Node <sub>2</sub>	List of Lists
A	B	[['A', 'B'],
A	D	['A', 'D'],
A	E	['A', 'E'],
A	G	['A', 'G'],
A	H	['A', 'H'],
B	C	['B', 'C'],
B	I	['B', 'I'],
B	H	['B', 'H'],
E	F	['E', 'F']]

## Sets

A set in Python is defined as an unordered collection with no duplicate elements. They are denoted with curly brackets. Sets only contain immutable objects so a set cannot contain a list. I used sets when making lists of items that did not need more than one entry. Below is an example of a set in Python:

```
{'A', 'B', 'C', 'D'}
```

## Tuples

Tuples are like lists except they are immutable. This means that they can be added to sets, but they cannot be changed after they are created. Tuples are denoted with parentheses. I used tuples when I needed to put things in sets and thus could not put things in a list. Below is an example of a tuple in Python:

```
('A', 'B', 'C', 'D')
```

Sets of tuples can represent the edges and nodes of a network just like lists of lists can. For example the set of tuples:

```
{('A', 'B'), ('C', 'D')}
```

Is equivalent to the table:

	Column <sub>1</sub>	Column <sub>2</sub>
Row <sub>1</sub>	A	B
Row <sub>2</sub>	C	D

However, since sets are unordered, that same set of tuples is also equal to the table:

	Column <sub>1</sub>	Column <sub>2</sub>
Row <sub>1</sub>	C	D
Row <sub>2</sub>	A	B

Fortunately for me, when working with networks the order of edges in the table is not important so an unordered grouping like a set can be used. For an example of a network represented as a set of tuples, see Table 1.4. I used sets of tuples when I wanted to have only unique edges in a network and get rid of duplicate edges.

Table 1.4: An alternative representation of the network in Figure 1.6, also showing how the table can be stored as a set of tuples.

Node <sub>1</sub>	Node <sub>2</sub>	Set of Tuples
A	B	{('A', 'B'),
A	D	('A', 'D'),
A	E	('A', 'E'),
A	G	('A', 'G'),
A	H	('A', 'H'),
B	C	('B', 'C'),
B	I	('B', 'I'),
B	H	('B', 'H'),
E	F	('E', 'F')}

## Dictionaries

A dictionary is an unordered set of key:value pairs. Key:value pairs are denoted with “:.” Metaphorically, the key can be thought of as a word, and the value as its definition. The dictionary key does not need to be a word though; it can be any immutable data structure. That is to say a set can be a dictionary key or value, but a list cannot. I used dictionaries to count the number of entries associated with an element of a list. When a key, “Key,” is put into a dictionary, “dict,” it takes the form of `dict[Key]`. The dictionary then produces the value corresponding to that key, in this case “Value.” In code a dictionary is created like this: `dict={'Value': 'Key'}`, and with the dictionary `dict`, `dict['Key']` would then produce `'Value'`.

## 1.7 Bioinformatics Tools and Resources

There are a number of tools that make applying networks to biological systems easier and I would like to introduce them here. UniProt Knowledgebase is the database that contains the nodes of the interactome I worked with. The Gene Ontology is a vocabulary for understanding what proteins are and do. FlyBase is resource for information on the fruit fly. Clustal Omega and BLAST are programs that align

biological sequences like proteins. GraphSpace is a way to visualize networks using Python. These tools were essential to my thesis.

### UniProt Knowledgebase

In this thesis, the interactome I worked with describes proteins by their UniProt Knowledgebase accession IDs. UniProt Knowledgebase, or UniProtKB, is a collection of annotated protein sequences for a variety of proteomes across many species, including *D. melanogaster* (Bateman et al., 2017). Because UniProtKB is a collection of protein sequences from many sources, redundant entries can be a problem and while Bateman *et al.* have taken steps to reduce the number of redundant entries in UniProtKB, I still encountered them in this thesis.

### The Gene Ontology

The Gene Ontology is an effort to create a unified database for eukaryotic genes and proteins with a structured vocabulary to describe their biological processes, molecular function, and cellular location (Ashburner et al., 2000). Entries in the Gene Ontology are annotated to provide a quantitative comparison of gene and protein function. These annotations can be either experimentally or computationally derived, and the majority are automatic and not curated. The Gene Ontology can be used to analyze the results of high throughput experiments. Protein protein interactions can be inferred through the Gene Ontology (du Plessis et al., 2011) .

In this thesis, I used a Gene Ontology tool called PANTHER (Protein ANalysis THrough Evolutionary Relationships) to classify the proteins found in the interactome. PANTHER is a system that aims to infer the function of genes and their proteins through their evolutionary relationship to other genes and proteins (Mi et al., 2013). It does this through combined hand curation and bioinformatics algorithms. Another Gene Ontology tool I use is AmiGO, the official web-based toolset for using the Gene Ontology database (Carbon et al., 2009).

### Other Bioinformatics Tools

FlyBase, BLAST, Clustal Omega and GraphSpace are other bioinformatics tools I used in this thesis. FlyBase is an online repository of information on *D. melanogaster* (Gramates et al., 2017). I used FlyBase for looking up information on *D. melanogaster* proteins. BLAST is a program that aligns sequences, and I used it to compare proteins to see where in their structure they are similar or different (Altschul et al., 1997). Clustal Omega is another program that aligns biological sequences, and I used it to compare proteins (Goujon et al., 2010; Sievers et al., 2011). BLAST is designed to compare two sequences, and Clustal Omega is for comparing three or more sequences. GraphSpace is a web based platform for creating and interacting with networks (Bharadwaj et al., 2017). I used GraphSpace to visualize the data produced by my program. See Figure 1.7 for an example of a network in GraphSpace.

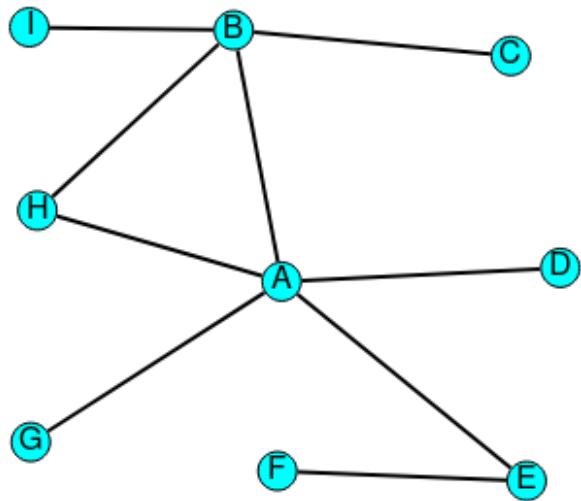


Figure 1.7: The same undirected network as in Figure 1.6 generated in GraphSpace.

# Chapter 2

## Materials and Methods

Dr. Anna Ritz created a new interactome from publicly available data (Ritz, 2017). I was one of the first to be able to work with this new interactome, I was lucky enough to have the opportunity to name it. Because Dr. Ritz’s interactome takes its information from an “eclectic” wide and varied array of sources and because birds and flies both fly, I named this new interactome the Eclectus Fly Interactome after the species of parrot *Eclectus roratus*. In this section I discuss the Eclectus Fly Interactome. Then, I introduce “collapsing” the interactome and why I think that is a useful thing to do. After that, I explain cytoscouts, the script to process the interactome.

### 2.1 The Eclectus Fly Interactome

The goal of the Eclectus Fly Interactome is to provide a wide-encompassing general collection of any interaction in any *D. melanogaster* cell observed from any source. The Eclectus interactome is the aggregate of six different interactomes. The six interactomes that make up the Eclectus interactome are listed in Table 2.1.

Table 2.1: Interactome sources with edge counts.

Name	Edges	Nodes	Source
DroID	262,179	11,444	(Murali et al., 2011)
FlyMine	278,370	15,049	(Lyne et al., 2007)
FlyReactome	612	133	(Matthews et al., 2009)
Mentha	45,669	10,819	(Calderone et al., 2013)
MyProteinNet	87,529	12,192	(Basha et al., 2015)
SignaLink 2	5,236	1,343	(Fazekas et al., 2013)
Eclectus Fly Interactome	542,637	18,162	

DroID is a collection of high throughput experimental data and interologs (Murali et al., 2011). FlyMine is also a high throughput experimental dataset interactome (Lyne et al., 2007). FlyReactome is a hand-curated interactome (Matthews et al., 2009). Mentha is a collection of interactions from other interactomes in the literature (Calderone et al., 2013). MyProteinNet is also a collection of interactions from other interactomes in the literature (Basha et al., 2015). SignaLink is a combination of high throughput experimental data and hand curation (Fazekas et al., 2013). The Eclectus Fly Interactome is a collection of unique edges from six constituent interactomes. This means that there are more edges in the sum of the Eclectus interactome's parts than there are in the Eclectus interactome itself.

The Eclectus interactome uses nodes that are UniProtKB IDs. The nodes are stored as edges in a text file, and they are processed as a list of lists in the Python script. UniProtKB, or UniProt, is a database of annotated protein sequences (Batemann et al., 2017). Using UniProt IDs gives specificity for each isoform but can inflate number of proteins in an interactome by including redundant proteins as separate nodes. The same protein and even isoform can have multiple UniProt ID entries if those isoforms were entered into the UniProt database from different sources. For example, from the Eclectus interactome, the protein Katanin p60 has two entries in the UniProtKB: one from homology data and one from RNA data, despite both sequences being for the same protein. When I used the Eclectus Fly Interactome, I ran into redundant UniProt IDs frequently.

### 2.1.1 The Concept of Collapsing the Interactome

Because of problems with redundant entries, I came up with an alternative way of reading the interactome with cytoscouts. This section is a conceptual explanation of the process of collapsing the Eclectus interactome. Instead of UniProt IDs, *e.g.* “A1Z9J1,” I made an interactome out of nodes listed by common name, *e.g.* “Shot,” by using another file that listed which UniProt ID corresponded to which common name. Then, since multiple nodes could share the same common name, I “collapsed” nodes with the same common names together to eliminate redundancies. The caveat to collapsing the interactome is that it assumes all nodes with the same common name are redundant. By only using the common names of proteins, isoforms are ignored. I used the collapsed and uncollapsed interactomes together in my thesis because of the limitations of both the collapsed and uncollapsed interactomes. I call the Electus interactome as it exists unmodified the “uncollapsed interactome” to contrast it to the collapsed interactome.

Collapsing works by combining nodes with equivalent names. In Figure 2.1 the top left toy network represents the interactome with UniProt ID nodes, and the top right toy network is that same interactome with common name nodes. Nodes “C,” “D” and “E” are equivalent to each other as “J” and are redundant, and nodes “G” and “H” are also equivalent as “K.” Since all nodes labeled “J” are equivalent to each other in the network, they can be pinched together and “collapsed” into one node. The same can be said for node “K.” The bottom panel of Figure 2.1 shows the collapsed network for the toy network.

In the uncollapsed Eclectus interactome there are 21 UniProt nodes that are equivalent to “Shot,” and this is like in Figure 2.1 there are three nodes that are equivalent to “J.” Collapsing the Eclectus Fly Interactome produces one node “Shot,” just like collapsing the network in Figure 2.1 produces one node “J.”

The way that I did this was by using a set of tuples to contain the interactome instead of a list of lists. Since it is a set, duplicate entries are omitted by definition. The interactome is handled in a set of tuples instead of a set of lists because lists cannot be stored in sets due to being mutable, while tuples are immutable and thus can be stored in sets. Please see Table 2.2 for how converting a list of lists into a set of tuples removes the redundant entries. To discuss this concept further I want to introduce my script, cytoscouts.

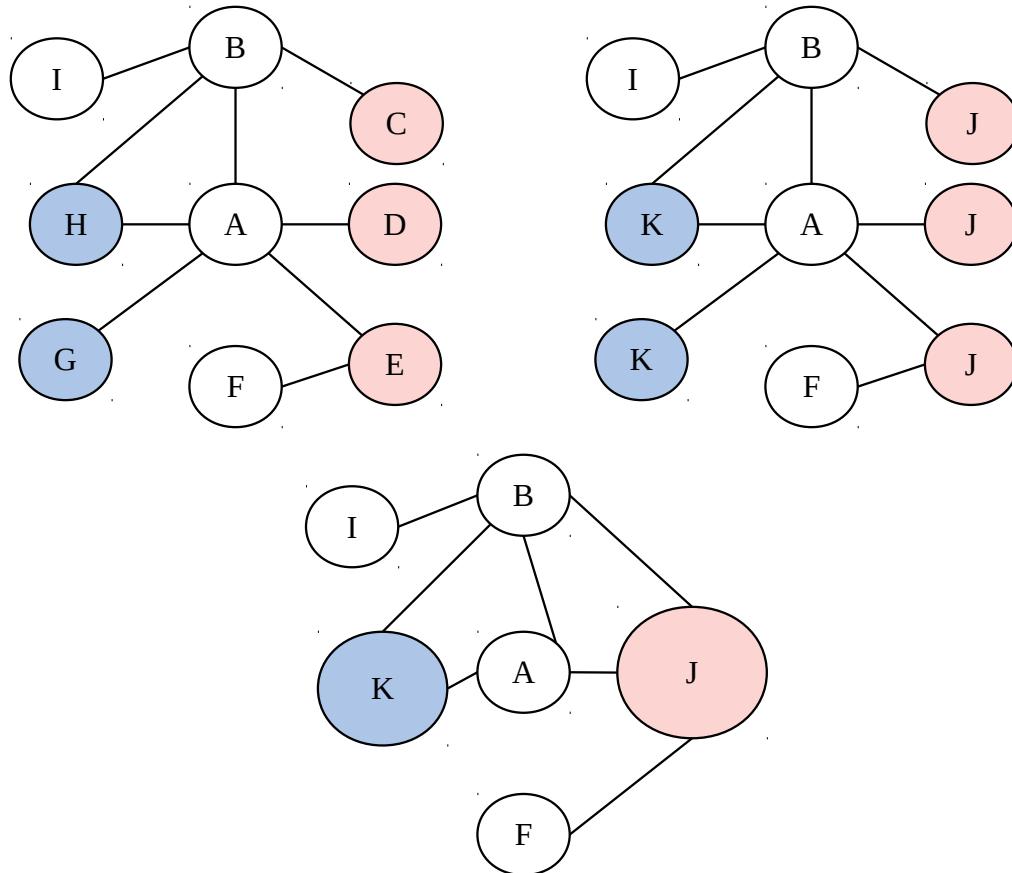
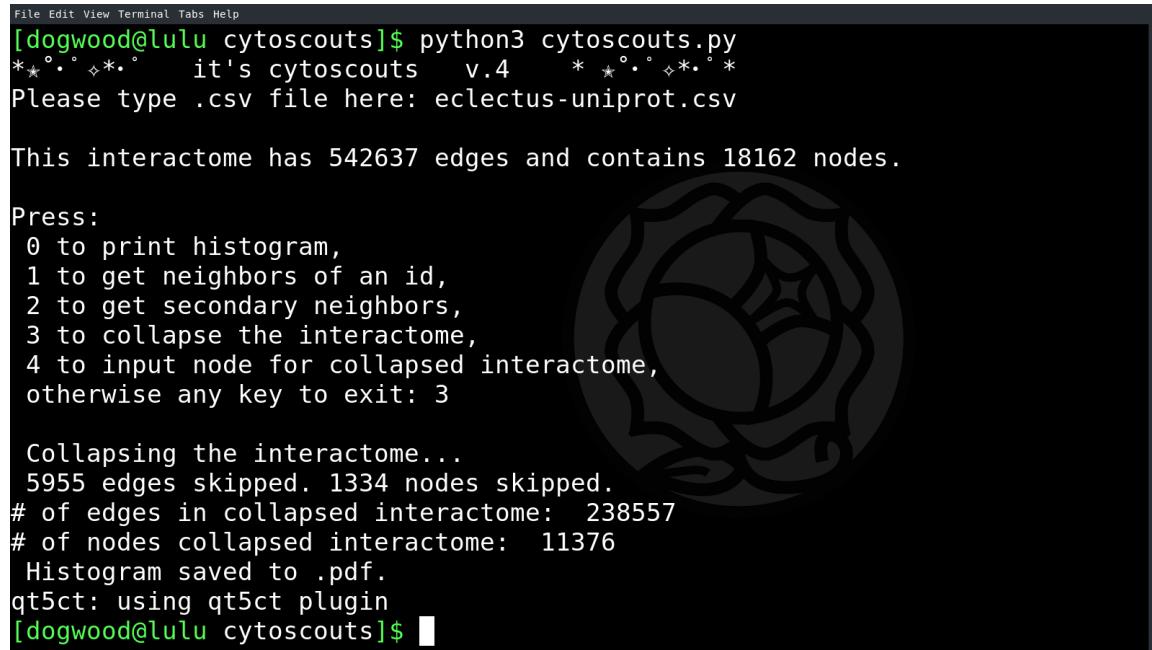


Figure 2.1: If nodes C, D and E are equivalent to node J, and nodes G and H are equivalent to node K, then the nodes can be collapsed. The top left network is uncollapsed and the bottom network is collapsed, and the top right network is in an intermediate stage.

Table 2.2: A table representation of the collapsing method in Figure 2.1. The redundant nodes are collapsed by treating edges as tuples in a set rather than as a list in a list, as a set cannot contain duplicate elements.

Node <sub>1</sub>	Node <sub>2</sub>	List of Lists	Set of Tuples
A	B	[['A', 'B'],	{('A', 'B'),
A	J	['A', 'J'],	('A', 'J'),
A	J	['A', 'J'],	
A	K	['A', 'K'],	('A', 'K'),
A	K	['A', 'K'],	
B	J	['B', 'J'],	('B', 'J'),
B	I	['B', 'I'],	('B', 'I'),
B	K	['B', 'K'],	('B', 'K'),
J	F	['J', 'F']]	('J', 'F') }

## 2.2 Cytoscouts



```

File Edit View Terminal Tabs Help
[dogwood@lulu cytoscouts]$ python3 cytoscouts.py
*.*.*.* it's cytoscouts v.4 *.*.*.**
Please type .csv file here: eclectus-uniprot.csv

This interactome has 542637 edges and contains 18162 nodes.

Press:
0 to print histogram,
1 to get neighbors of an id,
2 to get secondary neighbors,
3 to collapse the interactome,
4 to input node for collapsed interactome,
otherwise any key to exit: 3

Collapsing the interactome...
5955 edges skipped. 1334 nodes skipped.
# of edges in collapsed interactome: 238557
# of nodes collapsed interactome: 11376
Histogram saved to .pdf.
qt5ct: using qt5ct plugin
[dogwood@lulu cytoscouts]$ █

```

Figure 2.2: A screenshot of cytoscouts collapsing the Eclectus Fly Interactome via the command line interface.

Cytoscouts is a Python script that can be run on any computer with Python 3. It has a rudimentary user interface where commands are entered by keyboard. It takes an interactome in .csv format, specified by user input, and processes the interactome. From the interactome, cytoscouts can make lists of neighbors of a given node and print

histograms, which are bar graphs that show the distribution of nodes with a given degree, for the uncollapsed or collapsed interactome. The full code for cytoscouts is available at

<https://github.com/pwijngaard/cytoscouts>. For an example of what a cytoscouts session looks like, please see Figure 2.2. When using the script the user has five options:

- 0 Get a histogram of the entire uncollapsed interactome and save it to a file in the current directory.
- 1 Get a list of the neighbors of a given UniProt ID in the uncollapsed interactome along with a histogram of the neighbors of the UniProt ID in the uncollapsed interactome and save each to a file in the current the directory.
- 2 Get a list of the neighbors of neighbors of a UniProt ID (this function is largely deprecated since I did not find it useful).
- 3 Collapse the interactome and then save a histogram of the collapsed interactome to a file in the current the directory.
- 4 Collapse the interactome then get a list of the neighbors of a given common name in the collapsed interactome along with a histogram of the neighbors of the common name in the collapsed interactome and save each to file in the current the directory.

In this section, I explain how cytoscouts works when each of these options is selected, except deprecated Option 2, and walk through the process of turning the interactome into graphs and human-readable lists. This section is split into two segments: one for the uncollapsed interactome and one for the collapsed interactome. Both segments use the same fundamental processes, but collapsing the interactome is an additional step.

### 2.2.1 The Uncollapsed Interactome

Cytoscouts in the uncollapsed interactome takes one or two inputs depending on the user's choice. If the user simply wants to generate statistics for an interactome, it takes the name of the interactome specified in a .csv file. If the user then selects Option 0, cytoscouts generates a histogram of the interactome and saves it to a file in the current directory. If the user is looking for the neighbors of a specific node, they would select Option 1, and then cytoscouts would ask for the UniProt ID of that node. Cytoscouts generates a list of the nodes neighboring the specified UniProt ID and a histogram of those nodes and saves each to a file in the current directory. The process cytoscouts goes through for Option 0 or Option 1 is described in Figure 2.3.

The first thing cytoscouts does is a user specified interactome stored as a .csv file and turns into a list of edges and a set of nodes. It does this using Function 2.1, `importCSV`. `importCSV` reads the first two columns of every line from the .csv file as a list of two elements, and then it adds that list of two elements to the list of lists

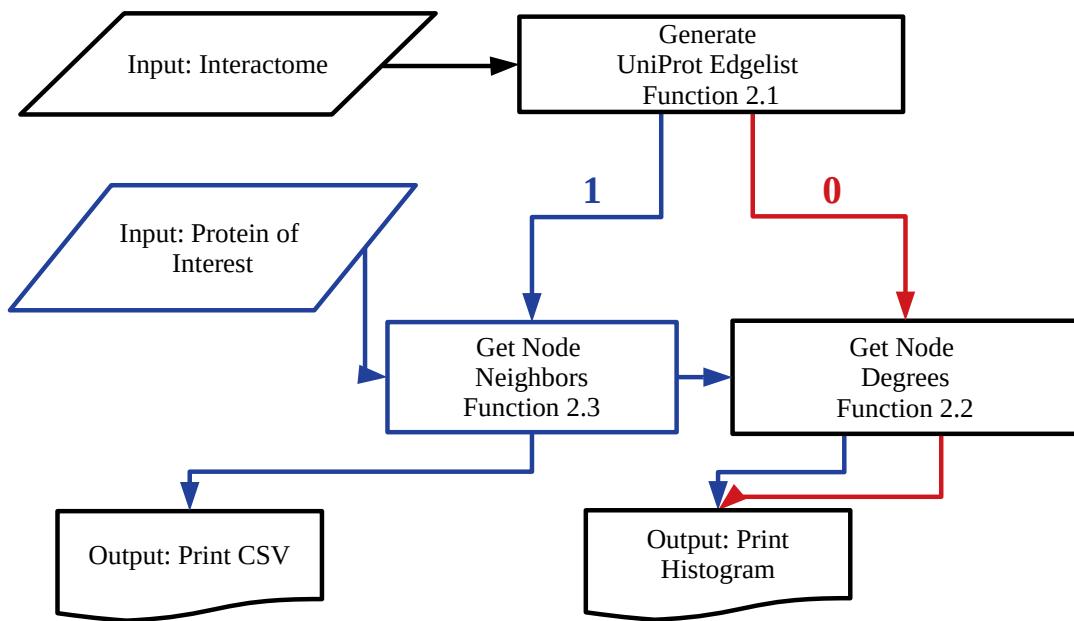


Figure 2.3: A flow chart of the process used to read the interactome and create histograms and lists. The path taken when Option 0 is selected is in red, and the path taken when Option 1 is selected is in blue.

`edgeList`. At the same time, it adds both elements to the set `nodeSet`. Since it is a set, any duplicate entries are ignored. By doing this the `importCSV` gathers all nodes and all edges from the .csv file. `importCSV` then returns the variables `edgeList` and `nodeSet` for use by other functions.

If the user selects Option 0, the function `getDegree`, seen in Function 2.2, starts and takes `edgeList` and `nodeSet` as inputs. `getDegree` runs a loop that builds a dictionary `deg` containing an entry for every element in `nodeSet`. Then it runs another loop where for every element in `nodeSet`, it looks through all of `edgeList` and every time it sees an element from set `nodeSet` in `edgeList` it adds one to the `deg` value of the key `nodeSet`. When it finishes this loop, `getDegree` returns `deg`. At that point `deg` has a key for every element in `nodeSet` with a value of that element's degree. Cytoscouts produces a histogram from `deg` of the distribution of nodes by node degree, which it writes to a file in the directory before it exits.

If the user has selected Option 1, the function `neighbors`, seen in Function 2.3, takes two inputs, the desired UniProt ID, `uniprot`, as well as `edgeList` from `importCSV`. Then it runs a loop in which it goes through every list of two elements in the list of lists `edgeList` and checks to see if one of the the elements in the list is `uniprot`. If one of the elements in the list is `uniprot`, then the function adds the other element that is not `uniprot` to the set `nNodes`. This way, `nNodes` becomes the set of every neighbor of `uniprot` in the interactome. `neighbors` then returns `nNodes` for use by other functions.

`getDegree` then takes `edgeList` and `nNodes` as inputs and using the same process

Function 2.1: `importCSV` creates a set of nodes and a list of edges from a user supplied .csv file.

---

```

1 import csv
2 def importCSV ():#get the CSV
3     fname = input('Please type .csv file here: ') #Prompt the user for a file name
4     nodeSet = set() #Initializes a set
5     edgeList = [] #Initializes an empty list
6     with open(fname, newline='') as csvfile: #Opens a file for reading
7         #Open a 'reader' object expecting in TSV format
8         reader = csv.reader(csvfile,dialect="excel-tab")
9         #A for loop where each line is iterated as a list of strings
10        #representing the rows of the table
11        #with each element of the list representing the columns of the table
12        for row in reader:
13            if row[0]=='UniProt1':#This if statement skips the header
14                continue
15            edgeList.append([row[0],row[1]]) #Appends to edgeList a pair of nodes: an edge
16            nodeSet.add(row[0])# Adds the first node to the set of nodes
17            nodeSet.add(row[1])# Adds the second node to the set of nodes
18    return edgeList,nodeSet#returns the list of edges and the set of nodes as variables

```

---

as it did with `edgeList` and `nodeSet`, produces the dictionary `deg` with values of degree for the key of every element in `nNodes`. Cytoscouts then saves every element in `nNodes` and that element's value from `deg` to a .csv file, produces a histogram from `deg` of the distribution of neighbor nodes by node degree and saves that to a file and exits.

Function 2.2: `getDegree` creates a dictionary of node:degree pairs.

---

```

1  def getDegree (edgeList,nodeSet):#Gather nodes by number of edges
2      deg = {} #Initializes the dictionary where seen nodes will be added
3      #the nodes are the keys and their degree is their value
4      for item in nodeSet:
5          deg[item]=0 #Every node starts with a degree of zero
6      for row in edgeList: #For every pair of nodes in the list of edges
7          deg[row[0]]+=1 #If the node is present in the first half of the edge
8              #add one to its degree
9          deg[row[1]]+=1 #If the node is present in the second half of the edge
10             #add one to its degree
11      return deg #returns the dictionary 'deg' as a variable

```

---

Function 2.3: `neighbors` creates a set of nodes that share an edge with a specified node.

---

```

1  def neighbors(uniprot,edgeList): #Gets the neighbors of a given node
2      nNodes=set() #Initializes an empty set of neighbor nodes
3      for row in edgeList:
4          #A for loop in which if a node in an edge is the uniprot id,
5          #then the other node in that edge is added to a set
6          if row[0] == uniprot:
7              nNodes.add(row[1])
8          if row[1] == uniprot:
9              nNodes.add(row[0])
10     return nNodes #Returns the set of neighboring nodes as the variable 'nNodes'

```

---

## 2.2.2 Collapsing the Interactome

Collapsing the interactome occurs if the user has selected Option 3 or 4. An additional file called `reference` that tells cytoscouts what common names correspond to what UniProt IDs is required to do this. The difference in process between the collapsed and uncollapsed interactome is that in the collapsed interactome, after the generation of `edgeList` and `nodeSet`, `edgeList` is collapsed into a set of tuples, `collapsedTupleSet`, and `nodeSet` is collapsed into `collapsedNodeSet`. See Figure 2.4 for the flowchart of the program with this added step. The set transformation produces an interactome consisting only of common names.

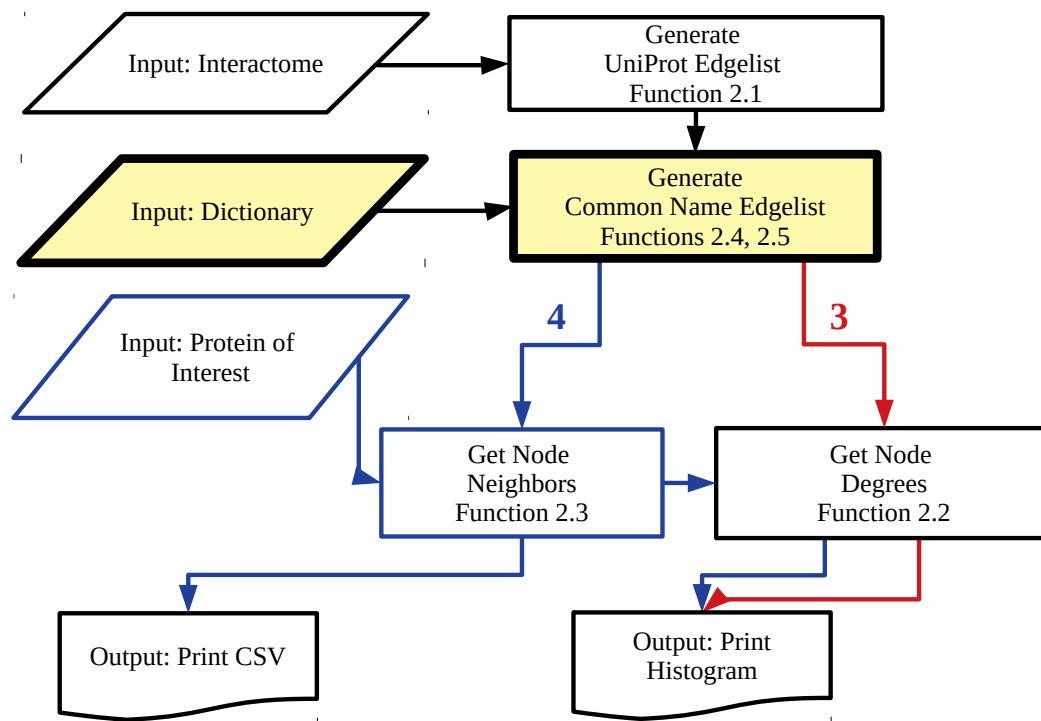


Figure 2.4: A flow chart of the process used to read and collapse the interactome and create histograms and lists. The highlighted nodes are the interposing step of collapsing the interactome. The path taken when Option 3 is selected is in red, and the path taken when Option 4 is selected is in blue.

Cytoscouts collapses the interactome by using Function 2.4, `collapseInteractome`. `collapseInteractome` takes the inputs `edgeList` as well as the dictionary `reference`. `collapseInteractome` goes through every list of two elements in `edgeList` and checks to see that both elements in the list are both keys in the dictionary `reference`. If both elements of the list are not keys in `reference`, then `collapseInteractome`

jumps to the next list in `edgeList` and adds one to the counter `skipped`. If both elements of the list are keys in `reference`, `collapseInteractome` then checks to see that a tuple of the reverse of the values of those two elements in `reference` is not already in the set of tuples, `collapsedTupleSet`. If the reverse order of the values of those two elements in `reference` is a tuple in `collapsedTupleSet`, `collapseInteractome` jumps to the next list in `edgeList`. If the reverse order of the values of those two elements in `reference` is not a tuple in `collapsedTupleSet`, `collapseInteractome` transforms the values of those two elements in `reference` into a tuple, and then adds that tuple to `collapsedTupleSet`. It loops through these steps until it has checked every list in `edgeList`. When it finishes checking every list in `edgeList`, `collapseInteractome` returns `collapsedTupleSet` and `skipped` for use by other functions.

Function 2.5, `collapseNodes`, takes the input of the set `nodeSet` and the dictionary `reference`. `collapseNodes` goes through every element in `nodeSet` and checks to see if that element is a key in `reference`. If that element is not a key in `reference` then `collapseNodes` jumps to the next element in `nodeSet` and adds one to the counter `skipped2`. If the element from `nodeSet` is a key in `reference`, then `collapseNodes` takes the value of that element in `reference` and adds the value to the set `collapsedNodeSet`. `collapseNodes` then returns `collapsedNodeSet` and `skipped2` for use by other functions.

Then the program continues with the collapsed interactome the same way as it did with the uncollapsed interactome. If the user selects Option 3, the program then continues as if they had selected Option 0 in the uncollapsed interactome, except now `getDegree` takes the inputs `collapsedTupleSet` and `collapsedNodeSet` instead of `edgeList` and `nodeSet`. If the user selects Option 4, the program continues almost the same as if they had selected Option 1 in the uncollapsed interactome, except now `getDegree` takes the inputs `collapsedTupleSet` and `collapsedNodeSet` instead of `edgeList` and `nodeSet`, and `neighbors` takes `commonName` and `collapsedTupleSet` as inputs instead of `uniprot` and `edgeList`. For the user, the only difference in the case of Option 4 is they must now use the protein's common name instead of its UniProt ID.

Cytoscouts can search for an individual node or visualize the entire interactome. It can do this with an interactome directly or it can collapse the interactome to remove equivalent nodes before searching. With the two ways of handling interactomes, I used cytoscouts to explore the Eclectus interactome and look for proteins that interact with Shot.

Function 2.4: `collapseInteractome` creates a set of tuples from the list of list `edgeList` and the dictionary `reference`, skipping edges that are not defined in `reference` and excluding duplicates that are tuples in reverse order.

---

```

1 def collapseInteractome (reference,edgeList):
2     collapsedTupleSet = set() #Initializes the set that will contain the tuple edges
3     skipped = 0 #Initializes the counter for skipped edges
4     #Below is a for loop that takes an edge and makes sure both nodes
5     #are in the reference
6     #then if they are, gets the dictionary value for both nodes
7     #and adds the dictionary values into a tuple and adds the tuple to the set
8     #collapsedTupleSet, provided the tuple is not a duplicate just flipped in order
9     for row in edgeList:
10         if row[0] in reference and row[1] in reference:
11             if (reference[row[1]],reference[row[0]]) not in collapsedTupleSet:
12                 collapsedTupleSet.add(tuple((reference[row[0]],reference[row[1]])))
13             else:
14                 skipped +=1 #Skips the nodes in the edge
15                     #if either one is missing from the reference,
16                     #then counts it as skipped
17     return collapsedTupleSet, skipped

```

---

Function 2.5: `collapseNodes` creates a set of common names from the set `nodeSet` and the dictionary `reference`, skipping names that do not have a definition in `reference`.

---

```

1 def collapseNodes (reference,nodeSet):
2     collapsedNodeSet = set() #Initializes the set of nodes
3     skipped2 = 0
4     #A for loop that adds the dictionary value for every element
5     #in nodeSet provided that that element has a dictionary value
6     for key in nodeSet:
7         if key in reference:
8             collapsedNodeSet.add(reference[key])
9         else:
10             #skipped2 += 1 Skips the node if it is missing from the reference,
11             #then counts it as skipped
12     return collapsedNodeSet, skipped2

```

---



# Chapter 3

## Results and Discussion

After creating cytoscouts, I used it to analyze the Eclectus Fly Interactome. This section is about what I found when I used cytoscouts to probe the interactome. I talk about the results of cytoscouts' analysis when I looked at the interactome in its uncollapsed state and what I found looking for Shot in the uncollapsed interactome. I also talk about cytoscouts results that led me to collapsing the interactome. Then I talk about what collapsing the interactome did to both the interactome and the results of the search for Shot. I talk about what did and *did not* show up in these searches for Shot. I discuss the results I got from using cytoscouts' output with the Gene Ontology tool PANTHER. Finally, I discuss the results and provide my interpretation of them.

Before discussing my results it is important to remember some of the shortcomings and assumptions of the interactome approach. The proteins are neighbors metaphorically, not literally. Two proteins being neighbors means that the proteins interact with each other, not that they share the same physical location. The nature of the interactions is also not discernible using the interactome. All one can know from the interactome results is that at some point in some circumstance the proteins are either observed to be capable of touching each other or are believed to do so from computational data.

### 3.1 The Uncollapsed Interactome

The uncollapsed interactome has 542,637 edges and contains 18,162 nodes. The histograms in Figure 3.1 show that the interactome has the characteristic of a scale free network, consisting of a few nodes with high degree and many nodes with low degree. In the histograms, with the X axis being node degree and the Y axis being node counts, the uncollapsed interactome (left) exhibits a sharp drop in node counts at the left side of the graph, meaning that the majority of the nodes in the interactome have a degree of less than 25. The log-log plot (bottom) makes it easier to see the overall trend of the plot. Having a roughly linear distribution in a log-log plot is another characteristic of a scale free network (Barabsi and Oltvai, 2004).

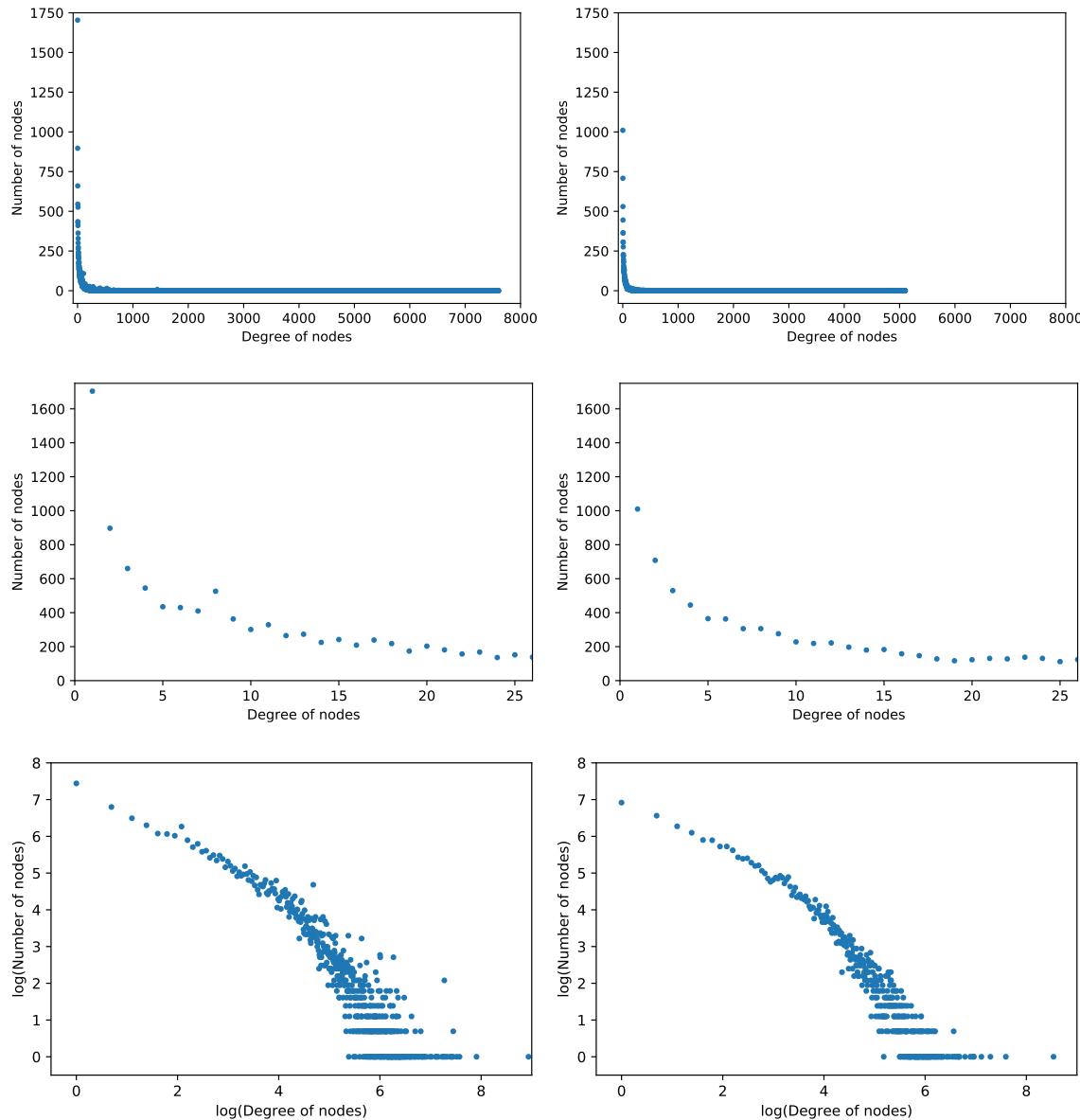


Figure 3.1: Histograms of the uncollapsed interactome on the left and the collapsed interactome on the right, by node degree (top), zoomed in to nodes with a degree of 25 or less (middle), and as a log-log transformation of the first pair of histograms (bottom).

### 3.1.1 Shot in the Uncollapsed Interactome

The UniProt ID for Shot isoform A, A1Z9J1, has 215 neighboring nodes. A full list of neighbors of Shot A with their common names and degrees is available online at [bit.ly/2JxJDpz](http://bit.ly/2JxJDpz).<sup>1</sup> These 215 neighboring nodes have 12,059 neighbors. That is to

<sup>1</sup>If for any reason in the future this bitly link or others like it should break, their destinations will still be online at <https://github.com/pwijngaard/cytoscouts/tree/master/results>.

say 66% of all nodes in the interactome are within two degrees of separation from Shot A. I did not find this result particularly useful, since in this interactome there are high-degree neighbors that link almost every protein. However, high degree of interconnectedness is another feature of a scale free network, so this statistic is further evidence that the interactome is scale free. Figure 3.2, top panel, is a histogram of the neighbors of Shot A by degree. The distribution has many nodes having a low degree and few nodes having a high degree, showing that Shot A’s neighbors also exhibit a scale free distribution.

In the neighbors of Shot A, 32 nodes had no characterization in the UniProtKB. Uncharacterized proteins are thought to exist based off of genomic, computational, or mass spectrometry data, but do not have explicit experimental evidence for their nature and function (Nadzirin and Firdaus-Raih, 2012). Uncharacterized proteins are where one would look for novel proteins. I present the list of these nodes in Table 3.1. These may be of interest for further research.

Table 3.1: The 32 uncharacterized nodes that interact with Shot A, A1Z9J1, by node degree.

UniProt ID	Common Name	Degree	UniProt ID	Common Name	Degree
Q9VPD6	CG5059	494	Q9VM67	CG18304	98
Q8IPU8	CG5059	493	Q9VFL0	CG3259	89
Q8IPU7	CG5059	490	Q9VP62	CG9389	85
Q9V677	CG8858	340	Q9VYP5	CG2750	79
Q9VTK9	CG6084	250	M9PHE1	CG2750	77
Q8IQF8	CG6084	249	Q9VMJ2	CG13996	70
M9PF61	CG6084	249	A0A0C4DHG7	CG42404	63
M9PI17	CG6084	249	Q9VFA8	CG42404	63
Q9VP84	CG10581	214	Q9VPK4	CG11454	61
Q7K2B0	CG7137	192	Q9VHB8	CG8199	61
Q7JUS1	CG30491	164	Q9W3X0	CG14441	59
Q9VR43	CG15625	156	Q7K1H9	CG10912	48
Q9VEF5	CG7379	136	Q9VJQ0	CG4168	48
M9NCS3	CG18304	98	Q9VQ11	CG31928	40
M9PCS5	CG18304	98	A1ZBK1	CG10051	32
M9PCE3	CG18304	98	Q9VCY2	CG33107	24

The top ten results for nodes that interact with Shot A can be seen in Table 3.2. These nodes have the highest degree of Shot A’s neighbor set, but that does not mean that they are necessarily more important or more strongly connected to Shot. Degree is simply a way of characterizing nodes that can be derived from the interactome itself.

Ubiquitin was the neighbor with the highest degree and can be seen in Figure 3.2 as the outlier on the far right. The high degree of Ubiquitin is not surprising given the

Table 3.2: Top ten UniProt IDs that interact with Shot A, A1Z9J1, by node degree.

UniProt ID	Common Name	Degree
P0CG69	Ubiquitin	7602
Q24568	Netrin B	1840
P18431	Shaggy	1502
O61307	Teneurin-M	1428
P34082	Fasciclin 2	1384
P92177	14-3-3epsilon	948
Q9W0L7	Ubiquitin-specific protease 10, isoform B	935
M9PDK3	Ubiquitin-specific protease 10, isoform B	932
Q9VLS7	Phosphatidic acid phospholipase A1, isoform B	908
M9MRD9	Phosphatidic acid phospholipase A1, isoform B	904

protein’s namesake ubiquity. Indeed, Ubiquitin has the highest degree in the entire interactome. In Figure 3.1 the protein furthest to the right in the top-left panel is Ubiquitin. Its high degree is a consequence of its role in protein degradation (Kleiger and Mayor, 2014). Ubiquitin-specific proteases act to remove Ubiquitin from proteins, so the high degree of Ubiquitin-specific protease 10 can be explained by that its function too (Komander, 2010). The two of these proteins together are evidence that Shot is regulated by Ubiquitination in some form, but given the high degree of Ubiquitin, most proteins are.

Netrin B, Teneurin-M and Fasciclin are proteins that regulate neuronal processes, and we would expect to see them on a list of Shot A’s neighbors as they are known to interact with Shot in its role in neuron development (Voelzmann et al., 2017). Shaggy is a kinase, 14-3-3epsilon is a signaling protein, and Phosphatidic Acid Phospholipase is an endoplasmic reticulum protein, according to their respective entries on FlyBase (Gramates et al., 2017). It is not surprising to see these proteins have a high degree given that they are either signaling proteins or involved with the endoplasmic reticulum, the “post office” of the cell.

Additionally, in Table 3.2 there are multiple UniProt IDs for proteins with the same common name. Ubiquitin-specific protease isoform B has two entries with different UniProt IDs, as does Phosphatidic acid phospholipase A1, isoform B. This is a consequence of the high degree of specificity that UniProt gives to each node. Due to this specificity, there is a risk that the neighbor counts, both in the number of neighbors and the degree of those neighbors, are inflated.

### Redundancies in the Uncollapsed Interactome

For example elsewhere in the interactome, Katanin p60 is in the interactome as a neighbor of Shot A twice, as Q9V989 from experimental evidence and as A0A0B4KFL3 from homology data. There are not two versions of Katanin p60 interacting with Shot A, but automatic data collection has put two entries for Katanin p60 in UniProtKB

and its curators have yet to remove the redundant copy. This result was the impetus for collapsing the interactome.

In some instances finding a redundant copy is less straightforward than the above example. An example of this problem is visible in Figure 3.2, in the top panel, where the outlier in the top left corner is due to there being 21 UniProt IDs for Shot with near identical neighbor counts. Table 3.3 shows those 21 neighbors in detail. Aside from Shot H, these Shot isoforms have identical neighbor counts. When I checked them with cytoscouts, the neighbors of these isoforms were identical to each other. In the case of Shot H, the three additional neighbors do not have common names that are unique from the other Shot isoforms. These isoforms should not be interacting with the same set of proteins because they have different functional groups.

However two things make it hard to rule these isoforms with identical neighbors out as redundant. The first is that I have not been able to find characterizations for most of these isoforms in the literature. The second reason is that due to an error on the part of the Eclectus interactome, none of Shot isoforms' neighbors are Actin. Shot binds directly to Actin through its calponin homology domains in its Actin binding domain, so the lack of an Actin interaction with any of the Shot isoforms is an omission in the interactome (Korenbaum and Rivero, 2002; Voelzmann et al., 2017). However, some isoforms, including some of the isoforms in Table 3.3, lack calponin homology domains and thus would not interact with Actin Based off of sequence alignments (available at [bit.ly/2HTuRfv](https://bit.ly/2HTuRfv)) with Clustal Omega, these isoforms differ mostly in their Actin binding domains (Sievers et al., 2011; Goujon et al., 2010). It may be possible that these isoforms' neighbor sets would differ if the Eclectus interactome correctly included interactions between Actin and Shot isoforms containing the Actin binding domain. That said, I interpret the fact that Shot H, which has a plakin repeat domain, has the same common name neighbors as Shot C, with no plakin domain, as being evidence that these nodes are redundant. Ultimately while I cannot rule out that these isoforms have the same number of neighbors, I find it more likely that the isoforms, despite being biologically distinct, were given redundant neighbor sets when they were entered in the interactome.

Another ambiguity appears at Table 3.2. Q9W0L7 and M9PDK3, the IDs for Ubiquitin-specific protease 10 isoform B share 100% local sequence identity with each other on UniProt (Bateman et al., 2017). However outside of that local alignment, they differ in amino acid length, and they are also listed as explicitly different isoforms: M9PDK3 is isform B and Q9W0L7 is isoform C. Similarly the entries for Q9VLS7 and M9MRD9 are for Phosphatidic acid phospholipase A1, and they also share 100% sequence identity with each other on UniProt but differ in amino acid length and they also are listed explicitly as different isoforms. I do not know what to make of this conflicting information, and I am unsure as to whether these are genuinely separate isoforms or redundant entries.

There are some definite instances of redundant entries and some instances where it is hard to tell if the entries are redundant or valid. I developed the collapse function to be used as an option in conjunction with the uncollapsed interactome knowing that it would erase some valid information. With these caveats in mind, I collapsed the interactome.

Table 3.3: Node A1Z9J1, Shot isoform A, has 21 neighbors named Shot in the uncollapsed interactome. These are all listed as separate isoforms in UniProt, but, except for node A1Z9J3, the isoforms have the same set of identical neighbors despite some isoforms missing functional domains, as evidenced by the difference in their number of amino acids. The degrees of each isoform are one higher in this table than they are elsewhere because this table includes interactions of the node with itself.

Entry	Protein Name	Length (a.a.)	Degree
A1Z9J3	Shot isoform H	8805	219
A1Z9J1	Shot isoform A	5501	216
A0A0B4K7Y9	Shot isoform AA	5408	216
A0A0B4K718	Shot isoform AB	4100	216
A0A0B4K788	Shot isoform AC	5462	216
A0A0B4K869	Shot isoform AD	5434	216
A1Z9I9	Shot isoform B	5390	216
A1Z9J2	Shot isoform C	5160	216
A1Z9J0	Shot isoform E	5201	216
A1Z9I8	Shot isoform G	5385	216
B7YZG3	Shot isoform I	5375	216
B7YZG4	Shot isoform J	5368	216
B7YZG2	Shot isoform K	5370	216
B7YZG5	Shot isoform L	5486	216
B7YZG6	Shot isoform M	5479	216
E1JH62	Shot isoform N	5388	216
E1JH63	Shot isoform O	5499	216
E1JH65	Shot isoform P	5458	216
E1JH64	Shot isoform Q	5155	216
A0A0B4K867	Shot isoform Y	5463	216
A0A0B4K7P8	Shot isoform Z	5394	216

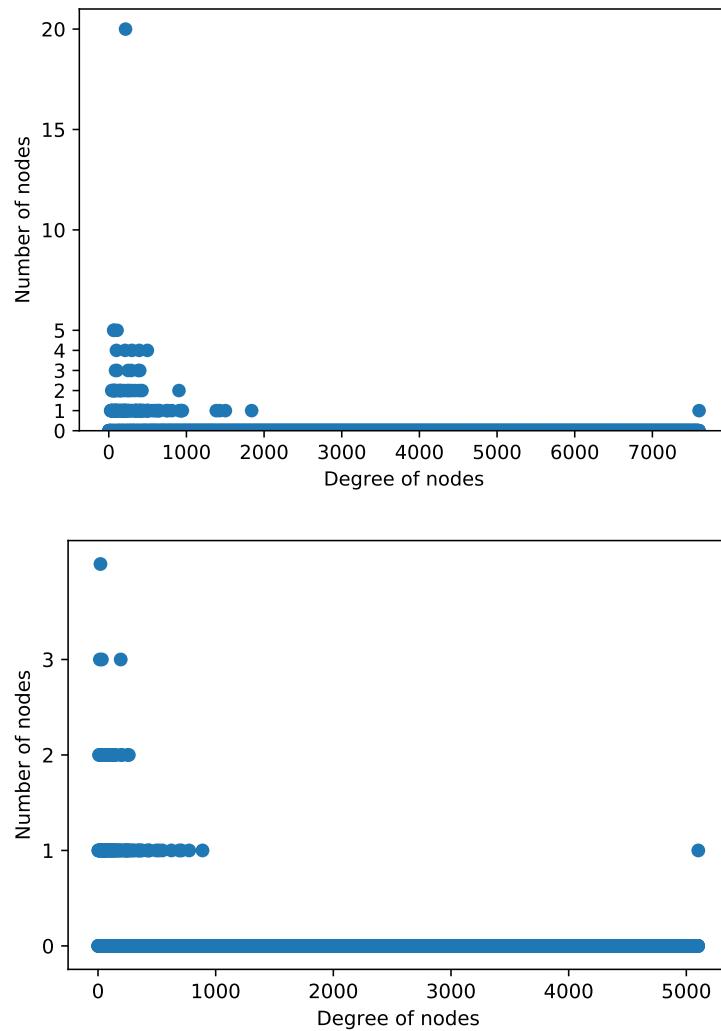


Figure 3.2: Top: A histogram of the all the neighbors of node Shot A, A1Z9J1, in the uncollapsed interactome by node degree. Bottom: A histogram of the all the neighbors of node “Shot” in the collapsed interactome by node degree.

## 3.2 The Collapsed Interactome

Cytoscouts collapsed the interactome by taking every set of UniProt nodes that shared a common name and combining them into one common name node. For example, in Table 3.3 the 21 nodes of each Shot isoform become one node, “Shot.” It did this by taking the interactome that was stored as a list of lists and turning it into a set of tuples. Since entries in a set only count once, this method discarded nodes with duplicate common names. Cytoscouts skipped 5,955 edges (2.1% of the total edges) due to there not being a common name listed for at least one of the nodes. Cytoscouts did not add the skipped edges to the collapsed interactome and discarded the information in those edges. The collapsed interactome has 282,962 edges and 11,376 nodes, making it 51% of the size of the uncollapsed interactome by edge count and 63% the size of the uncollapsed interactome by node count. Despite the reduction in size, the collapsed interactome retains its scale-free character, based on comparison of the two histograms in Figure 3.1.

### 3.2.1 Shot in the Collapsed Interactome

The node with the common name “Shot” in the collapsed interactome has 127 neighboring nodes. Figure 3.3 is a visual representation of Shot and its 127 neighbors made in GraphSpace. Of these, 105 are named and 22 are uncharacterized. Table 3.5 is a list of the named neighbors of Shot. In the Appendix, Table B.1 is the list of named proteins along with descriptions taken from the literature.

The name and proportion of degrees of the neighbors in the collapsed interactome are similar to those of the neighbors in the uncollapsed interactome, with the absolute degree counts being approximately 50%-60% less. For example, compare the top ten highest common name node neighbors of Shot in Table 3.4 with the top ten highest degree Shot node neighbors of A1Z9J1 in Table 3.2. The top five nodes in both tables are the same proteins with the same order with degrees that are approximately 50%-70% smaller. The addition of Sirt7 and Cdc5 to Table 3.4 are from the space freed up by removing common name duplicates, and these proteins were just outside of the top ten in the uncollapsed interactome. The bottom panel of Figure 3.2 shows the histogram of these neighbor nodes. As a consequence of the collapse of the interactome, the cluster of Shot nodes in the top panel of Figure 3.2 has disappeared. Aside from this, the distribution of Shot’s neighbors remains scale-free in the collapsed interactome. The far right outlier is again Ubiquitin. While collapsing removes clusters of equivalent nodes, taken together the data suggest that the general character of the neighbor set survived collapsing.

Table 3.4: Top ten common name neighbor nodes of Shot by node degree.

Common Name	Degree
Ubiquitin-p63E	5102
NetrinB	888
Shaggy	775
Teneurin-M	709
Fasciclin 2	691
Ubiquitin-specific protease 10	624
14-3-3epsilon	548
Phosphatidic acid phospholipase A1	518
Sirt7	495
Cdc5	436

Table 3.5: The 105 characterized common-name proteins neighboring Shot in the interactome. Proteins that are identified in the literature in the background of this thesis are marked in **bold**.

14-3-3epsilon	Ci	<b>Kat80</b>	Nrx-1	Sals	Thor
Abi	Cold	Kdn	Nrx-IV	Scny	Tip60
Alc	Dally	<b>Kra</b>	P115	Sec3	TwdlB
Alpha-Spec	Dap160	Kst	P32	Sema-1a	TwdlD
<b>AlphaTub84B</b>	DCTN1-p150	Ktub	Pak	Sema-2a	TwdlF
<b>alphaTub84D</b>	Der-2	Lap1	PAPLA1	Set1	TwdlJ
<b>Apc</b>	Dysb	Lsd-2	<b>Patronin</b>	<b>Shaggy</b>	TwdlN
Arm	<b>Eb1</b>	Mad	PKD	<b>Shot</b>	TwdlO
ASPP	UIF-2beta	Mats	PUf68	Sirt7	U2af50
<b>BetaTub56D</b>	ERR	Mi-2	Rab21	Skd	<b>Ubi-p63E</b>
<b>BetaTub60D</b>	<b>Fas2</b>	Mms19	Rcd5	Snapin	UGP
Bgb	Flw	Mrt	Rel	Snoo	<b>Usp10</b>
Bin1	Gint3	Msn	RnpS1	SoxN	Vha55
Bro	Golgin245	Mtpalpha	RpL10Aa	Srp72	Wde
Bsk	Hpo	N	RpL5	Stan	Zormin
Cam	Hrs	<b>NetB</b>	Rpn5	Tango6	
Cbs	Jra	Nf-YC	RpS5b	<b>Ten-m</b>	
Cdc5	<b>Kat60</b>	Nrg	Rtnl1	Thiolase	

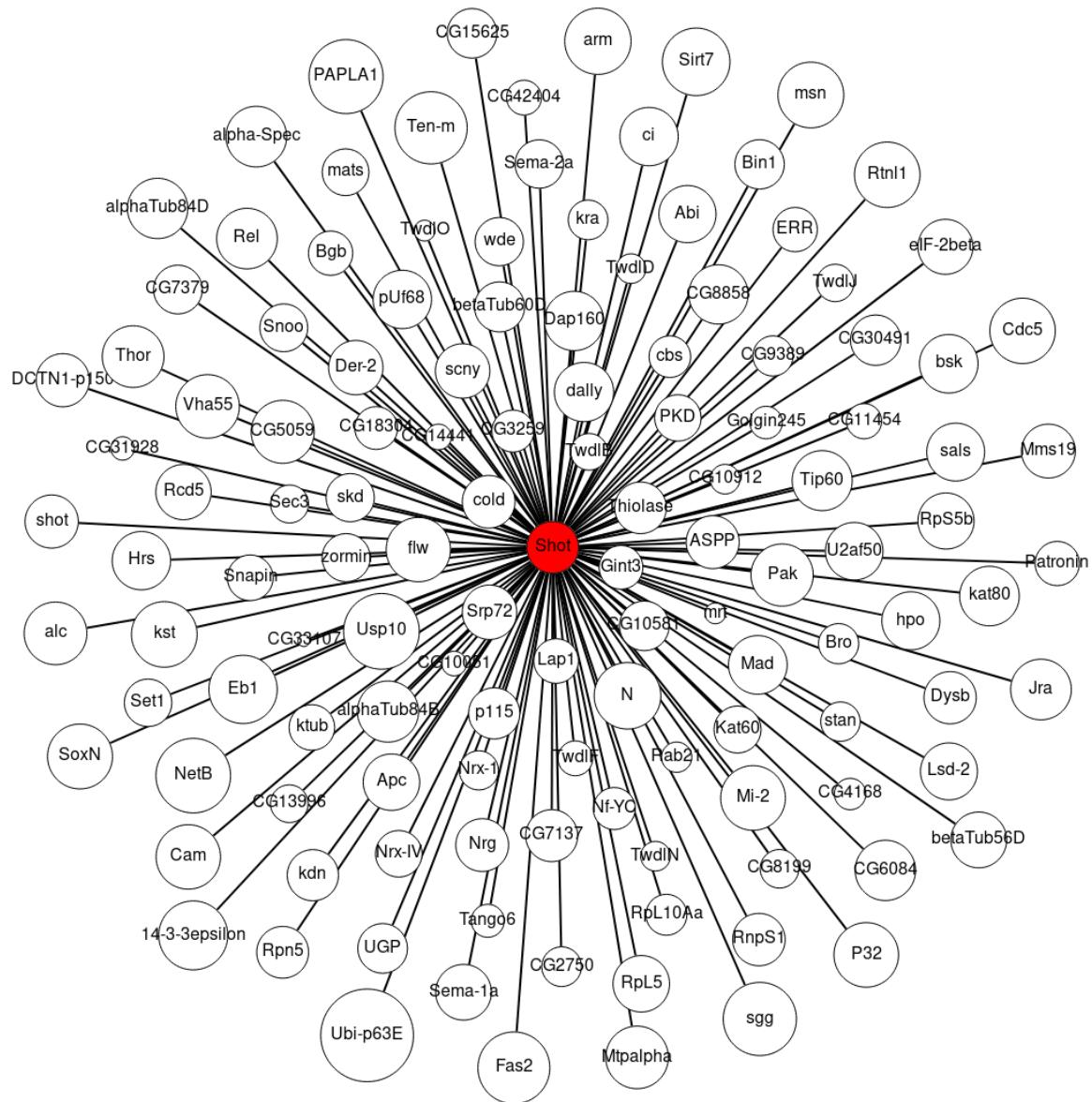


Figure 3.3: A PathLinker diagram of Shot and its 127 neighbors in the collapsed interactome, with node size proportional to degree. All names for proteins in the figure should be capitalized but the common names in the dictionary were not consistently capitalized.

### 3.3 The Gene Ontology

Once I had the neighbors of Shot, I used the Gene Ontology to generate a list of annotations for the neighbors. To study Shot using the Gene Ontology, I took the UniProt nodes from the output of the uncollapsed interactome search for neighbors of Shot A and pasted it into PANTHER, an online classification system for high throughput analysis. Only the uncollapsed interactome could be evaluated with the Gene Ontology because PANTHER does not accept common names as input. In addition to providing a protein classification, PANTHER annotates proteins using the Gene Ontology terms of molecular function, cellular component and biological process (Mi et al., 2013). Molecular function is the activity performed by the protein. Cellular component is the location where the protein performs its molecular function. Biological process is the system to which a protein contributes.

PANTHER was able to classify 118 of the 215 neighbors of Shot A and provided available annotations for protein class, molecular function, cell component and biological process. Table 3.6 is a list of all the proteins PANTHER classified. An abridged table of the 118 nodes and their Gene Ontology classifications is in the Appendix in Table D.1. The complete table with all Gene Ontology annotations is available for download to be viewed in Excel or another spreadsheet program at [bit.ly/2Hxb3vJ](https://bit.ly/2Hxb3vJ). I found keeping the full table open useful when working with Gene Ontology results, because I could quickly reference which protein belonged to what Gene Ontology term and vice versa.

Table 3.7 is a list of all the UniProt IDs PANTHER was unable to classify along with their common names, and the majority of these are either redundant entries or isoforms. There were four named proteins collected in the collapsed interactome that were not classified by PANTHER in Table 3.6. Those proteins are Dysb, Lsd-2, mrt and Sals. I identify them here according to their entries in the *D. melanogaster* resource FlyBase (Gramates et al., 2017). Dysb is a presynaptic neuronal protein that regulates the release of vesicles. Lsd-2 associates with lipid droplets and regulates lipid storage. Mrt has no data. Sals is an Actin binding protein that promotes elongation of Actin filaments.

The results of the PANTHER search are visualized in Figure 3.4 for protein class and Figure 3.5 for molecular function. These data represent a broad survey of the terms found for the 118 neighbor nodes for the kinds of proteins that the neighbors are and what kinds of interactions take place with them. While not quantitative, the data give an overview of the character of Shot A's neighbors.

To quantitatively assess Shot A's neighbors, I performed an overrepresentation test for the Gene Ontology terms for biological process and molecular function, seen in Tables 3.8 and 3.9 respectively. An overrepresentation test checks to see if any terms show up more than they would by chance given the distribution of terms for *D. melanogaster*. Count is the number of proteins with a given classification observed in the neighbor set. The expected count is the number of proteins from the neighbor set one would expect to see with that Gene Ontology term if the neighbor set was randomly taken from all proteins in the PANTHER *D. melanogaster* dataset. Enrichment is the number of times that the observed count is seen divided by the expected

Table 3.6: The common names of the 118 proteins neighboring Shot recognized by PANTHER. Minus the 17 uncharacterized proteins, there are 101 proteins in this list. The proteins from Table 3.5 not in this list are Dysb, Lsd-2, mrt, and Sals. GI is DCTN1-p150 by another name.

14-3-3epsilon	CG10051-RA	Eb1	Mi-2	RnpS1	Tango6
142336_at	CG11454	EIF-2beta	Mms19	RpL10Aa	Ten-m
34F4T	CG13996-RA	ERR	Msn	RpL5	Thiolase
Abi	CG15625	Fas2	Mtpalpha	Rpn5	Thor
Alc	CG30491	Flw	N	RpS5b	Tip60
Alpha-Spec	CG31928	Gint3	NetB	Rtnl1	TwdlB
AlphaTub84B	CG3259-RA	Gl	Nf-YC	Scny	TwdlD
AlphaTub84D	CG4168	Golgin245	Nrg	Sec3	TwdlF
Apc	CG42404	Hpo	Nrx-1	Sema-1a	TwdlJ
Arm	CG7137	Hrs	Nrx-IV	Sema2a	TwdlN
ASPP	CG7379	Jra	P115	Set1	TwdlO
BetaTub56D	CG8199	Kat60	P32	Shaggy	U2af50
BetaTub60D	CG8858	Kat80	Pak	Shot	Ubi-p63E
Bgb	Ci	Kdn	PAPLA1	Sirt7	UGP
Bin1	Cold	Kra	Patronin	Skd	Usp10
Bro	Dally	Kst	PKD	Snapin	Vha55
Bsk	Dap160	Ktub	PUf68	Snoo	Wde
Cam	Der-2	Lap1	Rab21	SoxN	Zormin
Cbs	DmelCG10581	Mad	Rcd5	Srp72	
Cdc5	DmelCG5059	Mats	Rel	Stan	

count. The raw P value is the probability that one would see the observed count for that term due to chance, given the size of the neighbor set. The False Discovery Rate is the P-Value adjusted for the expected number of false positives (Benjamini and Hochberg, 1995). Gene Ontology terms with low false discovery rates suggest that Shot has a role involving those terms. The Gene Ontology terms with the lowest false discovery rates in Tables 3.8 and 3.9 were transcription factors. I compiled a list which of Shot A's neighbors PANTHER classified as transcription factors (Table 3.10). Transcription factors had the most strongly overrepresented Gene Ontology terms in Shot A's neighbors, and I made a list of the neighbors with those terms.

Since the Gene Ontology is a specific and structured vocabulary, I listed the Gene Ontology definitions of each term used in Gene Ontology results in Appendix C. Definitions for protein class and molecular function using the Gene Ontology resource's own words for protein class can be found in Table C.1 for Figure 3.4. Specific Gene Ontology definitions for molecular function are in Table C.2 corresponding to Figure 3.5 and Table 3.9. Table C.3 gives the Gene Ontology definitions for the biological process terms used in Table 3.8.

Table 3.7: The 97 UniProt IDs that were unable to be classified by PANTHER and their common names, sorted by common name. Most of these unclassifiable IDs had an ID with the same common name classified by PANTHER, but the ones that did not are highlighted in **bold**.

UniProt ID	Common Name	UniProt ID	Common Name	UniProt ID	Common Name
Q9Y0S9	Abi	X2JKM9	Kat80	<b>Q9VGN5</b>	<b>Sals</b>
A0A0B4KH51	Abi	Q9VZQ3	Kst	E1JH64	Shot
Q95SG9	Alc	Q7KV70	Kst	E1JH62	Shot
A0A126GV30	Apc	M9PBL6	Kst	E1JH63	Shot
Q9W2J2	ASPP	Q7KV69	Kst	E1JH65	Shot
A0A0B4KF17	ASPP	<b>Q9VXY7</b>	<b>Lsd-2</b>	B7YZG6	Shot
A0A0B4K775	Cbs	<b>Q9VB49</b>	<b>Mrt</b>	B7YZG5	Shot
A0A0B4K864	Cbs	Q7KV90	Msn	B7YZG4	Shot
Q7K1H9	CG10912	Q7KV88	Msn	B7YZG3	Shot
<b>Q9W3X0</b>	<b>CG14441</b>	Q7KV89	Msn	A0A0B4K867	Shot
M9NCS3	CG18304	M9PH19	Msn	A0A0B4K869	Shot
M9PCS5	CG18304	M9PBJ7	Msn	A0A0B4K788	Shot
M9PCE3	CG18304	Q8IPE8	Mtpalpha	A1Z9J1	Shot
Q9VM67	CG18304	A0A0B4KGQ3	Nrx-1	A1Z9J0	Shot
Q9VYP5	CG2750	Q9VCZ9	Nrx-1	A1Z9J2	Shot
M9PHE1	CG2750	Q3KN41	Nrx-1	A0A0B4K7Y9	Shot
Q9VCY2	CG33107	A0A0B4KH61	Nrx-1	A0A0B4K7P8	Shot
A0A0C4DHG7	CG42404	Q9VI13	Pak	B7YZG2	Shot
Q8IPU8	CG5059	M9MRD9	PAPLA1	A0A0B4K718	Shot
Q8IPU7	CG5059	X2JDJ3	PAPLA1	A1Z9I8	Shot
Q9VTK9	CG6084	Q9VE91	PKD	A1Z9I9	Shot
Q8IQF8	CG6084	A0A0B4KGF0	PKD	Q9VLH8	SoxN
M9PI17	CG6084	Q7PLE8	Rab21	A5XCL5	UGP
Q8INU2	Dap160	Q9VMW4	Rtnl1	Q9VSW2	UGP
Q9VIF7	Dap160	Q7KTP4	Rtnl1	E1JI91	UGP
Q1WWC9	Dap160	Q9VMV9	Rtnl1	M9PDK3	Usp10
M9PDC7	Dap160	E1JHT9	Rtnl1	Q9W053	Zormin
M9PBF6	Dap160	Q9VMW2	Rtnl1	Q0E8J3	Zormin
<b>Q9VVT5</b>	<b>Dysb</b>	Q9VMW1	Rtnl1	M9PE03	Zormin
Q9XZ57	Eb1	<b>Q9VGN4</b>	<b>Sals</b>	Q0E8J4	Zormin
Q7JZD3	Eb1	<b>A0A0B4K633</b>	<b>Sals</b>	Q0E8J5	Zormin
Q9VSE9	ERR	<b>A0A0B4LI42</b>	<b>Sals</b>		
Q9VN89	Kat60	<b>Q58CJ5</b>	<b>Sals</b>		

Table 3.8: Abridged PANTHER overrepresentation test for the 118 classified neighbors of A1Z9J1 by Gene Ontology terms for biological processes, ordered by False Discovery Rate (lower is better). This tests to see if any biological process terms are occurring more often in this subset than they would according to chance given the distribution of Gene Ontology terms for biological process in *D. melanogaster*.

PANTHER GO-Slim Biological Process	Count	Expected Count	Enrichment	Fischer's Test Raw P Value	False Discovery Rate
Regulation of Transcription from RNA Polymerase II Promoter	10	1.93	5.18	3.16E-05	0.00238
Cytoskeleton Organization	8	1.24	6.48	4.56E-05	0.00258
Cell Cycle	12	3.09	3.89	7.65E-05	0.00346
Nervous System Development	5	0.57	8.83	0.00034	0.00961
Cell Differentiation	6	0.95	6.3	0.000483	0.0109
Apoptotic Process	5	0.69	7.2	0.000822	0.0143
Cellular Component Movement	7	1.59	4.41	0.00126	0.0178
Cellular Component Morphogenesis	5	0.81	6.14	0.00162	0.0215
Cytokinesis	4	0.55	7.29	0.00266	0.0273
Locomotion	4	0.53	7.52	0.00239	0.0284
Carbohydrate Metabolic Process	7	1.99	3.52	0.00431	0.0423
MAPK cascade	4	0.65	6.14	0.00478	0.045

Table 3.9: Abridged PANTHER overrepresentation test for the 118 classified neighbors of A1Z9J1 by Gene Ontology terms for molecular function, ordered by False Discovery Rate (lower is better). This tests to see if any molecular function terms are occurring more often in this subset than they would according to chance given the distribution of Gene Ontology terms for molecular function in *D. melanogaster*.

PANTHER GO-Slim Molecular Function	Count	Expected Count	Enrichment	Fischer's Test Raw P Value	False Discovery Rate
Transcription Cofactor Activity	6	0.85	7.07	0.00027	0.016
Structural Constituent Of Cytoskeleton	6	1.09	5.51	0.00095	0.0241
Kinase Activity	9	2.51	3.58	0.00109	0.0243
Nucleic Acid Binding	16	7.02	2.28	0.0025	0.0494

Table 3.10: PANTHER classified these Neighbors of Shot A as transcription factors.

Classified by biological process	Classified by molecular function
Arm	Arm
Bgb	Bgb
Bro	Bro
Mad	CG7379
Mms19	Mms19
Nf-YC	Skd
Rcd5	
SAP18	
Snoo	
Tip60	

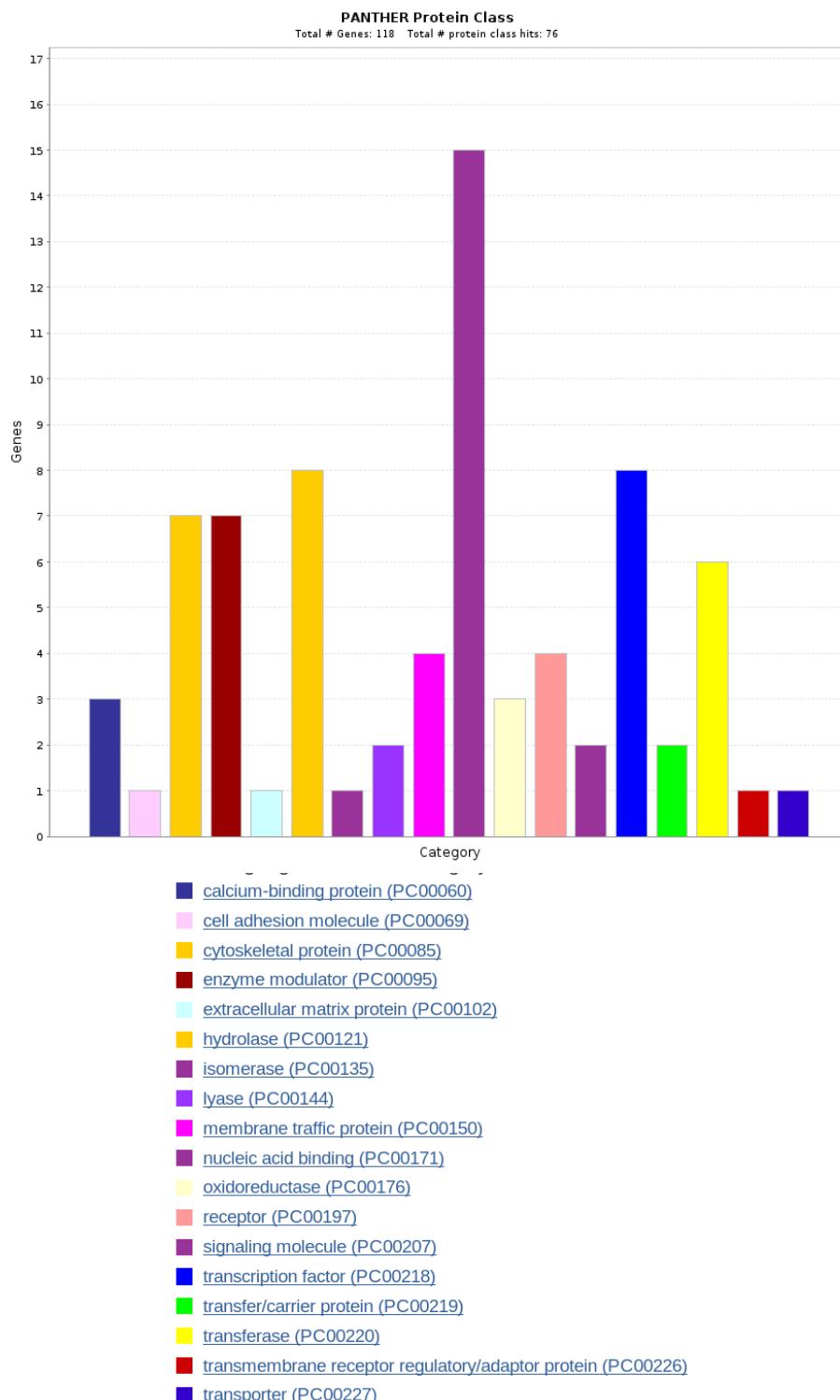


Figure 3.4: PANTHER Gene Ontology results for 118 nodes, for protein class. 97 nodes were unable to be classified.

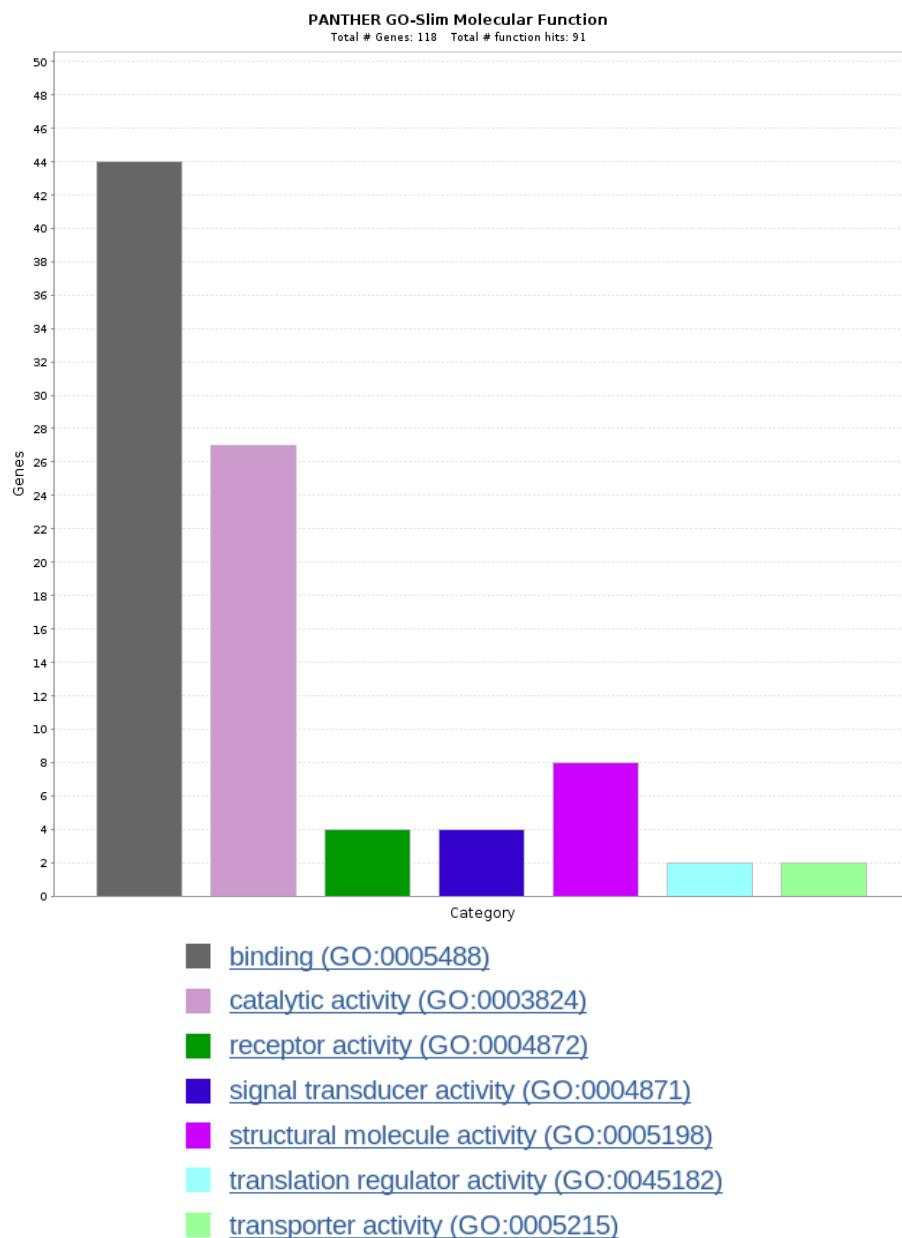


Figure 3.5: PANTHER Gene Ontology results for molecular function for 118 nodes. 97 nodes were unable to be classified.

## 3.4 Discussion

There are salient points in the data from cytoscouts that I want to discuss. Some insights are from the interactions in the interactome in itself, and some are from the Gene Ontology classifications of the proteins in the interactome. Together they describe the Eclectus interactome and Shot, and they hint at new avenues for research.

The biggest result from cytoscouts is that no Actin proteins show up as neighbors of Shot in either the uncollapsed or collapsed interactome. Actin proteins were in the interactome when I searched for them, but they were not recorded as having an interaction with any kind of Shot. This is surprising because Shot is documented to have a direct interaction with Actin by its calponin homology domains in the literature (Korenbaum and Rivero, 2002). The Eclectus interactome has a clear omission. The absence of a Shot-Actin interaction in the interactome is a reminder that interactomes are not exhaustive and are not replacements for experiments or research.

Actin was not the only protein I was expecting to see that did not appear in the list of neighbors. Some regulatory mechanisms laid out in the background, Fak/Src, Par1 and HectD1, did not show up in the interactome. I had hoped to see FAK/Src in the results as an interolog of the interaction FAK/Src has with ACF7, but FAK/Src is not among the neighbors. Although given that Wu *et al.* observed it to have no interaction with ACF7 in regards to cell migration, it may be its appearance in the interactome would not be related to cell migration (2008). Par1's regulation of Shot is unknown and it may not have a direct interaction with Shot. The purported *D. Melanogaster* homolog of HectD1, CG5604, is also absent from the list (Ketosugbo *et al.*, 2017). These absences, however, do not contradict experimental data, and they do not undermine my confidence in the Eclectus interactome like the absence of Actin does.

While the above proteins were not listed as neighbors, there were some regulatory candidates that did show up. Shot interacting with itself may be evidence for the autoinhibition model. Shaggy, also known as Zw3, was on the list of Shot neighbors, too. Although the fact that Applewhite *et al.* overexpressed Shaggy and did not observe any affect on Shot localization or dynamics means that this interaction may not have to do with migration (2013).

Two neighbors I was surprised to find were Mtpalpha and Thiolase. Mtpalpha and Thiolase are both proteins that exist inside the mitochondrion (Kishita *et al.*, 2012). The inside of the mitochondrion is a specialized environment generally kept isolated from the rest of the cell (Chipuk *et al.*, 2006). One explanation for these interactions is that they are simply a false positive. Maybe these proteins have other roles outside the mitochondrion. Interactions between Shot and Mtpalpha or Thiolase may be chemically possible but prohibited *in vivo* due to the separation of the two environments. I do not have anything more than speculation, but this result has captured my imagination. It makes me want to know what the source of these interactions are in the interactome, since that would give me more of an answer on if the interactions are noise rather than signal.

The finding I am most excited about is the interaction between and APC1. In

the literature, Shot colocalizes with EB1 and APC1 at the muscle-tendon junction in developing flies, where they promote microtubule assembly (Subramanian et al., 2003). APC1 is the *D. melanogaster* equivalent of mammalian APC (Gramates et al., 2017). APC is required for ACF7 to guide microtubules to focal adhesions and is regulated by ErbB2 (Zaoui et al., 2010). ErbB2's *D. melanogaster* equivalent is dEGFR (Alvarado et al., 2009). It would be interesting to see if a homolog of the ErbB2/APC/ACF7 regulation pathway existed in *D. melanogaster* for Shot.

The Gene Ontology proved to be a useful toolset for interpreting the interactome. With so much data, being able to get descriptive terms for all of Shot A's neighbors instantly was helpful in stimulating my thinking. At the same time, the amount of information the Gene Ontology supplies can be overwhelming, and the automated nature of its annotations can sometimes be arbitrary or misleading.

One frustration in working with the Gene Ontology was that it fell short of achieving its goal of a universal database of terms and annotations. I wanted the online table at [bit.ly/2Hxb3vj](http://bit.ly/2Hxb3vj) to act as a searchable companion to this thesis. However, I noticed some limitations of Gene Ontology while working with this thesis and the online table. The Gene Ontology annotation within PANTHER was not consistent across PANTHER's features. For example, searching the online table for the term "MAPK Cascade," or its GO code (GO:0000165), should yield four results consistent with Table 3.8, but the search only yields three because annotations for the Gene Ontology terms by PANTHER are not consistent. Still, even with the flawed tools I had, I found having the online table open to search for terms useful when interpreting the results in Tables 3.8 and 3.9.

PANTHER only recognized the Shot H isoform, which surprised me, since Shot A is the most commonly discussed isoform in the literature (Voelzmann et al., 2017). Shot H is 3,300 amino acids longer than Shot A, but outside of that the two proteins are 98% identical according to BLAST (Altschul et al., 1997). The difference in size is due to Shot H's plakin repeat region. My best guess is that the H isoform was used over the A isoform because of the presence of the plakin repeat region, making Shot H the most "complete" isoform for use in databases.

In Figure 3.4, excluding transcription regulation, the overrepresented terms in Shot's neighbor set in Tables 3.8 and 3.9 are not surprising. The terms are either directly related to the cytoskeleton or related to processes that the cytoskeleton is involved with. The exception is carbohydrate metabolic process, but given that the proteins annotated as being involved in carbohydrate metabolic process all have other annotations that are also in Table 3.8 and the relatively high false discovery rate, I am inclined see the presence of carbohydrate metabolic process as an artifact rather than a novel role for Shot.

In Figure 3.4, I was surprised to see that the most common two protein classes for Shot A's neighbors were nucleic acid binding and transcription factor. Table 3.8 and Table 3.9 show that transcription regulators are overrepresented in Shot A's neighbors. The neighbors of Shot A with transcription factor annotation are listed in Table 3.10. Transcription factors are found in the nucleus, a specialized environment separate from the rest of the cell (Wilson and Berk, 2010). The overrepresentation implies either that Shot is having these interactions in the nucleus, that the transcription

factors are interacting with Shot before they are imported into the nucleus or that the transcription factors serve some other function in the cytosol. I found evidence in the literature that Shot associates with the nucleus but only around the nucleus and not within it (Wang and Volk, 2015; Wang et al., 2015). I would be interested in seeing an experiment to determine if Shot localizes inside the nucleus or if Shot is interacting with these proteins in the cytosol.

There are three things that I take away from the results: one is a limitation of the interactome and the other two are interesting avenues for further research. The first is that interactomes are liable to omit important data. The absence of Actin's interaction with Shot is a reminder of the limitations of interactomes. The other two are potential avenues for further research. Searching the interactome made me aware of the potential for APC1 to regulate Shot. The interactome also made me aware that Shot is interacting with transcription factors, and I want to see an experiment that determines if that means Shot is in the nucleus. Despite the limitations of the Eclectus interactome, I am glad that I chose to use it as a research tool.

# Conclusion

I started this thesis wanting to learn more about Shot and cell migration. Applying cytoscouts to the Eclectus Fly Interactome gave me a place to start: motivation for an experiment to see if APC1 regulates Shot-mediated focal adhesion dynamics. Along the way, I learned about interactomes, computer programming and bioinformatics tools.

I learned a lot about the Eclectus interactome itself. The uncollapsed Eclectus interactome has 542,637 edges and 18,612 nodes. In the uncollapsed interactome the node for Shot isoform A has 215 neighbors. 32 of those nodes are uncharacterized. For 118 of those neighbors, PANTHER provided Gene Ontology classification for the proteins' function and class. PANTHER failed to provide results for 97 nodes. Note that PANTHER is capable of classifying uncharacterized proteins through phylogeny and orthologs so there is overlap between the 32 uncharacterized nodes and the 118 classified nodes. Collapsed, the Eclectus interactome has 282,962 edges and 11,376 nodes. In the collapsed interactome the node Shot has 127 neighbors. Collapsing the interactome is meant to solve the problem of redundant entries in the UniProt knowledgebase and Eclectus interactome. The collapsed interactome looks to be similar to the uncollapsed interactome in character, just with around 60% of the entries. I have made the assumption that the duplicate nodes in the uncollapsed interactome are redundant. The creators of UniProtKB acknowledge that redundancy is a problem in their database, and I have anecdotal evidence that there are redundant entries in the uncollapsed interactome by looking at a few cases by hand, *e.g.* Katanin p60 or the 21 Shot nodes. However, I have no systematic rigorous method for evaluating the redundancy of a node. Right now I treat the two interactomes as a bracketing tool. For the neighbors of a protein in an interactome, there are no fewer nodes than the number of nodes in the collapsed interactome, and there are no more nodes than the number of nodes in the uncollapsed interactome. The uncollapsed interactome is the upper bracket bound, and the collapsed interactome is the lower bracket bound. The "true" number of nodes lies somewhere between the two but no further information is available about where in between specifically. I do not know if this is an appropriate method, or if doing things this way curtails the usefulness of a computational approach. I hope that the question of the number of nodes in the interactome can be resolved and the Eclectus interactome can be characterized more accurately.

I also learned a lot about computer programming. When I started this thesis, I was a complete beginner to the field and knew nothing about Python. Now I have a Python script, cytoscouts, and enough Python expertise that I can continue to work

on it. I want to improve cytoscouts by adding the ability to add the sources for each interactome edge into cytoscouts' results. The interactomes I have encountered all use the first two columns to contain nodes and edges, which makes it simple for cytoscouts to read different interactomes. Many of the interactomes also contain information on the source of the edge. It is more complicated to get that information from the script, because the column in which that information is stored varies from interactome to interactome. The next step for cytoscouts is to develop a function that can find and read the source information column. Another function I want to add to cytoscouts is the ability to quantify the degree of a node relative to the degrees of the other nodes in the interactome. Currently, one can roughly determine the relative degree of a node by taking its degree and comparing it to the histogram of the interactome, but this is awkward and imprecise. I want to expand cytoscouts by adding this statistic automatically to its output. Additionally, when collapsing the interactome, I want to give cytoscouts a quantitative method to check if the entries removed during the collapsing process were truly redundant and if the two interactomes are truly similar in all regards except for size. Reflecting on my experience working with various bioinformatics tools with this thesis, I would also like to add a feature to more quickly determine which proteins in the Gene Ontology results are not contained in the common names of the interactome and vice versa. Doing this manually was time consuming, and I am sure that the process can be automated. Now that I have been introduced to programming, I look forward to learning more and improving cytoscouts.

Finally, I learned about the limitations and benefits of working with machine generated annotation tools like UniProt and FlyBase. I find the biggest problem is that the entries lack context. The FlyBase entry for APC1 describes it as a member of the  $\beta$  Catenin destruction complex, and this is true, but that is not the only function of APC1 in the cell. APC1's interactions with the cytoskeleton are left out of FlyBase's annotation. If I had relied solely on automatic annotation, I would have missed out on an critical part of my thesis. I wonder what other details are left out by the annotations, and I worry that the opaque nature of machine generated annotation is more confusing to the reader than it is informative. At the same time, I would never have thought to investigate APC1 by hand had it not shown up in the interactome. Being able to instantly generate annotations for hundreds of proteins is convenient and a useful platform for getting started with research, but clearly nothing can replace manual research of the literature. I would like to see what is uncovered by simply going through all of Shot's neighbors in the interactome by hand.

I learned a lot in this thesis and now it is done. I hope that someone else will continue the work I began here. For someone else's thesis, I think the work I have done here could be extended by investigating if Egfr controls APC1-mediated regulation of Shot in *D. melanogaster* like Erb2B does for APC-mediated regulation of ACF7 in mammals. Another thesis project that could stem from the work I have done here is to test to see if Shot is present inside the nucleus and determine the nature of the interactions it has with nuclear proteins. My sincere wish is that I have gathered enough information to provide a useful roadmap to other students who want to study Shot for their senior thesis.

# Appendix A

## RhoGTPases

The movements of the four phases of cell migration require consistent, specific and timely regulation to produce migration. The RhoGTPase family of proteins is the element thought to provide that regulation. I would like to introduce the protein families that regulate RhoGTPases before discussing Rac1, RhoA and Cdc42.

The RhoGTPases Rac1, RhoA and Cdc42 together coordinate cell adhesion and migration during lamellipodium-driven migration. Additionally there are several other RhoGTPases both upstream and downstream of Rac1, RhoA and Cdc42 that also affect or are speculated to affect the mechanics of lamellipodium-based migration (Ridley, 2015)

### The Regulation of RhoGTPases

GTPases signal when bound to guanidine triphosphate, but a GTPase has the intrinsic ability hydrolyze their bound GTP into GDP, guanine diphosphate. That is to say that a GTPases can act as a regulator as long as it is GTP bound. If something were to activate its ability to turn GTP into GDP, it would turn itself off. It would then need another GTP molecule before it can act again as a regulator.

This characteristic is the mechanism by which RhoGTPases are controlled. The proteins that regulate RhoGTPases are the RhoGEF protein family, the RhoGAP protein family, and the RhoGDI protein family:

- RhoGEFs, guanine nucleotide exchange factors, activate GTPases by exchanging an inactive GTPase's GDP with GTP.
- RhoGAPs, GTPase activating proteins, silence GTPases as regulators by triggering their intrinsic GTPase activity.
- RhoGDIs, Guanine dissociation inhibitors, maintain stable inactive GTP ready for use by GTPases.

GEFs can be thought of as the "on switch" and GAPs as the "off switch," and GDIs ensure that there is enough GTP available to make this system work.

GEFs interact with a diverse multitude of protein-protein regulatory interaction networks. There are approximately four times the number of RhoGEFs than there are RhoGTPases, and GEFs that interact on the same GTPase can be regulated by different signaling pathways. Additionally, GEFs can act on multiple GTPases

selectively at the same time. Different GEFs can even act on the the same GTPase at the same time (Goicoechea et al., 2014).

### **Rac1, RhoA and Cdc42 in Cell Migration**

At the start of cell adhesion, Rac1 is high and RhoA is low. At the end of cell adhesion, Rac1 is low and RhoA is high. GEFs and GAPs regulate and coordinate this biphasic reaction (Lawson and Burridge, 2014). Rac1 and Cdc42 are high at the leading edge of the migrating cells, and promote protrusions and adhesions (Etienne-Manneville, 2013). Rac1 requires Paks to effect changes in Actin and microtubules. There are coordinated Actin and microtubule dynamics in the leading edge of migrating cells and a positive feedback loop between Rac1 and microtubules (Wittmann et al., 2003)

# Appendix B

## Common Name Protein Description

Table B.1: Descriptions of the 105 named neighbors of Shot in the collapsed interactome. Descriptions are taken directly from either FlyBase (Gramates et al., 2017) or UniProt (Bateman et al., 2017) when available. If neither resource has a description entry, descriptions are taken from the literature if available. If no literature is available, descriptions are taken from computational data on FlyBase.

Name	Definition
14-3-3epsilon	14-3-3 $\epsilon$ is an acidic protein that preferentially heterodimerizes with other members of the family but also can homodimerize. It functions in multiple signaling pathways, most prominently in the Ras/MAPK cascade. It is involved in embryonic hatching, germ cell migration, gonad formation, wing venation and eye development (Gramates et al., 2017).
Abi	Regulates WASP-WAVE (Lin et al., 2009).
alc	AMPK homolog, Neuronal maintenance factor (Spasi et al., 2008).
alpha-Spec	$\alpha$ Spectrin ( $\alpha$ -Spec) is an essential protein that interacts with $\beta$ -Spec or Kst to form a heterotetramer. It is typically associated with the plasma membrane. It functions in a lipoprotein pathway that delivers dietary fat to the larval fat body for storage. It is also believed to function in asymmetric division of germ line stem cells via cytoplasmic structures called spectrosomes and fusomes (Gramates et al., 2017).
alpha-Tub84B	Subunit of microtubules (Gramates et al., 2017).

Name	Definition
alpha Tub84D	Subunit of microtubules (Gramates et al., 2017).
Apc	APC-like is one of two Drosophila APC family proteins. It is a key negative regulator of Wingless signaling, as a critical component of the destruction complex that phosphorylates beta-catenin and thus targets it for ubiquitination and proteasomal destruction. It is the primary family member during CNS and eye development, and functions redundantly with Apc2 in adult development (Gramates et al., 2017).
arm	Armadillo is the Drosophila homolog of beta-catenin. It plays separable roles in cell adhesion and Wingless signaling. It links classic cadherin cell adhesion receptors to alpha-catenin and the Actin cytoskeleton, and it acts as the key regulated effector of Wingless signaling, working with TCF/LEF proteins as a transcriptional co-activator (Gramates et al., 2017).
ASPP	Positive regulator of dCsk kinase (Langton et al., 2007).
beta Tub56D	Subunit of microtubules (Gramates et al., 2017).
beta Tub60D	Subunit of microtubules (Gramates et al., 2017).
Bgb	Big brother (Bgb) is a beta-subunit of the transcription factor complex Core Binding Factor (CBF), which is involved in transcription regulation. Bgb regulates hemocyte proliferation and acts redundantly with Bro in embryonic segmentation (Gramates et al., 2017).
Bin1	Bicoid interacting protein 1 (Bin1) interacts with sequence-specific transcription factors (e.g. bcd, Kr, Trl) to help recruit the Sin3A-HDAC1 histone deacetylase co-repressor complex to target genes. Bin1 contributes to axis patterning and embryonic segmentation via transcription regulation (Gramates et al., 2017).
Bro	Brother (Bro) is a beta-subunit of the transcription factor complex Core Binding Factor (CBF), which is involved in transcription regulation. Bro regulates hemocyte proliferation and acts redundantly with Bgb in embryonic segmentation (Gramates et al., 2017).
bsk	Basket (Bsk) is a serine/threonine-protein kinase, a key component of JNK pathway that phosphorylates the Jra transcription factor. Bsk roles include regulation of cell shape and stress response (Gramates et al., 2017).

Name	Definition
Cam	Calmodulin is a Calcium-binding messenger protein. It acts as a part of the calcium signal transduction pathway and interacts with various target proteins such as kinases or phosphatases. It is involved in Actin filament-based movement, mitotic spindle assembly, axon guidance, DNA damage response, autophagy, cell death, sensory perception, and muscle homeostasis (Gramates et al., 2017).
cbs	trans-Golgi protein that links Golgi inheritance to the cell cycle via centrosome cycle (Eisman et al., 2006).
Cdc5	Splicosome complex member (Guilgur et al., 2014).
ci	Cubitus interruptus (Ci) is a Zn-finger family of transcription factor. It contributes to Hedgehog (Hh) signaling pathway that is involved in pattern formation and growth control. Ci undergoes limited proteolysis to be converted into a truncated form that function as a transcription repressor in the absence of hh but is converted into a full-length transcription activator in the presence of hh (Gramates et al., 2017).
cold	Helps form cellular junctions (Nilton et al., 2010)
dally	Division abnormally delayed (Dally) is a core protein of heparan sulfate proteoglycans of the glypcan family. It acts as a co-receptor for growth factors and morphogens, such as FGFs, dpp, wg, hh and upd1, and affects signaling and distribution of these ligands. Dally's roles include wing development and germline stem cell maintenance (Gramates et al., 2017).
Dap160	Positvely regulates aPKC to polarize cells (Chabu and Doe, 2008)
DCTN1-p150	Dynactin 1, p150 subunit (DCTN1-p150) is the p150 subunit of the dynactin complex. Dynactin is thought to act as a dynein receptor that modulates binding of dynein to cellular cargoes and enhances the processivity of dynein movement. DCTN1-p150 roles include oocyte polarity, mitotic cell division, embryonic development, neuronal transport and neurogenesis (Gramates et al., 2017).
Der-2	Endoplasmic Reticulum Associatied Degredation Regulator (Oda et al., 2006).
Dysb	Presynaptic protein that regulates vesicle release (Dickman and Davis, 2009).
Eb1	Microtubule Plus End Tracking Protein (Gramates et al., 2017).

Name	Definition
EIF-2beta	IF-2 functions in the early steps of protein synthesis by forming a ternary complex with GTP and initiator tRNA. This preinitiation complex mediates ribosomal recognition of a start codon during the scanning process of the leader region (Bateman et al., 2017).
ERR	Estrogen-related receptor directly induces a transcriptional switch in mid-embryogenesis, up-regulating the genes that act in biosynthetic pathways associated with aerobic glycolysis. This nuclear receptor establishes the metabolic state that supports growth during larval stages (Gramates et al., 2017).
Fas2	Neuronal recognition molecule for the MP1 axon pathway, pathway recognition for axons during the development of nerve fascicles (Bateman et al., 2017).
flw	Required for cell adhesion in non-muscle tissues and in maintenance of muscle attachment. Vital for larval development (Bateman et al., 2017).
Gint3	Contains ubiquitin like domain (Gramates et al., 2017).
Golgin245	Golgins are long, parallel coiled-coil homodimers important in regulating vesicle mediated transport and Golgi architecture. They are anchored to specific sites on the Golgi complex surface either by a C-terminal transmembrane domain or interacting with Arf/Arf-like GTPases via a GRIP/GRAB domain. Golgins are thought to act as tethering factors, interacting with vesicles by binding Rab GTPases and guiding them to the correct Golgi cisternae (Gramates et al., 2017).
hpo	Hippo (Hpo) is a kinase in the Salvador-Warts-Hippo pathway. It controls tissue growth by controlling cell growth, proliferation and apoptosis. Hpo has several roles in post-mitotic cells including fate specification of photoreceptors and tiling of dendritic neurons (Gramates et al., 2017).
Hrs	Essential role in endosome membrane invagination and formation of multivesicular bodies, MVBs. Required during gastrulation and appears to regulate early embryonic signaling pathways. Inhibits tyrosine kinase receptor signaling by promoting degradation of the tyrosine-phosphorylated, active receptor, potentially by sorting activated receptors into MVBs. The MVBs are then trafficked to the lysosome where their contents are degraded (Bateman et al., 2017).
Jra	Transcription factor that recognizes and binds to enhancer heptamer motif. Plays a role in dorsal closure (Bateman et al., 2017).

Name	Definition
Kat60	Katanin 60 (Kat60) encodes a microtubule depolymerase involved in multiple microtubule-related processes including chromatid segregation, cell migration and dendrite development (Gramates et al., 2017).
Kat80	Augments microtubule severing by Kat60 (Yu et al., 2005)
kdn	Plays a role in controlling neuronal activity and seizure susceptibility. (Bateman et al., 2017)
kra	Krasavietz (Kra) is a translational regulator that is able to interact with the translation initiation factor eIF-2 $\beta$ . Kra is involved in axon guidance, long-term memory, and alcohol addiction (Gramates et al., 2017).
Kst	Karst (Kst), together with $\alpha$ -Spec, crosslinks F-actin and acts as a molecular scaffold. Kst is recruited to the apical membrane of epithelial cells by the apical determinant crb. Kst roles include endosomal trafficking, cell adhesion, cell signaling and growth control (Gramates et al., 2017).
ktub	King tubby (Ktub) belongs to the tubby and tubby-like protein family. Ktub functions in endocytosis of the rhodopsin ninaE, retinal degeneration upon light stimulation, and the localization of trp channel in cilia (Gramates et al., 2017).
Lap1	May have a role in assembling adherens junctions. (Bateman et al., 2017)
Lsd-2	Lipid storage droplet-2 (Lsd-2) is a protein associated with lipid droplets. It acts as a barrier for lipases (such as bmm) and thus prevents the mobilization of lipid stores. Lsd-2 is involved in regulation of lipid storage amount and energy homeostasis and acts in concert with Lsd-1 (Gramates et al., 2017).
Mad	Mothers against dpp (Mad) is the primary transcription factor that mediates cellular response to the BMP like ligands dpp, scw and gbb. Upon phosphorylation by either sax or tkv (type I BMP receptors), it forms a complex with Med and translocates to the nucleus where, together with cofactors, it regulates expression of BMP response target genes (Gramates et al., 2017).
mats	Coactivator of Warts (Wts) kinase in the Hippo/SWH (Sav/Wts/Hpo)signaling pathway, a signaling pathway that plays a pivotal role in organ size control and tumor suppression by restricting proliferation and promoting apoptosis. (Bateman et al., 2017)

Name	Definition
Mi-2	Mi-2 is a nuclear ATP-dependent nucleosome remodeler of the CHD family. It associates with active chromatin and utilizes the energy of ATP hydrolysis to move nucleosomes along DNA. It is required for repression of cell type-specific genes, full activation of heat shock genes and regulates higher order chromatin structure of polytene chromosomes (Gramates et al., 2017).
Mms19	Needed for mitotic divisions. Thought to activate CAK (Nag et al., 2018).
mrt	No data.
Msn	Missshapen is a Sterile 20 MAP kinase kinase kinase (Gramates et al., 2017).
Mtpalpha	Subunit of mitochondrial trifunctional protein which catalyzes fatty acid oxidation (Kishita et al., 2012).
N	Signaling protein, which regulates, with both positive and negative signals, the differentiation of at least central and peripheral nervous system and eye, wing disk, oogenesis, segmental appendages such as antennae and legs, and muscles, through lateral inhibition or induction. Functions as a receptor for membrane-bound ligands Delta and Serrate to regulate cell-fate determination (Bateman et al., 2017).
NetB	Netrins control guidance of CNS commissural axons and peripheral motor axons. Its association with either fra or unc-5 receptors will lead to axon attraction or repulsion, respectively. While short-range repulsion requires both fra and unc-5 receptors, long-range repulsion only requires unc-5 (Gramates et al., 2017).
Nf-YC	Transcription factor (Gramates et al., 2017).
Nrg	The long isoform may play a role in neural and glial cell adhesion in the developing embryo. The short isoform may be a more general cell adhesion molecule involved in other tissues and imaginal disk morphogenesis. Vital for embryonic development. Essential for septate junctions (Bateman et al., 2017).
Nrx-1	Neurexin 1 (Nrx-1) is a transmembrane synaptic adhesive molecule. It regulates the synaptic architecture and function in the brain and neuromuscular junction. Neurexin-1 roles include synaptic growth, transmission, synaptic formation and also regulation of learning and memory, locomotion and visual function (Gramates et al., 2017).

Name	Definition
Nrx-IV	Neurexin-IV is a transmembrane protein that is critical for septate junction formation in epithelia and between ensheathing glial cells and neurons (Gramates et al., 2017).
p115	P115 is a key protein of the early secretory pathway where it interacts with GM130 and Rab proteins. Depletion from S2 Cells leads to a fragmentation of the Golgi but without impairing transport efficiency (Gramates et al., 2017).
P32	P32 is an evolutionarily conserved mitochondrial protein. P32 functions in presynaptic calcium signaling and neurotransmitter release as well as chromatin metabolism (Gramates et al., 2017).
Pak	p21-activated kinase (Pak) is a serine/threonine effector kinase for the small GTPases Rac and Cdc42 involved in cytoskeletal regulation. Pak roles include growth cone guidance, synaptic development, and epithelial morphogenesis in both the ovary and embryo (Gramates et al., 2017).
PAPLA1	Phosphatidic Acid Phospholipase A1 is required for the endoplasmic reticulum to Golgi trafficking of a family of G-protein coupled receptors (Gramates et al., 2017).
Patronin	Involved in mitotic spindle assembly . Regulates microtubule (MT) severing . Antagonizes the activity of the kinesin-13 depolymerase Klp10A thereby switching off the depolymerization of the MTs at their pole-associated minus ends, which turns off poleward flux and induces anaphase B spindle elongation (Bateman et al., 2017).
PKD	Protein Kinase D is a Ser/Thr kinase of the PKC family of Ca <sup>2+</sup> /calmodulin-dependent protein kinases. It regulates actin-dynamics by controlling ssh phosphatase activity (Gramates et al., 2017).
pUf68	pUf68/Half pint (Hfp) is a single stranded nucleic acid binding protein with roles in transcription and RNA splicing. Hfp functions in Myc transcriptional repression and cell growth control, via interaction with the transcription factor hay (Gramates et al., 2017).
Rab21	The Rab family are members of the Ras superfamily of small GTPases. Rabs regulate vesicle trafficking including cargo selection, vesicle budding, transport, docking and targeting. They localize to different intracellular compartments directed by specific isoprenylation of C-terminal motifs (Gramates et al., 2017).

Name	Definition
Rcd5	Reduction in Cnn dots 5 (Rcd5) belongs to the family of forkhead-associated domain proteins. As part of the Non-Specific Lethal complex, it regulates transcription of housekeeping genes by facilitating recruitment of RNA polymerase II to promoters. During mitosis it plays a key role in chromosomal-dependent microtubule assembly, centriole duplication and pericentriolar material maturation (Gramates et al., 2017).
Rel	Relish is a transcription factor and the downstream component of the Immune Deficiency pathway, which regulates the antibacterial response and other less characterized cellular processes (Gramates et al., 2017).
RnpS1	RNA binding protein (?)
RpL10Aa	Large ribosomal protein (Gramates et al., 2017).
RpL5	Ribosomal protein L5 is a component of the large subunit of cytoplasmic ribosomes, which translate mRNAs encoded by the nuclear genome (Gramates et al., 2017).
Rpn5	Subunit of the proteasome (Gramates et al., 2017).
RpS5b	Small ribosomal protein (Gramates et al., 2017).
Rtnl1	ER organizing protein, ER tubulator (O'Sullivan et al., 2012).
Sals	Sarcomere length short is an Actin binding protein that promotes sarcomeric Actin filament elongation from pointed ends during muscle growth (Gramates et al., 2017).
scny	Required for maintaining multiple types of adult stem cells, including male and female germline, epithelial follicle cell and intestinal stem cells. May function as a transcriptional repressor by continually deubiquiting histone H2B at the promoters of genes critical for cellular differentiation, thereby preventing histone H3 'Lys-4' trimethylation (H3K4) (Bateman et al., 2017).
Sec3	Component of the exocyst complex involved in the docking of exocytic vesicles with fusion sites on the plasma membrane (Bateman et al., 2017).
Sema-1a	Sema-1a is a transmembrane protein belonging to the semaphorin protein family. Sema-1a is a repulsive axon guidance cue, signaling through direct interactions with the PlexA receptor, and also functions as a receptor to regulate dendrite targeting and axon guidance. Secreted semaphorins Sema-2a and Sema-2b can act as Sema-1a ligands (Gramates et al., 2017).

Name	Definition
Sema-2a	Sema-2a is a secreted member of the semaphorin protein family. Sema-2a utilizes the PlexB receptor to mediate both repulsive and attractive axon guidance. Sema-2a also utilizes the transmembrane semaphorin Sema-1a as a receptor for dendrite process guidance (Gramates et al., 2017).
Set1	SET domain containing 1 methylates histone H3 on lysine 4 at promoter-proximal regions of most RNA Polymerase II transcribed genes. It is the catalytic subunit of the COMPASS complex (Gramates et al., 2017).
Shaggy	Shaggy is a Glycogen Synthase Kinase 3, and a key component of the $\beta$ -catenin destruction complex. It functions in the canonical Wnt cascade (Gramates et al., 2017).
Shot	See Background
Sirt7	Sirtuin family members are NAD+-dependent lysine deacetylases that share a common catalytic core domain (Gramates et al., 2017).
skd	Skuld (Skd) is a subunit of the kinase module of the mediator complex. Skd is not required for all the transcriptional functions of mediator, but links it to a set of transcription factors that are involved in developmental signaling (Gramates et al., 2017).
Snatin	Component of the biogenesis of lysosome-related organelles complex-1 (BLOC-1) involved in pigment granule biogenesis. May participate in the coupling of lysosomes to microtubule plus-end-directed kinesin motor (By similarity) (Gramates et al., 2017).
Snoo	Sno oncogene (Snoo) belongs to the highly conserved Sno/Ski family of Smad binding proteins. Snoo specifically binds Med and switches its affinity from Mad to Smox, resulting in antagonism of dpp and facilitation of Activin signaling. Snoo overexpression antagonizes both pathways (Gramates et al., 2017).
SoxN	DNA Binding Protein, controls cell fate decisions. Part of the Wnt pathway (Chao et al., 2007).
Srp72	RNA binding protein, part of the signal recognition particle complex (Gramates et al., 2017).
stan	Starry night is the seven-pass transmembrane cadherin and controls planar cell polarity, neuronal dendrite morphogenesis, and axon guidance (Gramates et al., 2017).
Tango6	May be involved in protein secretion and Golgi organization (Bateman et al., 2017).

Name	Definition
Ten-m	Involved in neural development, regulating the establishment of proper connectivity within the nervous system. Acts as a homophilic and heterophilic synaptic cell adhesion molecule that drives synapse assembly. Promotes bi-directional trans-synaptic signaling with Ten-a to organize neuromuscular synapses. Functions in olfactory synaptic partner matching by promoting homophilic cell adhesion between pre-synaptic olfactory receptor neurons (ORN) axons and post-synaptic projection neurons (PN) dendrites partner in the developing antennal lobe to form stable connections. Also required for peripheral axon growth cone guidance and target recognition of motor neurons (Bateman et al., 2017).
Thiolase	Involved in mitochondrial $\beta$ oxidation (Kishita et al., 2012).
Thor	Thor is a eukaryotic translation initiation factor 4E binding protein that is controlled by tor. It contributes to translation regulation, response to environmental stress and cell growth regulation (Gramates et al., 2017).
Tip60	Tat interactive protein 60kDa (Tip60) is a lysine acetyltransferase that acetylates histone proteins to regulate chromatin packaging and epigenetic gene control. Tip60 also acetylates non-histone proteins, and plays a role in apoptosis, DNA repair and various neural processes (Gramates et al., 2017).
TwdlB	Chitin regulating protein, controls morphogenesis (Guan et al., 2006).
TwdlD	Chitin regulating protein, controls morphogenesis (Guan et al., 2006).
TwdlF	Chitin regulating protein, controls morphogenesis (Guan et al., 2006).
TwdlJ	Chitin regulating protein, controls morphogenesis (Guan et al., 2006).
TwdlN	Chitin regulating protein, controls morphogenesis (Guan et al., 2006)
TwdlO	Chitin regulating protein, controls morphogenesis (Guan et al., 2006).
U2af50	Necessary for the splicing of pre-mRNA. Binds to the polypyrimidine tract of introns early during spliceosome assembly (By similarity) (Bateman et al., 2017).

---

Name	Definition
Ubi-p63E	Ubiquitin exists either covalently attached to another protein, or free (unanchored). When covalently bound, it is conjugated to target proteins via an isopeptide bond either as a monomer (monoubiquitin), a polymer linked via different Lys residues of the ubiquitin (polyubiquitin chains) or a linear polymer linked via the initiator Met of the ubiquitin (linear polyubiquitin chains). Polyubiquitin chains, when attached to a target protein, have different functions depending on the Lys residue of the ubiquitin that is linked: Lys-48-linked is involved in protein degradation via the proteasome. Linear polymer chains formed via attachment by the initiator Met lead to cell signaling (Bateman et al., 2017).
UGP	Involved in carbohydrate metabolism (Gramates et al., 2017).
Usp10	The Ubiquitin Specific Proteases (USP) is the largest sub-family of deubiquitinases. They are cysteine proteases and catalyze the removal of ubiquitin from ubiquitin chains and ubiquitinated proteins (Gramates et al., 2017).
Vha55	Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase. V-ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells (Bateman et al., 2017)
wde	Methyl transferase cofactor (Koch et al., 2009).
Zormin	Zormin is found in the Z-disc and the M-line of muscles. It probably has a structural function (Gramates et al., 2017).

---



# **Appendix C**

## **Gene Ontology Definitions**

Table C.1: Gene Ontology definitions taken directly from PANTHER for the protein class terms in Figure 3.4 (Mi et al., 2013).

**Calcium-binding Protein:** A protein that participates in calcium cell signaling pathways by binding to Ca<sup>2+</sup>.

**Cell Adhesion Molecule:** A protein that mediates cell-to-cell adhesion.

**Cytoskeletal Protein:** Major constituent of the cytoskeleton found in the cytoplasm of eukaryotic cells. They form a flexible framework for the cell, provide attachment points for organelles and formed bodies, and make communication between parts of the cell possible.

**Enzyme Modulator:** A protein that modulates the activity of a select group of an enzyme such as kinases, phosphatases, protease, and G-proteins.

**Extracellular Matrix Protein:** A protein that is produced and secreted by cells and forms an intricate extracellular meshwork in which cells are embedded to construct tissues.

**Hydrolase:** Enzymes catalyzing hydrolysis of a variety of bonds, such as esters, glycosides, or peptides.

**Isomerase:** A class of enzymes that catalyze geometric or structural changes within a molecule to form a single product. The reactions do not involve a net change in the concentrations of compounds other than the substrate and the product.

**Lyase:** Enzymes that either strip a group from a molecule to form a double bond or add a group to a double bond.

**Membrane Traffic Protein:** A protein that is located on an intracellular vesicle membrane, and is responsible for the docking or fusion of the vesicle to the cytoplasma membrane.

**Nucleic Acid Binding:** A molecule that binds a nucleic acid. It can be an enzyme or a binding protein.

**Oxidoreductase:** An enzyme that catalyzes a redox reaction.

**Receptor:** A molecular structure within a cell or on the cell surface characterized by selective binding of a specific substance and a specific physiologic effect that accompanies the binding.

**Signalling Molecule:** A molecule that transduces a signal between cells.

**Transcription Factor:** A protein required for the regulation of RNA polymerase by specific regulatory sequences in or near a gene.

**Transfer/carrier Protein:** Proteins that carry specific substances in the blood or in the cell. They usually are not involved in transmembrane transport.

**Transferase:** Enzymes transferring a group from one compound (donor) to another compound (acceptor). Kinase is a separate category, so it is not included here.

**Transmembrane Receptor Regulatory/Adaptor Protein:** An intracellular protein that interacts with a transmembrane receptor, serving as either a regulator or adaptor to the receptor.

**Transporter:** A class of transmembrane proteins that allows substances to cross plasma membranes far faster than would be possible by diffusion alone. Please note that ion channel has its own category and is not included here.

Table C.2: Gene Ontology definitions taken directly from AmiGO for the molecular function terms in Figure 3.5 and Table 3.9 (Carbon et al., 2009).

**Binding:** The selective, non-covalent, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule.

**Catalysis:** Catalysis of a biochemical reaction at physiological temperatures.

**Kinase Activity** Catalysis of the transfer of a phosphate group, usually from ATP, to a substrate molecule.

**Nucleic Acid Binding** Interacting selectively and non-covalently with any nucleic acid.

**Receptor Activity:** Receiving a signal and transmitting it in the cell to initiate a change in cell activity.

**Signal Transducer Activity:** Conveys a signal across a cell to trigger a change in cell function or state.

**Structural Constituent Of Cytoskeleton** The action of a molecule that contributes to the structural integrity of a cytoskeletal structure.

**Structural Molecule Activity:** The action of a molecule that contributes to the structural integrity of a complex or its assembly within or outside a cell.

**Transcription Cofactor Activity** Interacting selectively and non-covalently with a regulatory transcription factor and also with the basal transcription machinery in order to modulate transcription.

**Translation Regulator Activity:** Any molecular function involved in the initiation, activation, perpetuation, repression or termination of polypeptide synthesis at the ribosome.

**Transporter Activity:** Enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of or within a cell, or between cells.

Table C.3: Gene Ontology definitions taken directly from AmiGO for the biological process terms in Table 3.8 (Carbon et al., 2009).

- Apoptotic Process** A programmed cell death process which begins when a cell receives an internal (e.g. DNA damage) or external signal (e.g. an extracellular death ligand), and proceeds through a series of biochemical events (signaling pathway phase) which trigger an execution phase.
- Carbohydrate Metabolic Process** The chemical reactions and pathways involving carbohydrates.
- Cell Cycle** The progression of biochemical and morphological phases and events that occur in a cell during successive cell replication or nuclear replication events.
- Cell Differentiation** The process in which relatively unspecialized cells, e.g. embryonic or regenerative cells, acquire specialized structural and/or functional features that characterize the cells, tissues, or organs of the mature organism or some other relatively stable phase of the organism's life history.
- Cellular Component Morphogenesis** The process in which cellular structures, including whole cells or cell parts, are generated and organized.
- Cellular Component Movement** The directed, self-propelled movement of a cell or subcellular component without the involvement of an external agent such as a transporter or a pore.
- Cytokinesis** The division of the cytoplasm and the plasma membrane of a cell and its partitioning into two daughter cells.
- Cytoskeletal organization** A process that is carried out at the cellular level which results in the assembly, arrangement of constituent parts, or disassembly of cytoskeletal structures.
- Locomotion** Self-propelled movement of a cell or organism from one location to another.
- MAPK Cascade** An intracellular protein kinase cascade containing at least a MAPK, a MAPKK and a MAP3K.
- Nervous System Development** The process whose specific outcome is the progression of nervous tissue over time, from its formation to its mature state.
- Transcription by RNA polymerase II promoter** The synthesis of RNA from a DNA template by RNA polymerase II, originating at an RNA polymerase II promoter. Includes transcription of messenger RNA (mRNA) and certain small nuclear RNAs (snRNAs).

# Appendix D

## PANTHER Results

Table D.1: Abridged PANTHER Results for 118 classified neighbors of the node A1Z9J1, Shot Isoform A in the Ritz Interactome.

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
P92177	14-3-3 protein epsilon; 14-3-3epsilon		protein binding; protein domain specific binding; protein heterodimerization activity; phosphoserine residue binding
Q9VFE4	40S ribosomal protein S5b; RpS5b	ribosomal protein	mRNA binding; structural constituent of ribosome; rRNA binding
Q9XZ56	4E-binding protein THOR; Thor	translation factor	eukaryotic initiation factor 4E binding
Q9W5R8	60S ribosomal protein L5; RpL5	ribosomal protein	RNA binding; structural constituent of ribosome; protein binding; 5S rRNA binding
A0A0B4K774	Abelson interacting protein, isoform D; Abi	G-protein modulator	protein binding; cadherin binding
A1Z7Q8	Alicorn, isoform A; alc	kinase modulator	

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q86BG1	Ankyrin-repeat, SH3-domain, and Proline-rich-region containing protein, isoform B; ASPP		p53 binding; protein tyrosine kinase activator activity
Q9VAS9	APC-like, isoform A; Apc		protein binding; beta-catenin binding; gamma-catenin binding
P18824	Armadillo segment polarity protein; arm		RNA polymerase II transcription coactivator activity; transcription coactivator activity; signal transducer activity; protein binding; transcription factor binding; kinase binding; protein phosphatase binding; nuclear hormone receptor binding; alpha-catenin binding; cadherin binding
Q7JUS1	AT09608p; CG30491; CG30495	dehydrogenase; reductase	oxidoreductase activity
Q9VR43	AT09986p; CG15625		
Q9VFE0	AT11516p; RpL10Aa	ribosomal protein	RNA binding; structural constituent of ribosome
Q9V397	BcDNA. GH12558; Mtpalpha	dehydrogenase; epimerase/racemase; hydratase	enoyl-CoA hydratase activity; long-chain-3-hydroxyacyl-CoA dehydrogenase activity
Q7KMH9	BcDNA. LD26050; wde		

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q24040	Big brother; Bgb		DNA binding; transcription cofactor activity; transcription coactivator activity; transcription factor binding; protein heterodimerization activity
Q24039	Brother; Bro		DNA binding; transcription cofactor activity; transcription coactivator activity; transcription factor binding; protein heterodimerization activity
P62152	Calmodulin; Cam	calmodulin	calcium ion binding; protein binding; myosin V binding; myosin heavy chain binding; myosin VI head/neck binding
Q9W0R0	Cell division cycle 5, isoform A; Cdc5		RNA polymerase II transcription factor activity, sequence-specific DNA binding; transcription factor activity, RNA polymerase II transcription factor recruiting; RNA binding; sequence-specific DNA binding; transcription regulatory region DNA binding
Q7K4K7	Centrosomin's beautiful sister, isoform A; cbs		Rab GTPase binding; ADP-ribosylation factor binding; protein homodimerization activity
Q9VFL0	CG3259; CG3259-RA		molecular_function; microtubule binding
Q9VJQ0	CG4168, isoform B; CG4168		
Q9VPD6	CG5059, isoform A; DmelCG5059		

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
M9PF61	CG6084, isoform C; 34F4T	reductase	alditol:NADP+ 1-oxidoreductase activity; alcohol dehydrogenase activity; oxidoreductase activity
Q7JXC4	CG6459 protein; P32		mRNA binding; translation activator activity; histone binding; mitochondrial ribosome binding
Q9VHB8	CG8199; CG8199		alpha-ketoacid dehydrogenase activity; 3-methyl-2-oxobutanoate dehydrogenase activity
Q9VP62	CG9389; 142336_at	phosphatase	inositol monophosphate 1-phosphatase activity
O97159	Chromodomain- helicase-DNA- binding protein Mi-2 homolog; Mi-2		DNA binding; chromatin binding; helicase activity; protein binding; ATP binding; zinc ion binding; ATPase activity; nucleosome binding; nucleosome-dependent ATPase activity
Q9VPR7	Coiled, isoform A; cold		
Q9VEU2	Derlin; Der-2	receptor	molecular_function
Q24114	Division abnormally delayed protein; dally		protein binding
P13496	Dynactin subunit 1; Gl	non-motor microtubule binding protein	microtubule binding; dynein intermediate chain binding; dynein complex binding

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
M9ND00	Dynamin associated protein 160, isoform D; Dap160	G-protein modulator; calcium-binding protein; membrane traffic protein	nucleic acid binding; calcium ion binding; protein binding
A1Z6P3	Eb1, isoform F; Eb1	non-motor microtubule binding protein	protein binding; microtubule binding; myosin binding; microtubule plus-end binding
Q8WS79	Estrogen-related receptor splice variant; ERR	C4 zinc finger nuclear receptor; nucleic acid binding; receptor	DNA binding transcription factor activity; steroid hormone receptor activity; nuclear receptor activity; steroid binding; zinc ion binding; sequence-specific DNA binding
P41375	Eukaryotic translation initiation factor 2 subunit 2; eIF-2beta	translation initiation factor	translation initiation factor activity
Q9VVG4	Exocyst complex component 1; Sec3	transfer/carrier protein	phosphatidylinositol-4,5-bisphosphate binding; GTP-Rho binding
P34082	Fasciclin-2; Fas2	transmembrane receptor regulatory/adaptor protein	protein binding
Q9VBD5	FI07246p; TwdlN		structural constituent of chitin-based cuticle
A1ZBK1	FI23922p1; CG10051-RA	metalloprotease	

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q7K0S5	GDI interacting protein 3, isoform B; Gint3		
Q9W3N6	General vesicular transport factor p115; p115	membrane traffic protein	protein transporter activity
Q9VZJ3	GH01794p; Rcd5		G-quadruplex RNA binding; protein binding; histone acetyltransferase activity ; histone acetyltransferase activity ; histone acetyltransferase activity
Q95U54	GH06271p; Mms19	nucleic acid binding; transcription cofactor	
Q9V3Z4	GH11341p; Rpn5	enzyme modulator	molecular function
Q9W1H8	GH13256p; Thiolase	acetyltransferase	3-hydroxyacyl-CoA dehydrogenase activity; acetyl-CoA C-acyltransferase activity; enoyl-CoA hydratase activity; long-chain-3-hydroxyacyl-CoA dehydrogenase activity
Q9W1R3	Golgin-245 ortholog; Golgin245	membrane trafficking regulatory protein; transfer/carrier protein	Rab GTPase binding; ADP-ribosylation factor binding; protein homodimerization activity

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q960X8	Hepatocyte growth factor-regulated tyrosine kinase substrate; Hrs		ubiquitin binding; metal ion binding
Q960X4	Histone acetyltransferase Tip60; Tip60	acetyltransferase; histone acetyltransferase chromatin binding protein; zinc finger transcription factor	activity
Q9VEX9	Histone deacetylase complex subunit SAP18; Bin1	chromatin chromatin-binding protein	transcription corepressor activity; protein binding; transcription factor binding
Q5LJZ2	Histone-lysine N-methyltransferase SETD1; Set1		nucleic acid binding; histone methyltransferase activity
Q9VEF5	Inhibitor of growth protein; CG7379	chromatin chromatin-binding protein; transcription cofactor; zinc finger transcription factor	
Q9VP84	IP06402p; DmelCG10581	kinase	nucleotide phosphatase activity, acting on free nucleotides
Q9VQ11	IP19042p; CG31928	aspartic protease	aspartic-type endopeptidase activity

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
A8JNJ6	Karst, isoform E; Kst		Actin binding; protein binding; phospholipid binding; microtubule binding; cytoskeletal protein binding
A0A0B4KFL3	Katanin p60 ATPase-containing subunit A1; Kat60	non-motor microtubule binding protein	ATP binding; microtubule binding; microtubule-severing ATPase activity; ATPase activity; myosin binding
Q0KHS0	Katanin p80 WD40 repeat-containing subunit B1; Kat80		microtubule binding; microtubule-severing ATPase activity
P18431	Kinase shaggy; Shaggy	non-receptor serine/threonine protein kinase	protein serine/threonine kinase activity; protein binding; ATP binding
Q86PC9	King tubby; ktub		phosphatidylinositol binding
Q9VNE2	Krasavietz; kra	nuclease; translation initiation factor	protein binding; translation initiation factor binding; ribosome binding; ribosomal small subunit binding
Q9V780	Lap1; Lap1		
Q9VSW1	LD13601p; UGP		UTP:glucose-1-phosphate uridylyltransferase activity
Q9VMZ5	LD19715p; TwdlF		structural constituent of chitin-based larval cuticle
Q9VLS7	LD21067p; PAPLA1	membrane traffic protein; phospholipase	phospholipase activity; metal ion binding
Q9VHC0	LD23870p; RnpS1	mRNA splicing factor	nucleic acid binding; mRNA binding

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q9W0L7	LD28815p; Usp10		thiol-dependent ubiquitin-specific protease activity
Q7KTX8	Mediator of RNA polymerase II transcription subunit 13; skd		RNA polymerase II transcription cofactor activity; protein binding
Q9W002	Missphen, isoform A; Msn		protein serine/threonine kinase activity; signal transducer, downstream of receptor, with serine/threonine kinase activity; ATP binding; MAP kinase kinase kinase kinase activity; JUN kinase kinase kinase kinase activity
Q95RA8	MOB kinase activator-like 1; mats	kinase activator	protein binding; protein kinase binding
P42003	Mothers against dpp; Mad	transcription factor	RNA polymerase II distal enhancer sequence-specific DNA binding; RNA polymerase II activating transcription factor binding; RNA polymerase II transcription coactivator activity; DNA binding; transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding; protein binding; transforming growth factor beta receptor, pathway-specific cytoplasmic mediator activity; sequence-specific DNA binding

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q9VAQ1	NAD-dependent protein deacetylase Sirt7; Sirt7		NAD+ binding
Q24568	Netrin-B; NetB	extracellular matrix linker protein; protease inhibitor; receptor	
A0A0B4KGL0	Neurexin 1, isoform F; Nrx-1		protein binding; protein kinase binding; apolipoprotein binding; neurexin family protein binding; neuroligin family protein binding
Q94887	Neurexin-4; Nrx-IV		
P07207	Neurogenic locus Notch protein; N		chromatin binding; transmembrane signaling receptor activity; calcium ion binding; protein binding
P20241	Neurogian; Nrg		calcium ion binding
Q94527	Nuclear factor NF-kappa-B p110 subunit; Rel		RNA polymerase II transcription factor activity, sequence-specific DNA binding; transcriptional activator activity, RNA polymerase II proximal promoter sequence-specific DNA binding; DNA binding; DNA binding transcription factor activity; protein binding; sequence-specific DNA binding

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q9W3V9	Nuclear factor Y-box C; Nf-YC	DNA-directed DNA polymerase; chromatin chromatin-binding protein; transcription factor	DNA binding transcription factor activity; sequence-specific DNA binding; protein heterodimerization activity
B7Z0W0	p21-activated kinase, isoform F; Pak		protein serine/threonine kinase activity; signal transducer, downstream of receptor, with serine/threonine kinase activity; ATP binding; Rho GTPase binding; SH3 domain binding; Rac GTPase binding
A1ZAU8	Patronin; Patronin		calmodulin binding; microtubule binding; spectrin binding; microtubule minus-end binding
Q8T6B9	Poly-binding-splicing factor half pint; pUf68		mRNA binding; protein binding; poly RNA binding
P0CG69	Polyubiquitin; Ubi-p63E	ribosomal protein	protein tag; ubiquitin protein ligase binding; proteasome binding
Q9W401	Probable citrate synthase, mitochondrial; kdn	lyase; transferase	citrate -synthase activity
Q9V677	Proteasome-associated protein ECM29 homolog; CG8858	kinase modulator	molecular_function; protein-containing complex scaffold activity

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q9V5N8	Protocadherin-like wing polarity protein stan; stan	G-protein coupled receptor; cadherin; protease	G-protein coupled receptor activity; signal transducer activity, downstream of receptor; calcium ion binding; cell adhesion molecule binding
Q7PLE9	Rab21, isoform C; Rab21		GTPase activity; protein binding; GTP binding; phosphatidylinositol 3-kinase catalytic subunit binding
Q9VBD2	RE49895p; Twd1D		structural constituent of chitin-based larval cuticle
E1JHT6	Reticulon-like protein; Rtnl1		
Q7K2B0	Ribosomal RNA-processing protein 8; CG7137		methylated histone binding
Q9VBD8	RT02919p; TwdlO		structural constituent of chitin-based cuticle
Q9VBD6	RT03118p; TwdlJ		structural constituent of chitin-based cuticle
Q9VMJ2	RT07854p; CG13996-RA		molecular function
Q24322	Semaphorin-1A; Sema-1a	membrane-bound signaling molecule	protein binding; heparin binding; semaphorin receptor binding; chemorepellent activity
Q24323	Semaphorin-2A; Sema2a	membrane-bound signaling molecule	protein binding

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q8T0S6	Serine/threonine-protein kinase hippo; hpo		protein serine/threonine kinase activity; signal transducer, downstream of receptor, with serine/threonine kinase activity; MAP kinase kinase kinase activity; protein binding; ATP binding; kinase activity; protein homodimerization activity
A0A0B4KHC3	Serine/threonine-protein kinase; PKD		protein serine/threonine kinase activity; protein kinase C activity; ATP binding
P48462	Serine/threonine-calcium-protein binding phosphatase beta isoform; flw	protein; protein phosphatase	protein serine/threonine phosphatase activity; myosin phosphatase activity
A1Z9J3	Shot, isoform H; Shot		Actin binding; calcium ion binding; protein binding; microtubule binding; cytoskeletal protein binding
Q9VDK7	Signal recognition particle subunit SRP72; Srp72		7S RNA binding
Q9VQF9	SNAPIN protein homolog; Snapin		protein binding
Q00G30	Sno oncogene, isoform B; Snoo		
M9PCZ7	SoxNeuro, isoform B; SoxN	HMG box transcription factor	transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding; protein binding; chromatin DNA binding

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
P13395	Spectrin alpha chain; alpha-Spec		Actin binding; calcium ion binding; microtubule binding; cytoskeletal protein binding
Q24562	Splicing factor U2AF 50 kDa subunit; U2af50		RNA binding; mRNA binding; protein binding; poly-pyrimidine tract binding; pre-mRNA 3'-splice site binding
P92208	Stress-activated protein kinase JNK; bsk	non-receptor serine/threonine protein kinase	protein kinase activity; protein serine/threonine kinase activity; JUN kinase activity; MAP kinase activity; protein binding; ATP binding
O61307	Teneurin-m; Ten-m		protein binding; filamin binding; identical protein binding; protein homodimerization activity; protein heterodimerization activity
P18289	Transcription factor AP-1; Jra	basic leucine zipper transcription factor	RNA polymerase II proximal promoter sequence-specific DNA binding; RNA polymerase II transcription factor activity, sequence-specific DNA binding; transcription coactivator activity; protein binding; transcription factor binding; protein heterodimerization activity

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
P19538	Transcriptional activator cubitus interruptus; ci		RNA polymerase II regulatory region sequence-specific DNA binding; RNA polymerase II distal enhancer sequence-specific DNA binding; transcriptional activator activity, RNA polymerase II proximal promoter sequence-specific DNA binding; transcriptional activator activity, RNA polymerase II distal enhancer sequence-specific DNA binding; DNA binding; DNA binding transcription factor activity; protein binding; protein homodimerization activity; transcription regulatory region DNA binding
Q9VJ29	Transport and Golgi organization protein 6; Tango6		
P06603	Tubulin alpha-1 chain; alphaTub84B	tubulin	GTPase activity; structural constituent of cytoskeleton; GTP binding
P06605	Tubulin alpha-3 chain; alphaTub84D	tubulin	GTPase activity; structural constituent of cytoskeleton; GTP binding; myosin binding
Q24560	Tubulin beta-1 chain; betaTub56D	tubulin	GTPase activity; structural constituent of cytoskeleton; GTP binding
P08841	Tubulin beta-3 chain; betaTub60D	tubulin	GTPase activity; structural constituent of cytoskeleton; GTP binding
Q9VBE0	TweedleB; TwdlB		structural constituent of chitin-based cuticle

UniProt ID	Common Name	PANTHER Protein Class	GO Molecular Function
Q9VRP5	Ubiquitin carboxyl-terminal hydrolase 36; scny	cysteine protease	thiol-dependent ubiquitin-specific protease activity; protein binding; thiol-dependent ubiquitinyl hydrolase activity
Q9VFA8	Uncharacterized protein, isoform D; CG42404		molecular_function
Q9VPK4	Uncharacterized protein; CG11454		single-stranded RNA binding; mRNA binding
P31409	V-type proton ATPase subunit B; Vha55	ATP synthase; DNA binding protein; anion channel; hydrolase; ligand-gated ion channel	ATP binding; proton-transporting ATPase activity, rotational mechanism
M9PBJ1	Zormin, isoform J; Zormin		Actin binding; protein binding

# References

- Alfaro-Aco, R. and Petry, S. (2015). Building the Microtubule Cytoskeleton Piece by Piece. *The Journal of Biological Chemistry* *290*, 17154–17162.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* *25*, 3389–3402.
- Alvarado, D., Klein, D. E. and Lemmon, M. A. (2009). ErbB2 resembles an autoinhibited invertebrate epidermal growth factor receptor. *Nature* *461*, 287–291.
- Applewhite, D. A., Grode, K. D., Duncan, M. C. and Rogers, S. L. (2013). The actin-microtubule cross-linking activity of Drosophila Short stop is regulated by intramolecular inhibition. *Molecular Biology of the Cell* *24*, 2885–2893.
- Applewhite, D. A., Grode, K. D., Keller, D., Zadeh, A. D., Slep, K. C. and Rogers, S. L. (2010). The Spectraplakin Short Stop Is an ActinMicrotubule Cross-Linker That Contributes to Organization of the Microtubule Network. *Molecular Biology of the Cell* *21*, 1714–1724.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics* *25*, 25–29.
- Barabsi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* *5*, 101.
- Bartolini, F. and Gundersen, G. G. (2006). Generation of noncentrosomal microtubule arrays. *Journal of Cell Science* *119*, 4155–4163.
- Basha, O., Flom, D., Barshir, R., Smoly, I., Tirman, S. and Yeger-Lotem, E. (2015). MyProteinNet: build up-to-date protein interaction networks for organisms, tissues and user-defined contexts. *Nucleic Acids Research* *43*, W258–W263.
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-A-Jee, H., Cowley, A., Silva, A. D., Giorgi, M. D., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L.,

Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Renaux, A., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.-C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., Castro, E. d., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermaier, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.-L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L.-S. and Zhang, J. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research* *45*, D158–D169.

Baumbach, J., Novak, Z. A., Raff, J. W. and Wainman, A. (2015). Dissecting the Function and Assembly of Acentriolar Microtubule Organizing Centers in Drosophila Cells In Vivo. *PLOS Genetics* *11*, e1005261.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* *57*, 289–300.

Bharadwaj, A., Singh, D. P., Ritz, A., Tegge, A. N., Poirel, C. L., Kraikivski, P., Adames, N., Luther, K., Kale, S. D., Peccoud, J., Tyson, J. J. and Murali, T. M. (2017). GraphSpace: stimulating interdisciplinary collaborations in network biology. *Bioinformatics (Oxford, England)* *33*, 3134–3136.

Blanchoin, L., Boujemaa-Paterski, R., Sykes, C. and Plastino, J. (2014). Actin Dynamics, Architecture, and Mechanics in Cell Motility. *Physiological Reviews* *94*, 235–263.

Bottenberg, W., Sanchez-Soriano, N., Alves-Silva, J., Hahn, I., Mende, M. and Prokop, A. (2009). Context-specific requirements of functional domains of the Spectraplakin Short stop in vivo. *Mechanisms of Development* *126*, 489–502.

Calderone, A., Castagnoli, L. and Cesareni, G. (2013). mentha: a resource for browsing integrated protein-interaction networks. *Nature Methods* *10*, 690–691.

Callan-Jones, A. C. and Voituriez, R. (2016). Actin flows in cell migration: from locomotion and polarity to trajectories. *Current Opinion in Cell Biology* *38*, 12–17.

- Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., AmiGO Hub and Web Presence Working Group (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* (Oxford, England) *25*, 288–289.
- Case, L. B. and Waterman, C. M. (2015). Integration of actin dynamics and cell adhesion by a three-dimensional, mechanosensitive molecular clutch. *Nature Cell Biology* *17*, 955.
- Chabu, C. and Doe, C. Q. (2008). Dap160/intersectin binds and activates aPKC to regulate cell polarity and cell cycle progression. *Development* (Cambridge, England) *135*, 2739–2746.
- Chao, A. T., Jones, W. M. and Bejsovec, A. (2007). The HMG-box transcription factor SoxNeuro acts with Tcf to control Wg/Wnt signaling activity. *Development* *134*, 989–997.
- Chipuk, J. E., Bouchier-Hayes, L. and Green, D. R. (2006). Mitochondrial outer membrane permeabilization during apoptosis: the innocent bystander scenario. *Cell Death and Differentiation* *13*, 1396–1402.
- Currie, J. D., Stewman, S., Schimizzi, G., Slep, K. C., Ma, A., Rogers, S. L. and Holzbaur, E. L. F. (2011). The microtubule lattice and plus-end association of Drosophila Mini spindles is spatially regulated to fine-tune microtubule dynamics. *Molecular Biology of the Cell* *22*, 4343–4361.
- De Las Rivas, J. and Fontanillo, C. (2010). ProteinProtein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *PLoS Computational Biology* *6*.
- Debec, A., Dtraves, C., Montmory, C., Graud, G. and Wright, M. (1995). Polar organization of gamma-tubulin in acentriolar mitotic spindles of *Drosophila melanogaster* cells. *Journal of Cell Science* *108* (Pt 7), 2645–2653.
- Dickman, D. K. and Davis, G. W. (2009). The Schizophrenia Susceptibility Gene Dysbindin Controls Synaptic Homeostasis. *Science* (New York, N.Y.) *326*, 1127–1130.
- docs.python.org (2018). 5. Data Structures Python 3.6.5rc1 documentation.
- du Plessis, L., kunca, N. and Dessimoz, C. (2011). The what, where, how and why of gene ontologya primer for bioinformaticians. *Briefings in Bioinformatics* *12*, 723–735.
- Duhamel, S., Goyette, M.-A., Thibault, M.-P., Filion, D., Gaboury, L. and Ct, J.-F. (2018). The E3Ubiquitin Ligase HectD1 Suppresses EMT and Metastasis by Targeting the+TIP ACF7 for Degradation. *Cell Reports* *22*, 1016–1030.
- Eisman, R. C., Stewart, N., Miller, D. and Kaufman, T. C. (2006). centrosomin's beautiful sister (cbs) encodes a GRIP-domain protein that marks Golgi inheritance and functions in the centrosome cycle in *Drosophila*. *Journal of Cell Science* *119*, 3399–3412.

- Etienne-Manneville, S. (2013). Microtubules in Cell Migration. *Annual Review of Cell and Developmental Biology* *29*, 471–499.
- Ezratty, E. J., Partridge, M. A. and Gundersen, G. G. (2005). Microtubule-induced focal adhesion disassembly is mediated by dynamin and focal adhesion kinase. *Nature Cell Biology* *7*, 581–590.
- Fazekas, D., Koltai, M., Trei, D., Mdos, D., Plfy, M., Dl, Z., Zskai, L., Szalay-Bek, M., Lenti, K., Farkas, I. J., Vellai, T., Csermely, P. and Korcsmros, T. (2013). SignalLink 2 a signaling pathway resource with multi-layered regulatory networks. *BMC Systems Biology* *7*, 7.
- Fletcher, D. A. and Mullins, R. D. (2010). Cell mechanics and the cytoskeleton. *Nature* *463*, 485–492.
- Fong, K.-W., Au, F. K. C., Jia, Y., Yang, S., Zhou, L. and Qi, R. Z. (2017). Microtubule plus-end tracking of end-binding protein 1 (EB1) is regulated by CDK5 regulatory subunit-associated protein 2. *Journal of Biological Chemistry* *292*, 7675–7687.
- Franz, C. M., Jones, G. E. and Ridley, A. J. (2002). Cell Migration in Development and Disease. *Developmental Cell* *2*, 153–158.
- Gardel, M. L., Schneider, I. C., Aratyn-Schaus, Y. and Waterman, C. M. (2010). Mechanical Integration of Actin and Adhesion Dynamics in Cell Migration. *Annual review of cell and developmental biology* *26*, 315–333.
- Gautreau, A. and Krause, M. (2014). Steering cell migration: lamellipodium dynamics and the regulation of directional persistence. *Nature Reviews Molecular Cell Biology* *15*, 577.
- Goicoechea, S. M., Awadia, S. and Garcia-Mata, R. (2014). Im coming to GEF you: Regulation of RhoGEFs during cell migration. *Cell Adhesion & Migration* *8*, 535–549.
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBLEBI. *Nucleic Acids Research* *38*, W695–W699.
- Gramates, L. S., Marygold, S. J., Santos, G. d., Urbano, J.-M., Antonazzo, G., Matthews, B. B., Rey, A. J., Tabone, C. J., Crosby, M. A., Emmert, D. B., Falls, K., Goodman, J. L., Hu, Y., Ponting, L., Schroeder, A. J., Strelets, V. B., Thurmond, J. and Zhou, P. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Research* *45*, D663–D671.
- Guan, X., Middlebrooks, B. W., Alexander, S. and Wasserman, S. A. (2006). Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*. *Proceedings of the National Academy of Sciences* *103*, 16794–16799.

- Guilgur, L. G., Prudncio, P., Sobral, D., Liszekova, D., Rosa, A. and Martinho, R. G. (2014). Requirement for highly efficient pre-mRNA splicing during *Drosophila* early embryonic development. *eLife* *3*, e02181.
- Gunning, P. W., Ghoshdastider, U., Whitaker, S., Popp, D. and Robinson, R. C. (2015). The evolution of compositionally and functionally distinct actin filaments. *J Cell Sci* *128*, 2009–2019.
- Gupton, S. L., Eisenmann, K., Alberts, A. S. and Waterman-Storer, C. M. (2007). mDia2 regulates actin and focal adhesion dynamics and organization in the lamella for efficient epithelial cell migration. *Journal of Cell Science* *120*, 3475–3487.
- Humphries, M. J. (2000). Integrin Structure. *Biochemical Society Transactions* *28*, 311–340.
- Kapur, M., Wang, W., Maloney, M. T., Millan, I., Lundin, V. F., Tran, T.-A. and Yang, Y. (2012). Calcium tips the balance: a microtubule plus end to lattice binding switch operates in the carboxyl terminus of BPAG1n4. *EMBO reports* *13*, 1021–1029.
- Ketosugbo, K. F., Bushnell, H. L. and Johnson, R. I. (2017). A screen for E3 ubiquitination ligases that genetically interact with the adaptor protein Cindr during *Drosophila* eye patterning. *PLOS ONE* *12*, e0187571.
- Kishita, Y., Tsuda, M. and Aigaki, T. (2012). Impaired fatty acid oxidation in a *Drosophila* model of mitochondrial trifunctional protein (MTP) deficiency. *Biochemical and Biophysical Research Communications* *419*, 344–349.
- Kleiger, G. and Mayor, T. (2014). Perilous journey: a tour of the ubiquitin-proteasome system. *Trends in Cell Biology* *24*, 352–359.
- Koch, C. M., Honemann-Capito, M., Egger-Adam, D. and Wodarz, A. (2009). Windei, the *Drosophila* Homolog of mAM/MCAF1, Is an Essential Cofactor of the H3K9 Methyl Transferase dSETDB1/Eggless in Germ Line Development. *PLOS Genetics* *5*, e1000644.
- Kollman, J. M., Merdes, A., Mourey, L. and Agard, D. A. (2011). Microtubule nucleation by -tubulin complexes. *Nature Reviews Molecular Cell Biology* *12*, 709–721.
- Komander, D. (2010). Mechanism, Specificity and Structure of the Deubiquitinases. In *Conjugation and Deconjugation of Ubiquitin Family Modifiers: Subcellular Biochemistry*, (Groettrup, M., ed.), pp. 69–87. Springer New York New York, NY.
- Korenbaum, E. and Rivero, F. (2002). Calponin homology domains at a glance. *Journal of Cell Science* *115*, 3543–3545.
- Koster, J., Geerts, D., Favre, B., Borradori, L. and Sonnenberg, A. (2003). Analysis of the interactions between BP180, BP230, plectin and the integrin 64 important for hemidesmosome assembly. *Journal of Cell Science* *116*, 387–399.

- Langton, P. F., Colombani, J., Aerne, B. L. and Tapon, N. (2007). Drosophila ASPP Regulates C-Terminal Src Kinase Activity. *Developmental Cell* *13*, 773–782.
- Lawson, C. D. and Burridge, K. (2014). The on-off relationship of Rho and Rac during integrin-mediated adhesion and cell migration. *Small GTPases* *5*, e27958.
- Lee, S. and Kolodziej, P. A. (2002). Short Stop provides an essential link between F-actin and microtubules during axon extension. *Development* *129*, 1195–1204.
- Lee, S., Nahm, M., Lee, M., Kwon, M., Kim, E., Zadeh, A. D., Cao, H., Kim, H.-J., Lee, Z. H., Oh, S. B., Yim, J., Kolodziej, P. A. and Lee, S. (2007). The F-actin-microtubule crosslinker Shot is a platform for Krasavietz-mediated translational regulation of midline axon repulsion. *Development* *134*, 1767–1777.
- Lin, T.-Y., Huang, C.-H., Kao, H.-H., Liou, G.-G., Yeh, S.-R., Cheng, C.-M., Chen, M.-H., Pan, R.-L. and Juang, J.-L. (2009). Abi plays an opposing role to Abl in Drosophila axonogenesis and synaptogenesis. *Development* *136*, 3099–3107.
- Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., McLaren, P., North, P., Rana, D., Riley, T., Sullivan, J., Watkins, X., Woodbridge, M., Lilley, K., Russell, S., Ashburner, M., Mizuguchi, K. and Micklem, G. (2007). FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biology* *8*, R129.
- Ma, Y., Yue, J., Zhang, Y., Shi, C., Odenwald, M., Liang, W. G., Wei, Q., Goel, A., Gou, X., Zhang, J., Chen, S.-Y., Tang, W.-J., Turner, J. R., Yang, F., Liang, H., Qin, H. and Wu, X. (2017). ACF7 regulates inflammatory colitis and intestinal wound response by orchestrating tight junction dynamics. *Nature Communications* *8*, 15375.
- Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., Kanapin, A., Lewis, S., Mahajan, S., May, B., Schmidt, E., Vastrik, I., Wu, G., Birney, E., Stein, L. and D'Eustachio, P. (2009). Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research* *37*, D619–622.
- Mi, H., Muruganujan, A., Casagrande, J. T. and Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols* *8*, 1551–1566.
- Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G. G. and Finley, R. L. (2011). DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Research* *39*, D736–D743.
- Nadzirin, N. and Firdaus-Raih, M. (2012). Proteins of Unknown Function in the Protein Data Bank (PDB): An Inventory of True Uncharacterized Proteins and Computational Tools for Their Analysis. *International Journal of Molecular Sciences* *13*, 12761–12772.

- Nag, R. N., Niggli, S., Sousa-Guimares, S., Vazquez-Pianzola, P. and Suter, B. (2018). Mms19 is a mitotic gene that permits Cdk7 to be fully active as a Cdk-activating kinase. *Development* *145*, dev156802.
- Nashchekin, D., Fernandes, A. R. and St Johnston, D. (2016). Patronin/Shot Cortical Foci Assemble the Noncentrosomal Microtubule Array that Specifies the Drosophila Anterior-Posterior Axis. *Developmental Cell* *38*, 61–72.
- Nilton, A., Oshima, K., Zare, F., Byri, S., Nannmark, U., Nyberg, K. G., Fehon, R. G. and Uv, A. E. (2010). Crooked, Coiled and Crimped are three Ly6-like proteins required for proper localization of septate junction components. *Development* *137*, 2427.
- Ning, W., Yu, Y., Xu, H., Liu, X., Wang, D., Wang, J., Wang, Y. and Meng, W. (2016). The CAMSAP3-ACF7 Complex Couples Noncentrosomal Microtubules with Actin Filaments to Coordinate Their Dynamics. *Developmental Cell* *39*, 61–74.
- Oda, Y., Okada, T., Yoshida, H., Kaufman, R. J., Nagata, K. and Mori, K. (2006). Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *The Journal of Cell Biology* *172*, 383–393.
- O'Sullivan, N. C., Jahn, T. R., Reid, E. and O'Kane, C. J. (2012). Reticulon-like-1, the Drosophila orthologue of the Hereditary Spastic Paraplegia gene reticulon 2, is required for organization of endoplasmic reticulum and of distal motor axons. *Human Molecular Genetics* *21*, 3356–3365.
- Pollard, T. D. and Cooper, J. A. (2009). Actin, a Central Player in Cell Shape and Movement. *Science* *326*, 1208–1212.
- Pufall, M. A. and Graves, a. B. J. (2002). Autoinhibitory Domains: Modular Effectors of Cellular Regulation. *Annual Review of Cell and Developmental Biology* *18*, 421–462.
- Ridley, A. J. (2015). Rho GTPase signalling in cell migration. *Current Opinion in Cell Biology* *36*, 103–112.
- Ritz, A. (2017). <https://github.com/annaritz/fly-interactome> Fly interactome compiled from a set of publicly-available sources. original-date: 2017-08-16T22:41:39Z.
- Roeper, K., Gregory, S. L. and Brown, N. H. (2002). The ‘Spectraplakins’: cytoskeletal giants with characteristics of both spectrin and plakin families. *Journal of Cell Science* *115*, 4215–4225.
- Rogers, G. C., Rusan, N. M., Peifer, M. and Rogers, S. L. (2008). A Multicomponent Assembly Pathway Contributes to the Formation of Acentrosomal Microtubule Arrays in Interphase Drosophila Cells. *Molecular Biology of the Cell* *19*, 3163–3178.

- Sevimoglu, T. and Arga, K. Y. (2014). The role of protein interaction networks in systems biomedicine. *Computational and Structural Biotechnology Journal* *11*, 22–27.
- Shemesh, T., Verkhovsky, A. B., Svitkina, T. M., Bershadsky, A. D. and Kozlov, M. M. (2009). Role of Focal Adhesions and Mechanical Stresses in the Formation and Progression of the Lamellum Interface. *Biophysical Journal* *97*, 1254–1264.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Sding, J., Thompson, J. D. and Higgins, D. G. (2011). Fast, scalable generation of highquality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* *7*, 539.
- Spasi, M. R., Callaerts, P. and Norga, K. K. (2008). *Drosophila* alicorn is a neuronal maintenance factor protecting against activity-induced retinal degeneration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* *28*, 6419–6429.
- Steuer, R. and Lopez, G. Z. (2008). Global Network Properties. In *Analysis of Biological Networks*, (Junker, B. H. and Schreiber, F., eds), pp. 29–63. John Wiley & Sons, Inc. DOI: 10.1002/9780470253489.ch3.
- Stoddard, P. R., Williams, T. A., Garner, E., Baum, B. and Drubin, D. G. (2017). Evolution of polymer formation within the actin superfamily. *Molecular Biology of the Cell* *28*, 2461–2469.
- Subramanian, A., Prokop, A., Yamamoto, M., Sugimura, K., Uemura, T., Betschinger, J., Knoblich, J. A. and Volk, T. (2003). Shortstop recruits EB1/APC1 and promotes microtubule assembly at the muscle-tendon junction. *Current biology: CB* *13*, 1086–1095.
- Sun, D., Leung, C. L. and Liem, R. K. (2001). Characterization of the microtubule binding domain of microtubule actin crosslinking factor (MACF): identification of a novel group of microtubule associated proteins. *Journal of Cell Science* *114*, 161–172.
- Sun, Z., Guo, S. S. and Fssler, R. (2016). Integrin-mediated mechanotransduction. *The Journal of Cell Biology* *215*, 445.
- Swaminathan, V., Kalappurakkal, J. M., Mehta, S. B., Nordenfelt, P., Moore, T. I., Koga, N., Baker, D. A., Oldenbourg, R., Tani, T., Mayor, S., Springer, T. A. and Waterman, C. M. (2017). Actin retrograde flow actively aligns and orients ligand-engaged integrins in focal adhesions. *Proceedings of the National Academy of Sciences* *114*, 10648–10653.
- Takcs, Z., Jankovics, F., Vilmos, P., Lnrt, P., Rper, K. and Erdlyi, M. (2017). The spectraplakin Short stop is an essential microtubule regulator involved in epithelial closure in *Drosophila*. *Journal of Cell Science* *130*, 712–724.

- Vidal, M., Cusick, M. and Barabsi, A.-L. (2011). Interactome Networks and Human Disease. *Cell* *144*, 986–998.
- Voelzmann, A., Liew, Y.-T., Qu, Y., Hahn, I., Melero, C., Snchez-Soriano, N. and Prokop, A. (2017). Drosophila Short stop as a paradigm for the role and regulation of spectraplakins. *Seminars in Cell & Developmental Biology* *69*, 40–57.
- Wang, S., Reuveny, A. and Volk, T. (2015). Nesprin provides elastic properties to muscle nuclei by cooperating with spectraplakin and EB1. *J Cell Biol* *209*, 529–538.
- Wang, S. and Volk, T. (2015). Composite biopolymer scaffolds shape muscle nucleus: Insights and perspectives from Drosophila. *Bioarchitecture* *5*, 35–43.
- Wilson, K. L. and Berk, J. M. (2010). The nuclear envelope at a glance. *J Cell Sci* *123*, 1973–1978.
- Wittmann, T., Bokoch, G. M. and Waterman-Storer, C. M. (2003). Regulation of leading edge microtubule and actin dynamics downstream of Rac1. *The Journal of Cell Biology* *161*, 845–851.
- Wu, X., Kodama, A. and Fuchs, E. (2008). ACF7 Regulates Cytoskeletal-Focal Adhesion Dynamics and Migration and Has ATPase Activity. *Cell* *135*, 137–148.
- Wu, X., Shen, Q.-t., Oristian, D. S., Lu, C. P., Zheng, Q., Wang, H.-W. and Fuchs, E. (2011). Skin Stem Cells Orchestrate Directional Migration By Regulating Microtubule-ACF7 Connections Through GSK3. *Cell* *144*, 341–352.
- Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.-F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A.-S., Motyl, A., Hudson, M. E., Park, J., Xin, X., Cusick, M. E., Moore, T., Boone, C., Snyder, M., Roth, F. P., Barabsi, A.-L., Tavernier, J., Hill, D. E. and Vidal, M. (2008). High Quality Binary Protein Interaction Map of the Yeast Interactome Network. *Science (New York, N.Y.)* *322*, 104–110.
- Yu, H., Luscombe, N. M., Lu, H. X., Zhu, X., Xia, Y., Han, J.-D. J., Bertin, N., Chung, S., Vidal, M. and Gerstein, M. (2004). Annotation Transfer Between Genomes: ProteinProtein Interologs and ProteinDNA Regulogs. *Genome Research* *14*, 1107–1118.
- Yu, W., Solowska, J. M., Qiang, L., Karabay, A., Baird, D. and Baas, P. W. (2005). Regulation of Microtubule Severing by Katanin Subunits during Neuronal Development. *The Journal of Neuroscience : the official journal of the Society for Neuroscience* *25*, 5573–5583.
- Yue, J., Xie, M., Gou, X., Lee, P., Schneider, M. and Wu, X. (2014). Microtubules Regulate Focal Adhesion Dynamics through MAP4K4. *Developmental Cell* *31*, 572–585.

- Yue, J., Zhang, Y., Liang, W. G., Gou, X., Lee, P., Liu, H., Lyu, W., Tang, W.-J., Chen, S.-Y., Yang, F., Liang, H. and Wu, X. (2016). In vivo epidermal migration requires focal adhesion targeting of ACF7. *Nature Communications* *7*, ncomms11692.
- Zaoui, K., Benseddik, K., Daou, P., Salan, D. and Badache, A. (2010). ErbB2 receptor controls microtubule capture by recruiting ACF7 to the plasma membrane of migrating cells. *Proceedings of the National Academy of Sciences* *107*, 18517–18522.
- Zhang, J., Yue, J. and Wu, X. (2017). Spectraplakin family proteins cytoskeletal crosslinkers with versatile roles. *J Cell Sci* *130*, 2447–2457.