Design, implementation and experimental validation of a network-based model to predict mitotic microtubule regulating proteins

Faisal F. Khan
St. Anne's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2013

College Affiliation

St. Anne's College

1st Supervisor

Professor Charlotte Deane,

Department of Statistics,

University of Oxford

2nd Supervisor

Professor James Wakefield

Department of Biosciences,

University of Exeter

Research funding

Pakistan Higher Education Commission

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

To Ami and Baba.

Acknowledgments

First and foremost, I am very grateful to my supervisors, Professor Charlotte Deane and Professor James Wakefield, for giving me the opportunity to work with them. I would thank Professor Deane in particular, for trusting my abilities and giving me the confidence to step into an area of science that was relatively new to me. Thank you for their patience, motivation and continuing support.

I also wish to thank both of them for bearing with me if I strayed and pursued other interests during my work, especially for their support in taking up the SIP offer to study at the Business School. All this has brought immense value and richness into my Oxford experience.

During the write up of this work, I am very grateful for their precious time and dedication in correcting and commenting on multiple versions of this manuscript and helping me improve my writing skills.

I would also like to express my gratitude to my examiners, Dr Clive Wilson and Dr Daimark Bennett, for agreeing to assess my thesis, for making my viva enjoyable experience and for all their questions, comments and suggestions.

Thank you to all the lab members of the Wakefield group at Oxford, including Yaseen, Simon, Alan and especially Tommy, who helped me generously in everything - from sharpening my pipetting skills and knocking over flies to cloning and silencing genes. I thoroughly enjoyed my stay with all of them for which I am very grateful.

Many thanks to all members of the Deane group who welcomed me warmly into the lab including Seb, Mireille, Anna, Rebecca, and especially Waqar who has been very patient as I learned many things from scratch. It was really great to have spent some quality time with Tiago, JP, Jamie, Konrad, Leila, James, Henry and Hannah and I really wish I had been with them for longer. Thank you to all for the great company in Vienna at ISMB 2011.

At Exeter, I had a great time in the new Wakefield lab, with Dan, Pete, Jack, Valeria, Magdalena and Sarah. Thank you each one of you for always being there, for all the stimulating discussions and those great fun moments in and outside

the lab (especially Jack's 'jokes'!). I am very grateful to all who helped me learn new methods in biology and for being patient with the tons of questions I ask – protein gels and embryo imaging (Dan), spin- and pull-downs (Pete) and fly crosses (Jack). Thanks to everyone for the antibodies they donated, for the reagents they shared and all the tips I managed to steal from their benches.

A big thank you and Jazakallah, to Ahmad and Amandla and the Exeter ISoc family for the great company outside the lab, especially when it used to get lonely and in Ramadan. Thank you to the MegaKebab owners and crew on Sidwell Street for bringing a taste and some diversity to my Kebab staple.

I will particularly remember 2011 with my house-mates Ahmar, Said and Zabair at 345 Iffley Road. It has easily been one of the most memorable ones in my life. Thank you for each single one of the great moments we had spent together including the heated debates, the weirdest Biryanis and the funny 'silence policy'.

My deepest thanks to all those friends and brothers who made my transition to a new country smoother than I could have ever imagined and for making me feel at home even though I was thousands of miles away. A very big thank you to Moudud (my first friend here in the UK!) for teaching me so much about the English culture and British Muslim etiquettes and Habib for brushing up my English and teaching me to differentiate between a 'V' and a 'W' when I speak. And of course, Zabair, for helping me massively with both and for always being there when I needed help.

Thank you very much Oxford and Exeter for a life-changing experience and all the inspiring friends and colleagues. I pray and hope to become capable enough to give back to you one day.

I can't be more grateful to my wife for her patience in my absence and for bringing up our lovely daughter when I was not there at all to help. I have missed some special years of her life and I hope somehow, magically, I make it worth it for them and for all of us.

Thank you to my wife, Maryam, for being there during my thesis write up and for her support. Those few months were the loveliest I spent at Oxford. Thank you to the crew at Café Nero on High Street for that great cup of Mocha, every morning, for the last couple of months of my write up, when Maryam had left.

My deepest gratitude to both my parents, Ami and Baba, for their unconditional love, care and support since the beginning of my time. Thank you for going beyond my needs and comfort, and fulfilling every single outrageous demand I made – from my first ever microscope when I was 10 and a brand new Pentium the day it was released, to all the mighty ones which I won't mention here! Thank you for each single one of them. Thank you for all the values, the freedom and the

confidence you brought into my life. Nothing would have ever been possible without your sacrifices, your prayers and your faith in me.

I would also like to thank all my teachers here, Miss Samina and Miss Aisha from school, Mr Nisar Ayub, Mr Naveed Tabassum and Mr Abdul Wahid from College and Dr Noor Muhammad from University, for showing me the light and giving me the confidence to pursue my dreams.

I am indebted to the Pakistan Higher Education Commission for funding my studies and stay in Oxford. I hope I can give back to all the taxpayers who made this possible. A big thank you to my father again for supplementing my stipend and making my life much easier. And of course, my Kebab excesses, my Union membership, my regular trips to Exeter and London and the piles of books from Amazon would not have existed without your help.

Finally, all praises to Allah, who blessed me with the gift of Islam. A gift which encourages reason and reflection. A gift which I found to be the most articulate in explaining life and which has given me a clear purpose. A gift which made me strong enough to go through difficult times during this project in particular, (alone, for five years, away from home and my 2 year-old!), and in life in general.

Abstract

The purpose of this thesis was to study mitosis in *Drosophila*, from a network biology perspective. The primary aim was to develop and test a network-based prediction model that could integrate available data in public databases (like Flybase) and, based on that, predict potential mitotic proteins.

The approach taken to design the protein interaction network included the use of *a priori* knowledge about the microtubule composition of the mitotic spindle and the higher likelihood of microtubule-associated proteins (MAPs) to have a putative mitotic function. The design also included the integration of different complementary datasets, from gene expression and functional RNAi screens to cross species conservation of MAPs for fitting a network-based model for predicting mitotic proteins.

I begin with the creation of the MAP interactome based on a MAP dataset in *Drosophila*. This initial network was extended by transferring homologs and interologues of MAP datasets from four other species, i.e. human, mouse, rat and *Arabidopsis*. These proteins were then used as seed proteins to conduct a virtual pull-down experiment, by adding indirect interactors into the network, i.e. proteins that directly bind to two or more MAPs within the network, which completed the MAP interactome. Data from genome-wide studies in *Drosophila* were gathered for each node in the MAP interactome. These 'layers' of data were then used as features to fit a prediction model that could score each node in the network, based on the likelihood of its role in mitosis. The final model performed with 96% accuracy after 10-fold cross validation and was used to rank all the proteins in the MAP interactome.

By analysing the top 100 high scoring predicted mitotic proteins, a highly connected cluster of 33 proteins was identified that was subject to experimental validation in the lab. The first approach was to conduct an *in vitro* analysis using an RNAi screen to test for any spindle, chromosome or centrosome phenotypes upon gene knockdown. After two independent RNAi screens, around 80% of the proteins produced mutant mitotic phenotypes strongly supporting the results of the MAP prediction model.

The second approach was to conduct an *in vivo* analysis by expressing GFP-fusion constructs of selected genes from the subcluster. These were expressed in *Drosophila* early embryos to study their subcellular localization during interphase and mitosis. A variety of localizations were observed ranging from chromatin and microtubules to more generic cytoplasmic localizations. These results suggested not all predicted proteins were co-localizing with microtubules, and therefore might not necessarily be microtubule associated proteins but can possibly be functioning as microtubule associated regulator proteins. Proteomics analysis of a subset of these genes showed a large proportion of false positive interactions but also picked new interactions between member proteins that highlighted a module within the subcluster.

The RNAi hits from the *in vitro* analysis and the members of the module within subcluster-16 from the *in vivo* analysis provide interesting subjects for further characterization.

Table of Contents

List of Abbreviations

AIC	Akaike information criterion
AP	Affinity purification
APC	Anaphase promoting complex
AUC	Area under the curve
BIND	Bio-molecular Interaction Network Databases
BioGRID	Biological General Repository for Interaction Datasets
BLAST	Basic Local Alignment Search Tool
CBP	Calmodulin-binding protein
CDD	Conserved domain database
CDK	Cyclin dependent kinase
co-IP	Co-immuno-precipitation
ConA	Concanavalin A
CPC	Chromosome passenger complex
DAPI	4,6-dimamidino-2-phenylindole
DGRC	Drosophila Genome Research Centre
DIP	Database for Interacting Proteins
ECL	Enhanced chemiluminescence
EMS	Ethyl methanesulphonate
GFP	Green fluorescent protein
GI	Genetic interactions
GO	Gene ontology
HT	High-throughput
LC	Liquid chromatography
MALDI	Matrix-assisted laser dissolution/ionization
MAP	Microtubule-associated protein
MCC	Mitotic checkpoint complex

MHC	Major histocompatibility complex
MINT	Molecular Interaction Database
MIPS	Mammalian Protein-protein Interaction Database
MS	Mass spectrometry
MT	Microtubule
MTOC	Microtubule organizing centre
NEB	Nuclear envelope breakdown
PBS	Phosphate buffer saline
PCM	Peri-centriolar material
PIN	Protein Interaction Network
PPI	Protein-protein interaction
RNAi	RNA interference
ROC	Receiver-operator curve
RSS	Residual sum of squares
SAC	Spindle assembly checkpoint
SILAC	Stable isotope labelling with amino acids in cell culture
SPL	Shortest path length
TAP	Tandem affinity purification
TEV	Tobacco etch virus
UAS	Upstream activation sequence
Y2H	Yeast-two-hybrid

List of Figures

List of Tables

Chapter 1
Introduction

Throughout the last century, genes and proteins have mainly been studied individually with a reductionist approach. But the advent of new high-throughput techniques in biology and the subsequent emergence of astonishing volumes of data in the post-genomic era has led to a shift in the attitude of biologists. It is becoming more common to look collectively at these individual components, whether genes or proteins, the interactions of these components and their resulting implications on a systems level. Since genes and proteins rarely act alone, it is likely that studying the interactive network they are part of is crucial to attain a deeper understanding of their biology.

It was after decades of progress made by scientists in understanding the molecular components of biological systems that a resurgence of the 'systems' perspective was observed, and towards the end of the 20th century Systems Biology emerged as a field. This did not happen solely because of the advent of new high throughput methodologies and unprecedented amounts of data but also due to an increasing trend of multidisciplinary strategies and cross talk between biology and different disciplines like physics, statistics, mathematics and even electronics (Agapakis & Silver 2009).

Despite the surge of genomic and post-genomic data and the multidisciplinary approach of studying it, many still argue that we have made little progress in

establishing a strong understanding of genotype-phenotype relationships and translating the abundance of data into meaningful knowledge (Houle et al. 2010).

One of the main aims of Systems Biology is to put together the knowledge we have about cellular components and try to make sense of the broader picture of events within the cell by studying the components, their dynamics, control mechanisms and underlying design principles (Kitano 2002). One way of approaching this is to construct networks of these interacting molecules and study their local and global patterns of interactions. Mathematical and statistical modelling can then follow, which can help in assessing these interactions and predicting new ones. Several types of interaction networks have been experimentally mapped including metabolic (Jeong et al. 2000; Kim et al. 2011), genetic (Boone et al. 2007; Dixon et al. 2009) and protein-protein (Uetz et al. 2000; LaCount et al. 2005; Rual et al. 2005) interactions.

This introductory chapter will provide an overview of protein interaction networks, touching upon the experimental sources of interaction data, their properties, methods of analysis and some of their applications in biology. This will be followed by an introduction to microtubules and microtubule-associated proteins and the role they play in the intricate cellular processes of mitosis. This chapter will end with a brief introduction to the model system used in this study – *Drosophila melanogaster*, commonly known as the fruit fly.

1.1 Protein Interaction Networks

Protein interaction networks, or PINs, are simplified diagrams of the actual protein-protein interactions taking place in the cell. These networks, mathematically known as graphs, started to emerge in the 1990s with an aim of summarizing complex cellular systems, in a way that could aid in their study. They contain a series of nodes and edges, which represent proteins and their interactions, respectively (Figure 1.1).

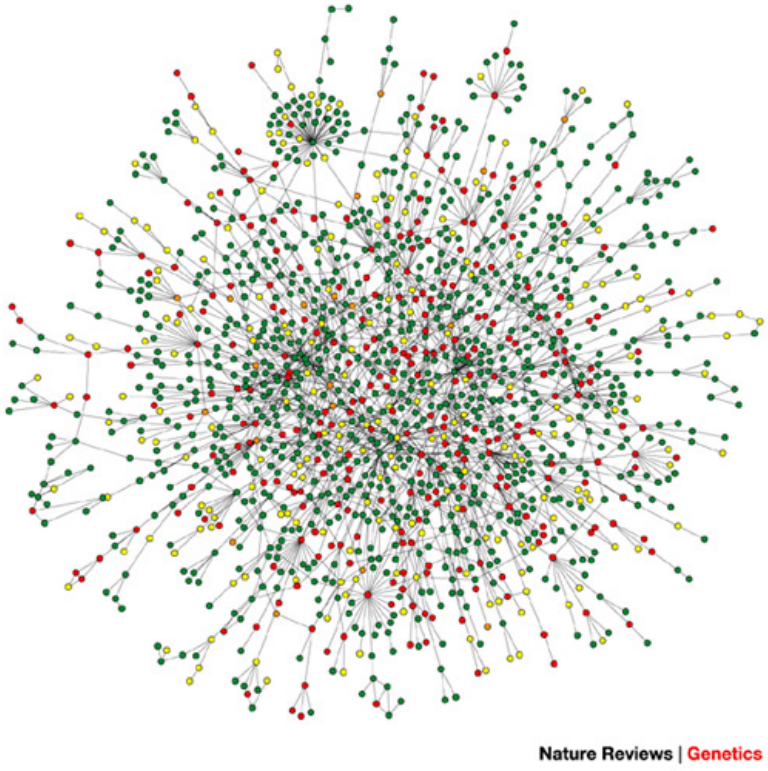


Figure 1.1: An example of a biological network – the *S. cerevisiae* interactome. The nodes represent individual proteins and the edges represent the interaction between the two proteins at its ends. Nodes are coloured based on phenotypic effects of the proteins: red (lethal), green (nonlethal), orange (slow growth), and yellow (unknown). Image taken from (Jeong et al. 2001).

1.1.1 Definition of a protein interaction

Defining a protein interaction is challenging and there has been much debate on the exact definition of a protein-protein interaction (PPI). This is because proteins are always in some sort of contact with many other proteins, e.g. during their synthesis, folding, transport, localization and ultimately their degradation within a cell. Questions arise as to whether PPIs include only the obligate contacts a protein makes with another, like in a protein complex, or whether it is inclusive of all the proteins 'touched' or 'bumped into' within a cell (De Las Rivas & Fontanillo 2010). For an interaction to hold any functional meaning, it is expected to be specific in nature, not accidental, and should be occurring via dedicated protein interfaces.

This assumption is also visible in the underlying design principles of the two main experimental sources of PPI data (explained in the following section). Protein-protein interactions can be direct in nature, or indirect, for example through scaffold proteins in large protein complexes like ribosomes or membrane proteins. The current consensus in the field is that PPIs are '*specific physical contacts between pairs of proteins that occur by selective docking in a particular biological context*' (De Las Rivas & Fontanillo 2010).

This means a PPI needs to be a specific, non-generic interaction between two proteins that have a common biological function. However, PPIs are not hard-wired and can change with changing environment. This means that the use of standardized experimental protocol and stating the biological context of PPI datasets is important when such data is reported.

1.1.2 Sources of PPI data

Protein interaction networks are often compiled using PPI data from many different sources. The two main high-throughput (HT) experimental sources of protein-protein interactions are yeast-2-hybrid and co-complex screens (De Las Rivas & Fontanillo 2010). Initially they were regarded as being biased and unsystematic but they have been improving in recent years as more standardized protocols are employed (Koegl & Uetz 2007; Orchard et al. 2007). Their speed and straightforward implementation has allowed the study of several near-complete whole-organism interactomes (Dreze et al. 2010; Giot et al. 2003; H. Yu et al. 2008).

Another source of PPI data is literature mining of published data from different laboratories that are working on individual proteins and complexes (He et al. 2009; Zhou & He 2008). This data is highly variable when it comes to experimental conditions and the systems they are studied in. In addition to these experimental sources, computational methods have been used to predict interaction data (Liu & Chen 2012).

Although PPI data from large-scale screens is error-prone and has several limitations depending on the method, they have the ability to provide a good estimate of the size of interactomes in particular organisms that can guide further research within these species. The coverage and quality of PPI data is still challenged, as we will discuss in the following sections (Section 2.2.3). This makes it important to understand the methodology used to generate PPI data

and the experimental conditions in order to grasp the strengths and limitations of each method.

The two main types of experimental approach that are currently in use, the binary approach and the co-complex approach (Figure 1.2), are described in more detail below.

Binary PPI data

The binary approach uses methods like yeast two-hybrid (Y2H) to establish pair-wise direct interactions between proteins within the proteome of a cell.

The yeast two-hybrid system is based on the GAL4/UAS system. The GAL4 protein regulates the transcription of genes that allow yeast cells to metabolise galactose. This transcription factor is made of two domains, i.e. an N-terminus DNA-binding domain that binds the upstream activation sequence (UAS) and a C-terminus activating domain. This system was exploited by Fields and Song, who developed the first two-hybrid system in yeast for detecting binary protein-protein interaction (Fields & Song 1989).

The DNA sequence encoding the two domains of the GAL4 proteins was bound to that of two (bait and prey) proteins that are being tested for an interaction. If the two test proteins have affinity towards each other, they bring together the two halves of the reporter protein, making it functional. The end product is a fluorescent readout or the ability to grow on selective media, which indicates the presence of an interaction.

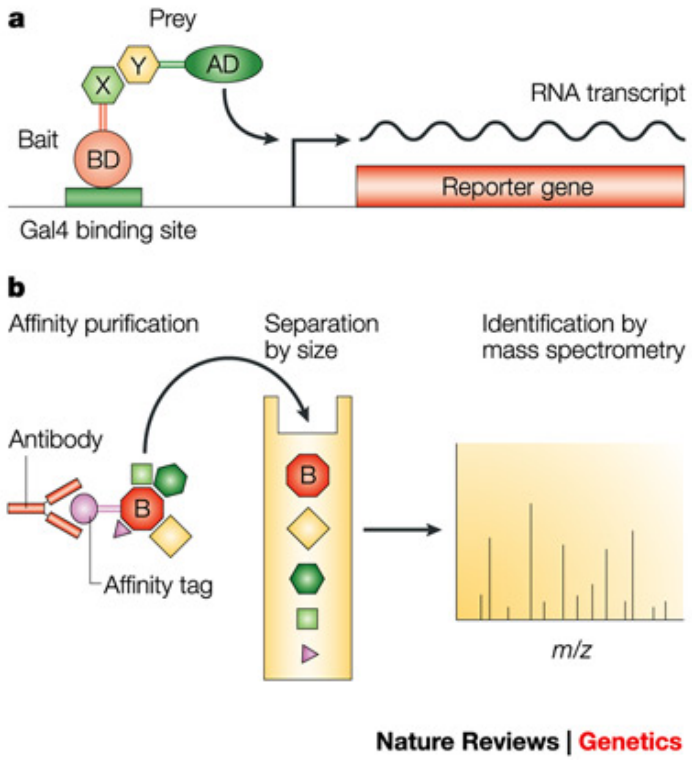


Figure 1.2: The two main HT sources of PPI data. (a) The yeast-two-hybrid system where the two putative interacting proteins (X and Y) bound as bait and prey activate the Gal4 hybrid system upon interaction. (b) The affinity purification is based on the affinity of the protein (bait) via a fused tag. The samples are then resolved and analysed by mass spectrometry. Image taken from (Grünenfelder & Winzeler 2002).

There are several modifications to the original protocol, which have been widely used for HT experiments such as screening large cDNA libraries for protein interactions (Koegl & Uetz 2007). Several key datasets have been published for model organisms like Yeast (Ito et al. 2001; Uetz et al. 2000; H. Yu et al. 2008), Fly (Giot et al. 2003), Worm (S. Li et al. 2004) and also Human (Dreze et al. 2010).

PPI data from co-complex experiments

Co-complex techniques coupled with mass spectrometry are the second main source of physical PPI data. Since this method has been employed to analyse the MAP interactome in our study (Chapter 4), we examine it in more detail.

Technological advances in mass spectrometry (MS) and recent developments in affinity purification (AP) techniques have allowed the production of near–complete proteomic datasets in different organisms. In fact the field is moving fast from resolution and identification of proteins in complex samples, towards quantification of protein abundances and studying temporal variations and post-translational modifications (Cox & Mann 2011).

Apart from different methodologies of *expression* and *modification proteomics* (Gstaiger & Aebersold 2009), *interaction proteomics* exploits this established protein analysis pipeline (Figure 1.3) to study samples from affinity purification or co-complex 'pull down' experiments to identify *in vivo* or *in vitro* interactors of proteins of interest. Affinity handles or tags are fused to 'bait' proteins and expressed in host systems of choice. Specific antibodies are then used to pull down these baits along with their interacting 'prey' proteins. This method has gained tremendous prominence in the field, especially the modified tandem affinity purification (TAP) method and has evolved into an established methodology over the last decade (Cox & Mann 2011).

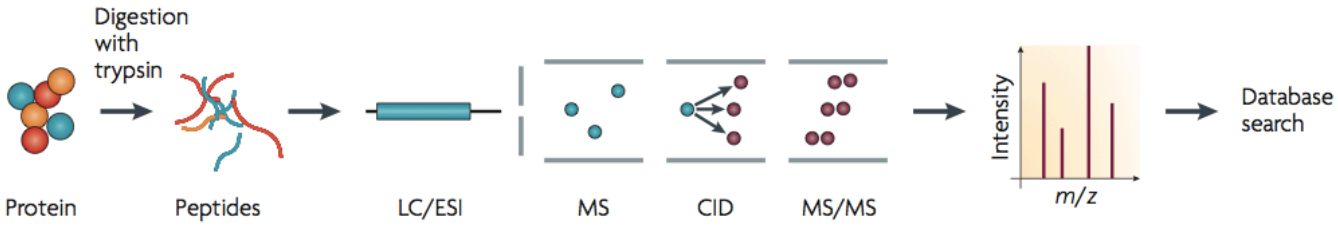


Figure 1.3: The workflow for a routine proteomics analysis of a protein sample including protein digestion, liquid chromatography, electrospray ionization and tandem mass spectrometry followed by analysis. Figure taken from (Gstaiger & Aebersold 2009).

In a standard co-complex experiment, a protein tag is fused with the bait protein and expressed in host cells or organisms. The tag can be as simple as the green fluorescent protein (GFP) or a variant of the popular TAP tags that can be captured by binding to two (tandem) affinity columns. The fusion protein is then pulled down based on affinity towards an (anti-tag) antibody along with its interactors and purified from the crowded cell lysate (Figure 1.2b). The purification step removes unbound proteins, but is gentle enough to leave bound ones. The pull down sample is then digested using the enzyme, trypsin, which results in a peptide fingerprint that can be analysed using MS. The digested sample is then resolved through a gradient of aqueous/organic solvent using high-pressure liquid chromatography.

These peptides are then ionised and converted into intact gas phase ions, which are then analysed by MS. Two different methods exist for conducting the ionization, i.e. electrospray methods (Fenn et al. 1989) and the matrix assisted laser dissolution/ionization (MALDI) method (Hillenkamp et al. 1991), differing in their efficiency and sensitivity.

Recent models of mass spectrometers take a fraction of a second to scan the entire range of masses producing mass spectra, which are then analysed. In the more recently adopted methodology called tandem mass spectrometry (MS/MS), a fraction of the peptides are dissociated (a process called CID, or collision-induced dissociation) and analysed in another simultaneous run of mass spectrometry, producing results with improved accuracy. The time taken by an average run depends on the length of the gradient and the number of fractions and replicates analysed in the experiment (Cox & Mann 2011).

TAP tags used in co-complex methods, especially HT screens, are derived from the classical ones used in yeast (Gavin et al. 2002; Ho et al. 2002). The ProtA-TEV-CBP tag is quite large in size increasing the chances of incorrect localization and/or folding of expressed proteins, and may also lead to expression and affinity problems in systems other than yeast. This led to the development of a diverse set of variants of the original TAP tags by different groups which are smaller in size (Zeghouf et al. 2004; Gloeckner et al. 2007), have higher binding affinities (Drakas et al. 2005), improved elution efficiencies (Schimanski et al. 2005), higher yield and purity (Bu¨rckstu¨mmer et al. 2006) and better stability (Guerrero et al. 2006).

Sample preparation in principle appears to be a simple procedure but is the most daunting task and has a lot of room for innovation. Preserving different types of interactions through a process that involves cell lysis and protein solubilisation is a huge challenge (Cox & Mann 2011; Chait 2011).

Apart from control bait proteins, many studies have used quantitative MS methods especially metabolic labelling like stable isotope labelling with amino acids in cell culture (SILAC) (Mann 2006; Ong et al. 2002) to distinguish between background binders (Wepf et al. 2009; Gingras et al. 2007). For better resolution, interactions from co-complex experiments can be coupled with binary data from Y2H experiments, to resolve into direct and indirect, and in some cases even to identify the contact points of interacting proteins (Boxem et al. 2008).

1.1.3 Limitations of experimental methods

Initially PPI data faced a number of serious challenges, which made biologists very sceptical about the information contained in such networks. One of the main inconsistencies was the poor overlap between different studies of the same system (Gentleman & Huber 2007). This was thought to be predominantly due to false negatives that were explained by the presence of under-represented proteins and the presence of transient interactions between proteins such as those during post-translational modifications like phosphorylation. False positives in many of these high-throughput screens were another inherent problem in the methods used at that time, for example self-activating baits in yeast two-hybrid screens and 'sticky' proteins in co-purification assays (Gentleman & Huber 2007; Hart et al. 2006).

As experimental methods have improved, the accuracy and coverage of studies have also improved. Experimental adaptations include regular auto-activation assays with Y2H screens, development of new purification tags for TAP pull-downs, use of statistical assessment of high-throughout (HT) data and the

enrichment of protein samples from cell extracts improving the coverage of datasets (Koegl & Uetz 2007). All these developments have helped to reduce the early scepticism of molecular biologists faced by interaction networks.

Technical limitations of the Y2H system, which are still faced, include the inability of some proteins to interact in the nucleus of yeast (most proteins function in the cytosolic region of the cell), along with its inability to test interactions between integral membrane proteins (Koegl & Uetz 2007). Binary data from Y2H experiments lack details such as whether the interaction occurs simultaneously as part of a complex or not and whether it requires any post-translational modification to occur, outside the experimental host cell, in its original cell type or organism. Another significant source of error, mainly false positives, is the fact that interactions reported by the Y2H system are independent of the endogenous expression levels of the protein in its native environment (Huang et al. 2007).

Another problem for data from Y2H screens is whether the interactions reported are mere biophysical interactions occurring in the experimental system or whether they are actual ones that occur inside the cell? Do these proteins even see each other in the cell? – something that depends on their expression, sub-cellular localization and copy number. Many of them can also be biophysically real and non-physiological interactions that might be occurring in not-yet-observed *in vivo* conditions. Attempts are now being made to address these questions by using co-expression and localization data to validate interactions (Zhang et al. 2010; Rüegg et al. 2008).

In co-complex experiments, the main sources of error include the inability of the method to distinguish between direct and indirect interaction – a problem which can be resolved by integration with pair-wise data from Y2H experiments (Gentleman & Huber 2007). Co-complex methods also have the inherent limitation of possible interference of the TAP tag in the folding, function, localisation and interactions of bait proteins (Blow 2009), hence contributing to false negatives. Datasets from co-complex/AP-MS experiments are always only partially over-lapping, even when context and conditions are carefully controlled (Tabb et al. 2010; Paulovich et al. 2009). This gap in reproducibility is attributed, in part, to the capacity of the LC-MS analytical system and the number of peptides present in the digest. One way to improve the coverage and reproducibility in such experiments is by increasing the fraction of precursors selected for CID during tandem mass spectrometry (Cox & Mann 2011).

A recent study described the errors found in interaction data by classifying it into two types. Stochastic errors occur at random and are variable in nature. They can easily be controlled by replication. Systematic errors are recurrent and consistent sources of bias. They are hard to diagnose and can only be addressed by improving experimental procedures and data analysis methods (Gentleman & Huber 2007).

1.1.4 Databases of PPI datasets

Multiple databases are available online that store and curate protein-protein interaction data from different organisms. Widely used primary databases that contain raw datasets include the Database for Interacting Proteins (DIP)

(Salwinski et al. 2004), the IntAct database (Aranda et al. 2010), the Molecular Interaction Database (MINT) (Ceol et al. 2010), the Bio-molecular Interaction Network Databases (BIND) (Bader 2003) , the Mammalian Protein-protein Interaction Database (MIPS) (Mewes et al. 2010) and the Biological General Repository for Interaction Datasets (BioGRID) (Stark et al. 2010). Primary databases very often contain redundant interactions, which are stored in different formats with varying protein identifiers. Mapping these identifiers alone, a major challenge in the field, consumes almost 70% of curation time according to some estimates (Orchard et al. 2007). This has led to the proposal of MIMIx, a standardized guideline for scientists to report interactions making their storage and retrieval more rapid and systematic (Orchard et al. 2007).

Another attempt to tackle these problems with primary databases was the emergence of 'meta-databases' that attempt to unify existing primary databases removing redundant interactions, minimizing the loss of molecules and interactions during manual mapping and thus making it faster and easier for the interactomics community to retrieve datasets. Two good examples are PINA (Wu et al. 2009) and APID (Prieto & De Las Rivas 2006; Hernandez-Toro et al. 2007).

There are many programmes for the visualization of PPIs, two of the most popular are Osprey (Breitkreutz et al. 2003) and Cytoscape (Smoot et al. 2010), with the latter gaining more prominence. Other programmes are also emerging like the 3D network visualization tool called BioLayout Express3D (Theocharidis et al. 2009). An excellent review has been provided on visualization of 'omics' data by Gehlenborg *et al* (Gehlenborg et al. 2010).

1.1.5 Properties of PINs

The analysis of biological networks relies heavily on our understanding of general networks through network theory (Albert & Barabassi 2002; Strogatz 2001). These insights from mathematics have given us an understanding that biological networks have core organisational principles and are not randomly connected groups of nodes.

The earliest networks that were studied in mathematics had random 'wiring', and were called random networks. Erdos and Renyi first studied these networks in 1960 (Erdos & Renyi 1960). Nodes in random networks roughly have the same degree, i.e. they are connected to more or less the same number of other nodes. In contrast, real networks like biological networks had some nodes (proteins), which interact with a greater number of other nodes compared to an average one in the network. This meant that the random network model could not explain some features of biological networks, and hence could not be useful in modelling them.

Biological networks are highly non-uniform networks having a few proteins/nodes that attach to many other proteins/nodes with the rest having a low number of interactions, and thus different network models were required to study them.

Hubs

The small number of proteins that have a very large number of connections are called hubs (Barabási & Oltvai 2004). Hubs are a good example of how network topology can be translated back to meaningful biology. Several studies in model

organisms suggest that hubs correspond to essential genes (Jeong et al. 2001), which upon deletion cause lethality in different cells. Hubs have also been reported to produce a larger number of mutant phenotypes upon deletion when compared to non-hub proteins (H. Yu et al. 2008).

These observations led people to test the assumption that disease-related genes might be hubs. Wachi *et al* showed that unregulated genes in lung cancer cells tended to have above average number of interactions (Wachi et al. 2005). Similarly in another study, proteins in cancer related cells were reported to have twice the number of partners compared to those in non-cancerous cells (Jonsson & Bates 2006). However not all essential hub-encoding genes are disease related, since their deletion would lead to the lethality of embryos in their early developmental stages, and would not allow the disease to be manifested. Goh *et al*, in their analysis of human data, summarise that essential genes were more likely to encode hub proteins. Disease genes on the other hand tend to encode non-hubs, supporting the fact that most disease genes in humans encode non-essential proteins (Feldman et al. 2008; Goh et al. 2007).

Although some studies have debated the role of hubs (Agarwal et al. 2010), pointing out the possibility that some of them might be products of investigative bias, they are still expected to have an impact on more proteins by virtue of the number of interaction they have, compared to non-hub proteins. Hubs have been further classified based on their spatial and temporal location in the network into those located within co-expressing or functionally related clusters (party hubs) or those that connect such clusters (date hubs) (Han et al. 2004).

Robustness

The degree of a protein/node is the number of interactions it has (i.e. a hub protein is a high degree protein). Figure 1.4 presents the degree distribution of a PPI network, which shows that a small number of proteins have high degrees and a high number of proteins (in the long tail) have a lower degree. This is typical of all PPI networks. The degree distribution of biological networks gives them another interesting property – robustness, since random deletions of nodes have a higher chance of affecting the predominant low-degree proteins (Barabási & Oltvai 2004). Hubs on the other hands have a low likelihood of being deleted at random. This explains the incredible integrity of biological systems and their robustness towards different intracellular and extracellular perturbations. On the other hand, deletion of hubs should it occur, would tend to disintegrate the network into smaller disconnected sub-graphs, severely damaging cell function. The presences of hubs in biological networks also make them susceptible to 'attack vulnerability' (Albert et al. 2000). Pathogens capitalize on this property and specifically attack hubs in cellular networks. Several viral proteins like the Influenza H5N1 (Shapira et al. 2009) and Epstein-Barr virus (Calderwood et al. 2007) have been reported to have preferentially targeted host hubs.

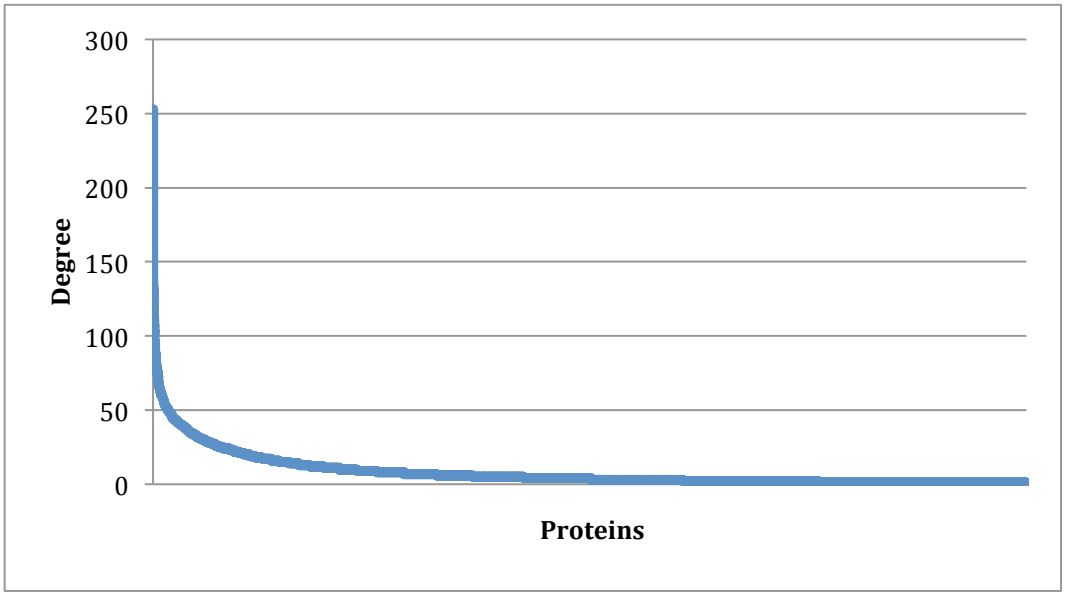


Figure 1.4: The degree distribution of the *Drosophila melanogaster* interactome from DIP and BioGRID combined (proteins on x-axis ordered by degree). Its long tail representing nodes with few interactors characterizes the graph. Few proteins make the 'hubs' or highly connected proteins.

Network modules

Biological systems can often be disassembled into modular components that are arranged in ordered hierarchies for example, co-expressing genes in genetics, homologous sequences in evolutionary biology and clustered nodes working towards a common and distinct function in network biology (Hartwell et al. 1999; Ravasz & Barabási 2003; Alon 2007).

The underlying assumption of modules in PINs is if a protein plays a role in a certain biological function, its immediate interacting neighbours can be expected to be involved in a similar function (Hartwell et al. 1999). Several studies have shown that genes involved in diseases with similar phenotypes have a greater tendency to interact with each other (Xu & Li 2006; Gandhi et al. 2006). This

supports the assumption that if a gene is involved in a particular biological function or whose disruption has a role in a disease phenotype, more relevant genes can be found in its vicinity in an interaction network. These groups or clusters of proteins/nodes are called modules.

A module in pure network terms is a highly clustered group of nodes. Such *topological* modules in networks can be identified using different clustering algorithms (Enright et al. 2002; Ravasz et al. 2002). Two other types of modules can be described with reference to biological networks. One is the *functional* module, which is a highly connected cluster of nodes that share a similar or relevant biological function within the cell and the other class of modules, known as *disease* modules, are highly interacting groups of proteins that contribute together to disease phenotypes when disrupted (Navlakha & Kingsford 2010).

As one would expect these different types of modules have considerable overlaps, for example a functional module in a network is likely to overlap with a topological module due to its higher connectivity, and also a disease module, if some of the proteins in it contribute to disease phenotypes. It should be noted however that disease modules can be different and are known to overlap with other disease modules as well (Barabási et al. 2011).

Modules can be broken down to smaller components or sub-graphs called motifs. Motifs can sometimes be assigned to particular functions, the study of which is common in synthetic biology (Alon 2007; Shoval & Alon 2010).

Network Summary Statistics

Several network statistic measures have been found useful to study and compare the characteristics of these large, unconnected and relatively sparse networks in biology (Costa et al. 2008). Some of the most popular summary statistics are briefly touched upon here.

Degrees and Degree distribution

As described earlier, degree refers to the number of direct interactors of a node in a network. For large networks the average degree is used as a common summary statistic, which is the average of the degrees of all nodes in the network.

A more useful statistic is the degree distribution represented by a graph. This can help in describing the type of the network for example in biological networks, the graph is characterised by a long tail representing the majority of nodes with a few direct interactors (Figure 1.4).

Clustering coefficient

The clustering coefficient is a measure of the connectivity between nodes within a network. This is essentially the number of triangles that pass through a node. The value of clustering coefficients range from 0 to 1, where 1 would imply that all nodes in the network are part of at least one triangle. Clustering coefficient is calculated by the following formula:

$$C_i = \frac{2\,n_i}{k_i\,(k_i - 1)}$$

where n_i denotes the number of links connecting the k_i neighbours of node *i.*

Biological networks have higher clustering coefficients compared to random networks. This statistic is one criterion that can be used to identify topological modules in networks. Nodes within a module tend to have higher clustering coefficients compared to others in the network.

Shortest Path Length and Diameter

The average or characteristic shortest path length (SPL) and diameter of a network give us an indication of the size and density of the network. The average shortest path length is the average of the shortest path lengths between all pairs of nodes in the network. A tightly packed network will have a smaller value for this statistic compared to a less dense one.

The network diameter is the longest path length between any pair of node on the network. This statistic helps in comparing the relative spread of networks.

Biological networks are usually large, but they still have relatively low average SPL values (and high clustering coefficients) compared to random networks, which implies that although the protein pairs might not be adjacent to one another, they are still connected through a short path – a property referred to as the small world phenomena Jeong et al. 2000).

Node Attributes

Networks can be analysed further by adding attributes to both the nodes and edges by integrating different datasets on top of the network. For nodes, these attributes can be derived from phylogenetic data like protein age and conservation across species (Weiss et al. 2010), functional data like gene

ontologies and RNAi phenotypes (Maere et al. 2005; Mohr et al. 2010a) and structural data like protein domain features (Marchler-Bauer et al. 2010). Integrating such data, which are increasingly being available as large-scale 'omics' datasets (described in detail in Chapter 2), not only helps in explaining the characteristics of nodes and their interactions in the network but also enhances our capacity to analyse the local neighbourhood of nodes in the network through prediction. One of the most-widely used node attribute in network analysis are Gene Ontology annotations, which are controlled descriptions (ontologies) that describe gene products based on their biological process, molecular process and cellular localization (The Gene Ontology Consortium 2000; Maere et al. 2005).

Predicting protein function

Function prediction is one of the fundamental questions faced in biology, and PPIs offer one avenue for attacking this problem (Barabási et al. 2011). Several methodologies have been reported that use information from interaction networks at different scales. For example, linkage analysis looks at proteins encoded by genes from the same locus. One study has reported a tenfold enrichment of disease-causing genes in such modules (Oti et al. 2006).

Other strategies include module based methods, which assume that member of the same module have a higher likelihood of being involved in the same biological function (Navlakha & Kingsford 2010; Lage et al. 2007). These can be modules enriched in a particular gene ontology, co-expressed genes or genes that carry disease-related mutations. Several important modules have been

established in studies that looked at diseases like breast cancer, obesity and diabetes (Taylor et al. 2009; Dobrin et al. 2009; Liu et al. 2007).

A broader approach is used by distance and diffusion based methods. Distance based methods include random walk methods which prioritize genes based on the shortest path length between candidate genes and known disease-related ones. Candidate genes are scored on the basis of distance (proximity) to known disease-related nodes on a particular disease network (Voevodski et al. 2009; Bebek et al. 2012a).

Diffusion-based or information flow approaches, take into account the multiplicity of paths between nodes. Instead of a single path, these methods compute the fraction of information flow across every path between a candidate and a known disease gene. A common analogy would be the flow of electric current across different resistors in an electric circuit. Different network flow methods have been developed to date, from simple neighbour counting to more complex network diffusion methods (Wang et al. 2010).

Linkage methods use pairwise linkage information only when prioritizing nodes while module-based methods extract signals by looking at the network neighbourhood. Diffusion and distance related methods analyse the topology of the entire network and the location of every known gene. Such full network-based methods are expected to have higher predictive power by virtue of the greater information they use, both topological and functional (Barabási et al. 2011). This assumption is supported by studies like a recent comparative

analysis which indicated higher predictive performance by diffusion based studies and lower predictive power for linkage analysis (Barabási et al. 2011).

With a growing repertoire of analytical methods, increasing predictive power and the popularity of networks as good 'scaffolds' for integrating other types of datasets, network-based methods hold great promise for the broader field of systems biology (Joyce & Palsson 2006).

1.1.6 Applications of PINs

PINs have not only helped biologists to develop a global systems-level understanding of cellular events, but has also aided in the study of complex diseases. With our understanding of networks, evidence is accumulating in support of the fact that disease can no longer be treated as the dysfunction of simple linear pathways, but as the consequences of malfunction in a complex web of non-linear pathways (Liu & Chen 2012). These methods are not only helping us understand the cross talk between these complex disease pathways but are also helping us in identifying the best targets for intervention (Sanz-Pamplona et al. 2012). Some important applications of PINs in different areas of biology and medicine are as follows.

- Protein interaction data has been used to predict interactions in different species (Yu et al. 2004; Matthews et al. 2001).
- PINs have also been used to predict protein function (Bebek et al. 2012b) and disease-related genes (Barabási et al. 2011).
- PINs can help us in studying the evolutionary origins of cellular networks by knocking down or over-expressing proteins in cells. Similarly, the

addition of new edges and its effects can also be studied using such networks (Isalan et al. 2008).

- The top-down study of protein interaction networks in Systems Biology has contributed to the bottom-up design of biological networks in Synthetic Biology by sharing tools and methodologies to study the properties and underlying design principles of networks within a cell (Lanza et al. 2012). This understanding is already translating into industrial applications like the production of drugs (Ro et al. 2006), chemicals (Kodumal et al. 2004; Steen et al. 2010) and biofuel (Atsumi et al. 2008; Inui et al. 2008; Steen et al. 2010).
- Network approaches are used to unravel the underlying pathways of complex diseases like cancer especially with the recent discovery of intra-tumour heterogeneity within individual patients (Roukos 2012), neurodegenerative disorders (Ghosh & Basu 2012), and even host-pathogen interactomes in infectious diseases like malaria (Winzeler 2006).
- In drug discovery research, network based studies have helped in identifying more rational points of intervention by understanding secondary targets and downstream affects (Papageorgiou & Wikman 2004).
- One emerging area, called *edgetics*, targets interactions (edges) instead of proteins (nodes) resulting in better drugs with more subtle affects and minimal side effects. (Vidal et al. 2011). This is being harnessed in anticancer drugs research involving novel p53 interactions by Roche

(Vassilev et al. 2004) and Bcl-2 family proteins by Abbott Labs (Oltersdorf et al. 2005).

- In vaccine development network biology has helped in identifying better predictors of immunogenicity through better understanding of the mechanisms and networks behind immune responses (Trautmann & Sekaly 2011).

1.2 Microtubules and MAPs

1.2.1 Microtubules

Microtubules (MTs) are highly dynamic, 25-nm wide hollow polymeric tubes that make the bulk of the cytoskeleton. These cylindrical polymers are made of protein monomers called tubulin. MTs are polar in nature, with a plus and a minus end. The proximal minus end is usually anchored in an insoluble proteinaceous matrix known as the MT organising centre (MTOC) from where the microtubule nucleates. The plus end is the distal end, which continuously grows and shrinks, probing the cytoplasmic space.

During the normal resting phase of a cell or interphase, microtubules provide structure and stability to the cell as part of the cytoskeleton. It also provides tracks for molecular motors to conduct cellular transport and helps in the positioning of several cellular components like the endoplasmic reticulum and the Golgi complex. When the cell approaches the dividing phase or mitosis (also called M-phase, described in more detail in the following section), the

microtubule network disassembles and reassembles into the mitotic spindle, the main cellular machinery, which drives cell division (Figure 1.5).

The tubulin monomer

The tubulin protein is the structural unit of microtubules. Tubulin is a family of globular proteins with isoforms of two related species, i.e. α-tubulin and β-tubulin (Figure 1.6), which are highly conserved proteins and found in almost all eukaryotes (Wade 2009). The two types of tubulin proteins exist as hetero-dimers (1 α-tubulin and 1 β-tubulin). Heterodimers polymerize end-to-end in a proto-filament, which form the building blocks of microtubules (Figure 1.7).

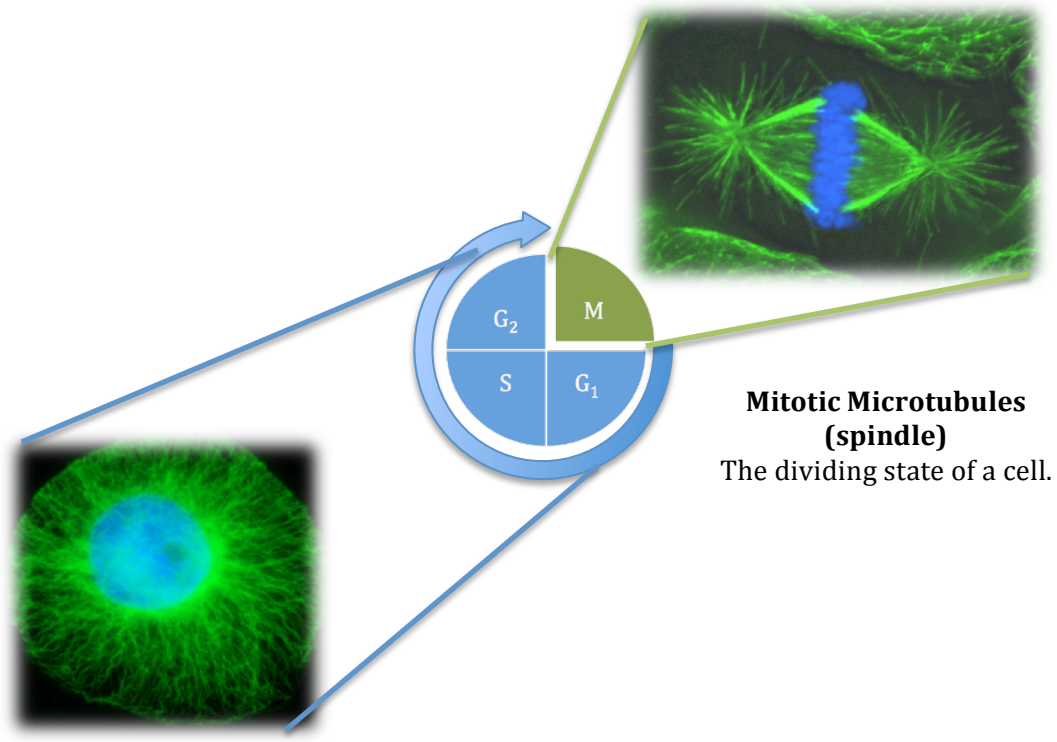


Figure 1.5: The microtubule network in the cell during interphase (blue) when they are part of the cytoskeleton and the mitotic or M-phase (green) where they make up the spindle. G_1, S and G_2 represent the three other phase of the cell cycle.

The tubulin proteins have approximately 50% sequence identity, and a strikingly similar structure under the electron microscope. Both have two globular domains separated by a helical domain. The larger globular domain also has a binding site for GTP at the plus-end (Figure 1.6). Under normal conditions, 13 proto-filaments align side by side and create a hollow tube (Figure 1.7b), which is known as the microtubule. Microtubules can be composed of about 9 to 17 proto-filaments, depending on the cell type and conditions (Wade 2009).

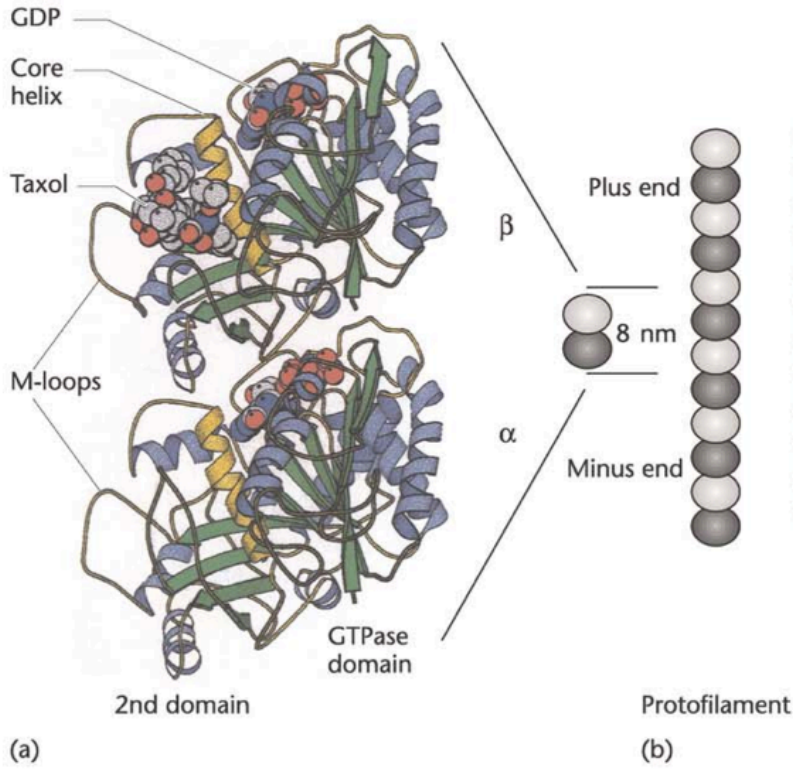


Figure 1.6: (a) Atomic structure of the αβ-tubulin heterodimer shown as a ribbon diagram; α-helices in blue, β-sheets in green. Taxol and GDP (bound to β-tubulin) and GTP (bound to a-tubulin in a position equivalent to that of GDP in β-tubulin) are shown as space-filling atoms. (b) Assembly of heterodimers into longitudinal proto-filaments. Diagram taken from (Amos 2005).

Another important type of tubulin is γ-tubulin, which has been shown to contribute to the nucleation of microtubules. γ-tubulins along with GCP/Grip proteins create ring-shape complexes in microtubule-organising centres like centrosomes where they play a role as templates by stabilizing the minus ends of proto-filaments (Wiese & Zheng 2006).

Microtubule Assembly and Disassembly

Microtubule assembly is a GTP-dependant process. Two heterodimer building blocks are added together when the GTP of the β-tubulin is hydrolysed into GDP (Figure 1.7c). Lateral interactions occur between proto-filaments until they complete a hollow tube. These lateral interactions that create a tubulin lattice are known to occur between M-loops within the tubulin structure (Figure 1.6). Microtubule assembly occurs when heterodimers are added at the ends of a proto-filament. MT assembly and disassembly events occur at both ends (Figure 1.7).

MT disassembly in the cell can occur stochastically, via destabilising proteins like Katanins and Spastins (W. Yu et al. 2008) or by the removal of MT stabilising cap proteins like proteins from the EB and CLIP families (Akhmanova & Steinmetz 2008). This happens when the ends of microtubules splay apart and bend outwards in curved formations, shedding spirals and rings of proto-filaments. Spontaneous conformational changes are believed to propagate across the length of proto-filaments and cause these stochastic events of disassembly. The presence of nucleotides (GTP, GDP) does not appear to affect the rate of disassembly, even in *in vitro* experiments.

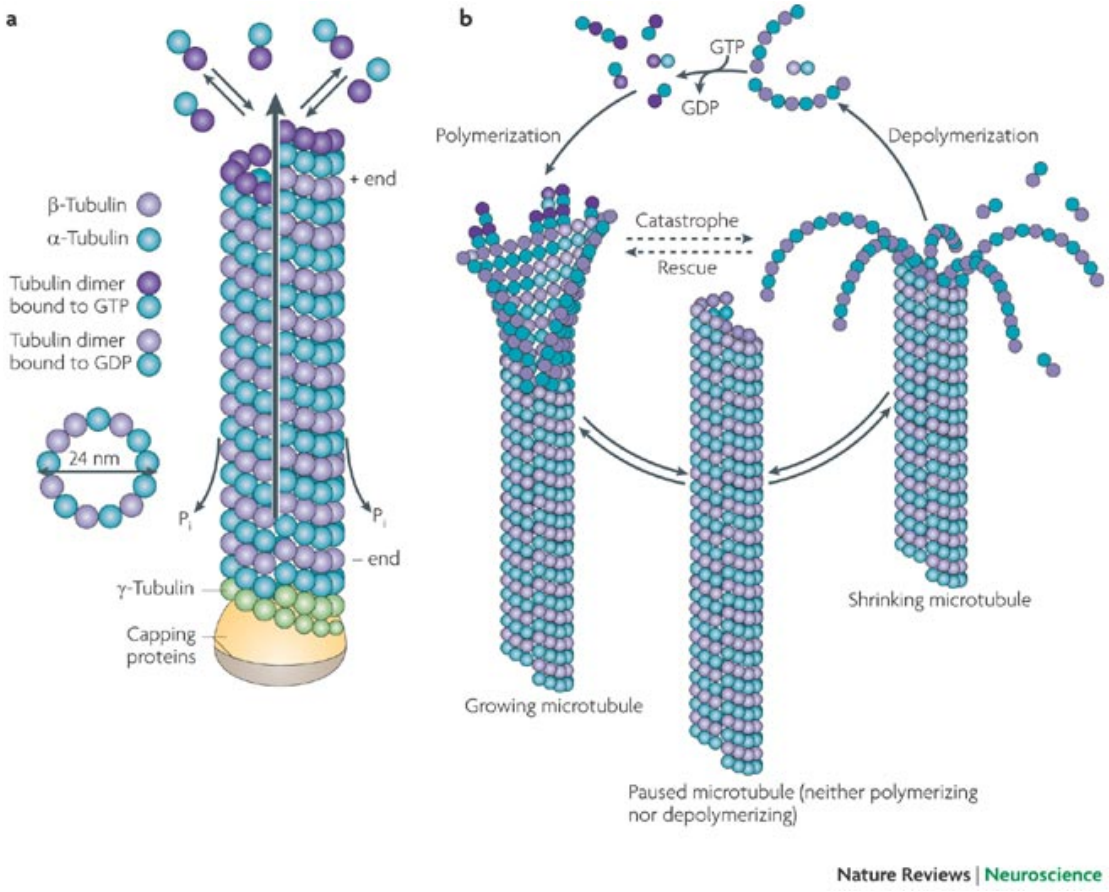


Figure 1.7: Dynamics of a microtubule. (a) The tubulin heterodimer, proto-filament, and the arrangement of 13 proto-filaments giving the standard 25nm hollow tubule with a plus and a minus end. (b) The polymerisation and depolymerisation of tubulin hetero-dimers and proto-filaments, causing the rescue and catastrophe of microtubules. Diagram taken from (Conde & Caceres 2009)

Effect of temperature and drugs

Temperature is an important physical condition for the assembly and disassembly of microtubules in the presence of free tubulin and GTP. Temperatures lower (4^{o}C) than the biological optimum promote the disassembly while higher temperatures (37^{o}C) bring about the re-assembly of MTs (Shelanski et al. 1973).

Drugs are also known to affect the assembly of microtubules. Two important ones are Colchicine derived from the plant, *Colchicum autumnale* (Wilson & Meza 1973) and Paclitaxel or Taxol® isolated from endophytic fungi found in the bark of the yew tree, *Taxus brevifolia* (Amos & Löwe 1999). Colchicine is known to inhibit the assembly of MTs by binding and sequestering unbound tubulin monomers. Taxol inhibits the disassembly of microtubules by binding and stabilising the hydrolysed β-tubulin in protofilaments. Both of these compounds have not only been used extensively in the molecular study of microtubules, but also as potential anti-mitotic drugs in diseases like cancer (Jackson et al. 2007).

Properties of Microtubules

Cytoskeletal proteins like microtubules and actin filaments and are structurally polar molecules and have two unique properties that play a key role in their cellular function.

As mentioned, microtubules are highly dynamic with assembly and disassembly of heterodimers happening at both ends. The plus end is more dynamic, with rapid rounds of assembly, disassembly and reassembly occurring stochastically. This property is referred to as dynamic instability (Wade 2009). The change from growth to shrinkage is termed *catastrophe* while the change from shrinkage to growth is called *rescue* (Figure 1.7). One contributing reason can be the relatively unstable state of polymerised tubulin heterodimers, which undergo faster GTP hydrolysis that results in conformational changes and weaker lateral interactions between protofilaments in the lattice structure.

This property enables microtubules to play an important cytoskeletal function by reaching out to distant regions in the cell during interphase, helping out in the locomotion of motile cells, and perhaps more crucially, in the 'search and capture' of kinetochores during mitosis.

The second property is treadmilling, which refers to the continuous loss of GDP-bound free tubulin units from the minus end that is replaced by the addition of GTP-bound tubulin units at the more dynamic plus-end of the microtubule. During 'steady-state' treadmilling the net length of the microtubule remains constant, as the rate of shrinkage at the minus end is equal to the rate of growth units at the plus-end (Wade 2009).

1.2.2 Microtubule-associated proteins (MAPs)

The dynamic behaviour of microtubules is regulated by their interaction with microtubule-associated proteins (MAPs). MAPs are defined as proteins that interact with MTs and have a structural, dynamic or regulatory role. Table 1.1 gives a summary of different types of MAPs with representative fly proteins from each family along with their human homologues. Since microtubules make different cellular apparatus, like the cytoskeleton, mitotic spindle, cilia and flagella, MAPs unsurprisingly have several direct and indirect roles in these processes. The regulation of these important proteins is commonly brought about by phosphorylation, which usually inhibits their activity.

The first microtubule-associated proteins (classical MAPs) studied were found in neuronal cells where they help in stabilizing the microtubules across the lengths of these long cells. These are structural MAPs that were often co-purified with

MTs, e.g. the tau (in axons), MAP1, MAP2 (in dendrites) and MAP4 proteins (Halpain & Dehmelt 2006; Dehmelt & Halpain 2004). Their presence significantly reduces the critical concentration of free tubulin required for assembly. These proteins are composed of a MT-binding repeat domain on the C-terminus and a variable domain at the N-terminus which protrudes out of the microtubule.

Other MAPs help in microtubule nucleation such as the multi-subunit γ-tubulin ring complex or γ-TuRC (made of γ-tubulin and the GCP proteins) (Wiese & Zheng 2006; Kollman et al. 2011), which is recruited into the pericentriolar matrix of centrosomes. Other proteins like Cnn, Cep192/Spd2 and the Augmin complex regulate the recruitment of these complexes (Goshima et al. 2008; Carvalho-Santos et al. 2010; Dobbelaere et al. 2008).

Microtubule motor proteins are another class of nucleotide-dependent MAPs that function as molecular motors using microtubules as tracks and conducting transport-related functions. The two well-studied classes are the plus-end directed Kinesin proteins and the minus-end directed Dyneins (Caviston & Holzbaur 2006). Another closely related class is the Kinesin-related proteins or KLPs. Most of these proteins have a motor domain with ATPase activity at the N'-terminus and a variable cargo binding tail domain at the C'-terminus.

Several MAPs are required for a process called microtubule bundling, which stacks microtubules in ordered anti-parallel arrays in the mitotic spindle. Common examples of these are the kinesin-5 and MAP65 family members

MAP type	Protein Family	*Drosophila* proteins	Human proteins
Classical/Structural MAPs	MAP1	Futsch	MAP1
	Tau/MAP2	Tau	MAP2/MAP4
MT nucleation	y-TuRC		
	Cnn	Cnn	
	Cep192	D-Spd2	Cep192
	Augmin	Dgt2-6	hDgt6
MT motors	Kinesins		MCAK, CENPs, etc
	Dynein	Dynein	Dynein
	KLPs	Klp38B, Klp59C, etc.	
MT bundling proteins	Kinesin-5	Klp61F	Eg5
	MAP65	Feo	PRC1
MT stabilizing proteins	CLASPs	Orbit/MAST	CLASP1/CLASP2
	Spektraplakins	Shot	MACF1/MACF2
Spindle stabilizing proteins	TPX2		TPX2
	NuMA	Mud	NUMA
	HURP		DLGAP5
	TACC	D-TACC	TACC1,TACC2,TACC3
MT polymerizing proteins	EB	EB1	EB1, EB2, EB3
	XMAP215	Msps	chTOG
	CLIPs	CLIP-190	CLIP-170
MT disassembly proteins	Stathmins	Stathmin	Op18
MT severing proteins	Katanins	D-Kat60	Katanin p60/p80
	Spastins	D-Spastin	SPG4

Table 1.1: Different types of microtubule associated-proteins, along with representative protein families and their members in *Drosophila* and Humans.

(Walczak & Shaw 2010). Other MAPs like the CLASP family (Galjart 2005) and the spectraplakins (Suozzi et al. 2012) stabilize microtubules by attaching them to molecules like actin and other cortical adhesion sites.

Since fidelity is important during mitosis, several additional proteins help in stabilizing the mitotic spindle like TPX2, NuMA, HURP and TACC proteins (Peset & Vernos 2008; Kwon & Scholey 2004). Other MAPs help in polymerizing microtubules, by binding plus-ends like the EB family (Akhmanova & Steinmetz 2008), by adding heterodimers to plus ends like Msps or by promoting rescue like CLIP-190 (Dzhindzhev et al. 2005).

Microtubule disassembly on the other hand occurs if the cytosolic pool of heterodimers is sequestered by proteins like Stathmin (Belmont & Mitchison 1996) turning the equilibrium towards catastrophe, by ATPase-dependent catalytic disassembly carried out by Kinesins, or by Microtubule severing proteins like Katanins and Spastins (W. Yu et al. 2008).

1.3 Mitosis

In order to reproduce, cells go through a highly regulated cycle of duplication and division, known as the cell cycle. There are two important steps in the cell cycle, i.e. the duplication of chromosomes (termed the S-phase of the cell cycle) and their equal segregation into two daughter cells (termed M-phase). Both these steps are separated by two gap phases (G_1 and G_2). The gaps not only allow for growth in between the S and M phases but also provide time for the cell to monitor each process and impose checkpoints in case of aberrant behaviour. Most eukaryotes generally have all four steps in the cell cycle.

DNA replication is the central step in the S-phase. It is highly regulated to ensure the accurate replication of each base pair in the DNA and also to allow for

replication to strictly happen only once. The duplication of chromosomal packaging proteins (histones) is also carried out during this phase. At the end of S-phase the nucleus has two identical sister chromatids held together by giant ring-shaped protein structures called Cohesin (Xiong & Gerton 2010; Nasmyth 2005).

Mitosis, during M-phase, is the process by which a eukaryotic cell faithfully divides its DNA content into two daughter nuclei. This is generally followed by the splitting of the cytoplasm (and its contents) giving rise to two equal daughter cells in a process called cytokinesis. Both mitosis and cytokinesis together constitute the mitotic (M) phase of the eukaryotic cell cycle (Figure 1.8). Mitosis is dependent on a microtubule-based machinery called the spindle, whereas cytokinesis is dependent both on the microtubule cytoskeleton and the actomyosin structure called the contractile ring (Rojas et al. 2012; Goshima et al. 2007a).

The process of mitosis, from entry until exit, is regulated mainly by two types of post-translational modifications, i.e. phosphorylation for transient regulation that is reversible in nature, and protein degradation for irreversible unidirectional progression. Phosphorylation is carried out by a variety of protein complexes composed of different Cyclin-dependant kinases (CDKs) and their positive regulators called Cyclins (Coudreuse & Nurse 2010; Tyson & Novak 2011; Guest et al. 2011a).

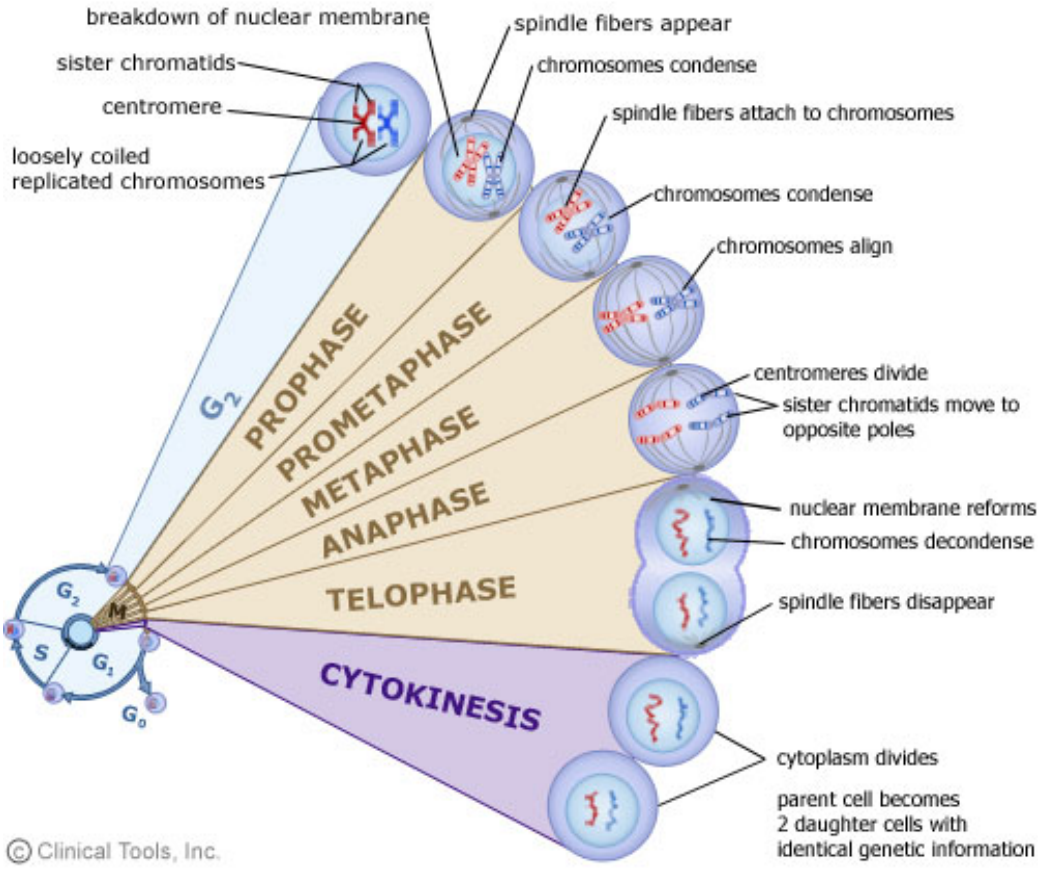


Figure 1.8: Stages of the Cell Cycle, i.e. G_1, S, G_2 and M. The M-phase or mitosis is shown in greater details with all its phases and key events happening in each phase. Image taken from (VGEC 2013).

1.3.1 G_2/M – Entry into mitosis

Before the onset of mitosis, Cyclin B levels increase, which in turn increases the levels of the Cdk1-CyclinB complex, also known as the M-phase promoting factor. The assembly of the Cdk1-CyclinB complex itself regulated by DNA replication and DNA damage checkpoints, which identify errors in DNA replication or mutations in the replicated DNA, caused by chemicals or radiations. This is carried out by two related protein kinases ATR and ATM, which bind to the site of damage and trigger a cascade of phosphorylation through the Chk1 and Chk2

proteins. These proteins phosphorylate other proteins including the p53 protein that stimulates Cdk kinase inhibitors to block the G_1/S and G_2/M transitions and promote cell cycle arrest (Samuel et al. 2002).

The CDK1-CyclinB complex in coordination with Polo-like kinases and Aurora kinases triggers the onset of the mitosis by phosphorylating several structural and functional proteins.

1.3.2 Prophase and centrosomes

The two sister chromatids in the nucleus during prophase are in the form of a thread-like network known as chromatin, which begins condensation. The phosphorylation of Condensin proteins by the CDK1-CyclinB complex stimulates the coiling of sister chromatids into compact rod-shaped chromosomes (Figure 1.8).

The centrosomes, which are the MT-organising centres in an animal cell (referred to as spindle pole bodies in fungi), are composed of two centrioles surrounded by an amorphous pericentriolar matrix or PCM. The PCM recruits γ-TuRC complexes, which are the templates for nucleating individual MTs (Wiese & Zheng 2006; Bettencourt-Dias & Glover 2007).

Prior to entry into mitosis, the centrosome cycle duplicates the centrosome in time for mitosis in M-phase. This cycle is triggered at the onset of G_1 phase though the Cdk2-CyclinE complex. The segregation of the mother centrioles and the nucleation of daughter centrioles occur during S-phase, while centriole elongation carries on till the end of G_2-phase.

During prophase the centrosomes start separating and start migrating towards opposite poles of the cell. This migration is partially carried out by motor proteins, like Kinesin-5 (activated though phosphorylation by CDK1-CyclinB and Aurora A kinase) and Dynein proteins (activated by Polo-like kinases and Aurora A).

Following separation, more γ-TuRC is recruited into the PCM, a process termed centrosome maturation. This promotes the nucleation of microtubules in preparation for the assembly of the mitotic spindle (Bettencourt-Dias & Glover 2007).

1.3.3 Pro-metaphase and the NEB

In most eukaryotes pro-metaphase starts with the abrupt breakdown of the nuclear envelope (termed nuclear envelope breakdown or NEB). This is triggered by the CDK1-CyclinB complex, which phosphorylates subunits of the nuclear pore complex catalysing the degradation of the nuclear membrane into smaller vesicles, and nuclear lamins which causes the disassembly of the nuclear lamina (Gerlich et al. 2002) (Figure 1.8). There are certain exceptions to this, for example fungal cells, in which mitosis occurs within an intact nucleus, and hence called 'closed mitosis' (De Souza & Osmani 2007).

The increase in γ-TuRC accumulation at centrosomes triggers microtubules nucleation, which radiate all over the cell from the two opposite poles. As they emanate, some microtubules from both sides reach out to the cell cortex (astral MTs) while others meet each other and overlap forming what will develop into the central spindle (polar MTs). A third class of microtubules is captured by

kinetochores – large multi-layered protein complexes located at the centromeric region on condensed chromosomes.

1.3.4 Metaphase and the SAC

Towards metaphase, all kinetochores on chromosomes are captured by microtubules from the two poles using a 'search and capture' mechanism.

The kinetochores are located back-to-back on opposite sides of the centromeric region to promote *biorientation* or the attachment and pulling apart of chromosomes by microtubules from both poles. This produces a state of physical tension at the kinetochores. This step is crucial as it ensures the fidelity of chromosome segregation and is regulated by the spindle-assembly checkpoint (SAC), which monitors the strength of MT attachment through SAC proteins including the mitotic checkpoint complex (MCC) composed of Mad2, BUBR1 and BUB3 (Musacchio & Salmon 2007).

Unattached kinetochores or incorrect attachment (like the attachment of both kinetochores to MTs from one pole) is detected by the lack or lower level of tension. Such interactions are unstable and are inhibited by proteins like the kinetochore-associated protein kinase Aurora B. The kinetochore captures the right microtubule through a 'trial and error' mechanism until the right level of tension is produced. This promotes the attachment of more kinetochore MTs producing thick *k-fibres*. The chromosomes after being tugged back and forth align at the central equatorial position called the metaphase plate (Figure 1.8).

1.3.5 Anaphase and the APC

Once the conditions for SAC are fulfilled, the anaphase-promoting complex or APC/C is released which triggers the start of anaphase with the sudden disruption of cohesion proteins that hold the sister chromatids together. This is carried out by the ubiquitination and subsequent degradation of Securin, which releases a protease called Separase that degrades the Cohesin ring structures. The release of sister chromatids coupled with microtubule shortening and the activity of minus-end directed microtubule motor proteins take the two separated chromatids to opposite poles (a sub-stage called anaphase A). The pole-ward travel is further facilitated by the growth of polar MTs, which push the two poles apart (anaphase B). The APC/C also degrades the Cyclins responsible for the G_1/S and G_2/M transitions marking the start of mitotic exit. This is carried out by the ubiquitin ligase activity of this protein complex, which triggers a proteolytic cascade via proteasomes (Peters 2006).

1.3.6 Telophase

Chromosome segregation completes towards late anaphase. The spindles starts shrinking and depolymerising while the nuclear envelope around each set of daughter chromatids, start reassembling in the last stage called telophase. Nuclear pore complex subunits are incorporated as chunks of nuclear envelope membrane coalesce and the nuclear lamina also reforms. Chromosomes decondensation begins, unwinding them into the interphase chromatin state that allows for the resumption of gene expression (Figure 1.8). These steps are the

reverse of those occurring in prophase and are partially triggered by the dephosphorylation of the same Cdk and Cyclin proteins (Yeong et al. 2002).

1.3.7 Cytokinesis

After the completion of mitosis and the division of the DNA contents of the cell, the splitting of cytoplasm, or cytokinesis, follows. This is carried out by a large dynamic protein structure made up of actin, myosin and other regulatory proteins called the contractile ring (Ebrahimi & Gregory 2011). This large ring structure is begins a rapid assembly at the start of anaphase and completes right after telophase. This is when the contractile ring creates a furrow by contraction, which begins to cleave the cytoplasm. The position of furrow generation (Cleavage plane positioning) is regulated by signals from the Anaphase spindle through astral microtubules and the central spindle through the protein RhoA (White & Glotzer 2012). When contraction is complete membrane insertion and fusion fills in the gap and completes cytokinesis.

Important deviations from this generalised process of mitosis, where the five mitotic steps are not always followed by cytokinesis, include meiosis and early embryonic stages in certain animals like *Drosophila*, in which cytokinesis is inhibited in the early blastoderm embryo during the first 13 rounds of S-M cycles and all nuclei lie in the same cytoplasm called syncytium. Another special case is the process of asymmetric cell division, which occurs in specialized cells like neuroblasts and stem cells (Wu et al. 2008).

In animal cells, another area of interest is the assembly of bipolar spindles in the absence of centrosomes. Several non-classical mechanisms of spindle assembly

in cell division have been highlighted by recent studies including the Ran-dependent chromatin-directed MT generation (Heald et al. 1996), Kinetochore-driven MT generation (Rieder 2008), CPC-directed MT generation (Colombié et al. 2008) and a MT-dependant microtubule nucleation pathway (Uehara et al. 2009).

1.4 *Drosophila melanogaster* – the fruit fly

Since the rediscovery of Mendel's laws of genetics by Thomas Hunt Morgan (1866-1945), the fruit fly (*Drosophila melanogaster*) has emerged as one of the workhorse model organisms for the study of genetic components in the cell.

Fruit flies have several key features, which made them the favourite organism for studying genetic principles in the early 20th century and later genomes at the dawn of the 21st century. It is an experimentally tractable organism which makes it suitable for studying components of the genomes of higher eukaryotes and also to model human diseases (Pandey & Nichols 2011; Reiter et al. 2001). The sequence of the fly genome was published almost a year before that of the human genome (Adams 2000) and is one of the best annotated genomes available. This is largely made possible by the presence of an up to date and regularly curated online database and resource called FlyBase, which has been serving the community of fruit fly researchers since 1992 (Consortium 2003).

Fruit flies offer the unique possibility to swiftly move from high-throughput screen to functional analysis of genes. This has become even more convenient with readily available mutant alleles, genetic constructs, RNAi reagents and

antibodies. Of particular importance is the ability to disrupt or mis-express genes *in vivo* through the GAL4 upstream activator sequence (UAS) transactivation system and the ability to gather loss-of-function alleles from libraries with p-element transposons, directed mutagenesis and also from ethyl methanesulphonate (EMS) suppressor screens (Spradling et al. 1999; Kelso et al. 2004; Matthews et al. 2005).

The lifecycle of fruit flies (Figure 1.9) has easily separable developmental stages and is short and convenient enough for growing synchronised populations in bulk for biochemical and even large-scale proteomics (AP-MS) analysis.

Fruit flies also have widely cultivated and well characterized cell lines, like S2, S2R+ and Kc167. The S2 cell line owing to its hemocyte ancestry are phagocytic in nature and highly susceptible to double-strand RNA induced gene silencing (RNAi) making it a very cost-effective and technically convenient system for functional analysis as compared to siRNA-induced RNAi in human cells. They can also be induced to immobilise on surfaces, making them excellent specimens for high-resolution microscopy (Rogers & Rogers 2008; Goshima 2010).

For studies with a systems biology approach, a number of whole genome high throughput datasets have been published over the last decade, which include gene expression datasets (Arbeitman et al. 2002), binary protein interaction data using yeast-two-hybrid methods (Giot et al. 2003), co-complex protein interaction data (Guruharsha et al. 2011), phospho-proteomics datasets (Zhai et al. 2008), and the recent publication of the highly annotated modENCODE dataset of regulatory DNA elements in the genome (Roy et al. 2010).

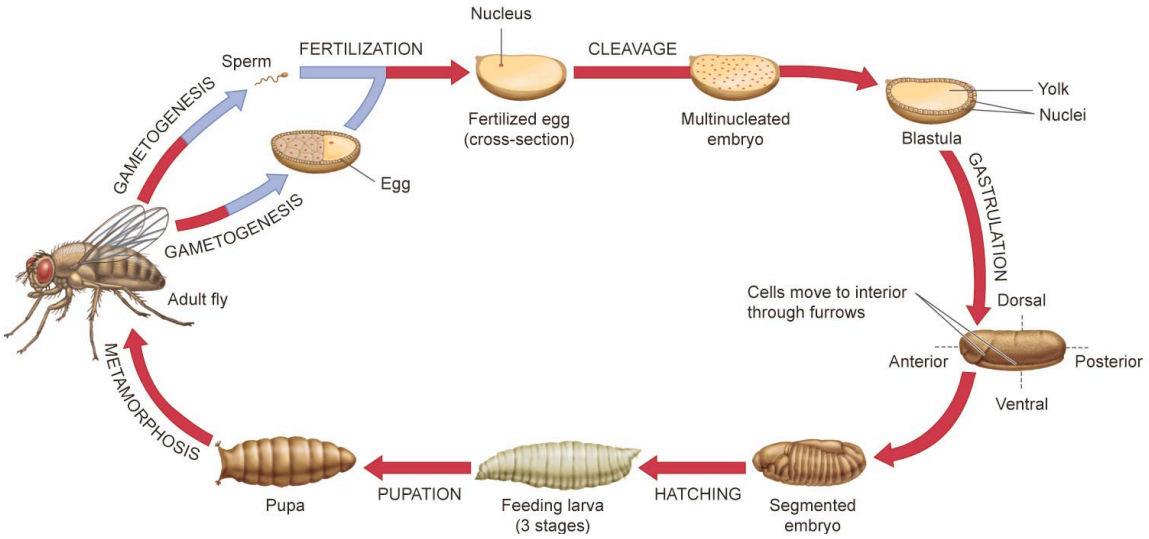


Figure 1.9: The life cycle of *Drosophila melanogaster*. The short life cycle of fruit flies (egg to adult in 10 days at 25°C) make it a very useful tool for studying genes and their products. The early embryos are excellent systems for studying microtubule dynamics as multiple nuclei divide synchronously parallel to the cortex slowing easy visualisation though fluorescence microscopy. Image taken from (KAP 2013).

1.5 Thesis Outline

Network biology, as described in this Chapter, offers the opportunity to analyse biological processes with a systems-level outlook. This thesis uses a network-driven approach to study mitosis in fruit flies (*D. melanogaster*). Using the *a priori* knowledge that microtubules are the building blocks of the mitotic spindle – the main cellular machinery that drives cell division – other putative microtubule-associated proteins (MAPs) are expected to play a role in mitosis.

Chapter 2 describes how I gather a set of experimentally validated MAPs from studies in different species and use these seed proteins to create a network. The network does not include a large subset of these MAPs owing to the poor coverage of PPI data around these proteins. We resolve this problem using two

strategies, 1) by transferring conserved interactions (interologs) from other species to the MAP network in *Drosophila* and 2) by adding proteins that bind two or more MAPs in the network. The later type of proteins, referred to as indirect interactors not only help in capturing back many known MAPs, but also add more putative MAPs into the network based on the assumption of common function due to direct interaction (guilt-by-association).

By adding attributes from different types of 'omics' and other datasets to each node in the network, we use a supervised machine learning method to fit a model and score each of our potential MAPs based on their likelihood of having a mitotic function.

Analysing the top 100 of the high scoring nodes in the model reveals interesting insights, including a well-connected subcluster. In Chapter 3 I conducted a comprehensive gene-knockdown screen using RNA interference against each member of this subcluster to test if the results of our prediction model can be replicated *in vitro*. The cells were stained using immuno-fluorescence against proteins related to the centrosomes, chromosomes and the spindle, and then analysed using microscopy.

Chapter 4 describes how a select cohort of proteins from the subcluster were cloned and tagged with GFP proteins for *in vivo* studies, which include the cellular localization of these proteins during mitosis and interphase, and a proteomics analysis of their interactions. In Chapter 5, I summarise our findings and highlight future directions.

Chapter 2
The integrated MAP interactome

2.1 Introduction

Over the last decade, increasing evidence has been reported that supports the notion that all diseases are complex in nature, even the simplest Mendelian disorders (Barabási et al. 2011). As an alternative to simple linear pathways, networks are natural candidates to study these complex systems and the intricate crosstalk between different underlying pathways.

The second phenomenon, which has emerged in the field over the last decade, as discussed in Chapter 1, is the surge in large-scale 'omics' datasets. This has provided an unprecedented opportunity for scientists and clinicians to study whole datasets and understand the underlying principles of biological systems on a global, systems-level scale. Genome-wide protein-protein interactions datasets also has the capacity to become a 'scaffold' or 'skeleton' to integrate other types of omics data (Braun 2012; Liu et al. 2012).

So far the standard work-flow for using protein interactions to study a specific biological problem, is to identify seed proteins relevant to the biology at hand and use these to gather interaction data from different databases in order to construct a relevant sub-network (Sanz-Pamplona et al. 2012).

This is followed by one or more different types of analysis that can help in deriving hypotheses about the system in question, in a well-informed manner (Ma'ayan 2009; Goh et al. 2012). These include, but are not restricted to, topology-based methods that study structural characteristics for example, network centrality measures (Gu et al. 2012; Scardoni et al. 2009) and network motif distributions (Milo et al. 2002; Alon 2007), module-based methods that identify and analyse densely connected clusters or modules (Luo et al. 2007) and Gene Ontology-based methods that search for over-representation of GO classes and terms (Maere et al. 2005).

2.1.1 Network-based integrative methods

Network biology broadly aims to integrate *omics* datasets, with interaction data as the template or scaffold, for a combined analysis with a holistic perspective (Bebek et al. 2012b). An area also referred to as Network-Enabled Wisdom-based or 'NEW' biology (Schadt & Björkegren 2012).

Network-based approaches are being used in a variety of systems, for example, in attempts to prioritize genes in human diseases (Vanunu et al. 2010), identifying the core cell cycle network in *Arabidposis* (Van Leene et al. 2010) and network-based analyses of genome-wide RNAi data in *Drosophila* (Wang et al. 2009).

In *Drosophila* (our model organism), several research groups have started using different integrative approaches. Guest *et al*, embarked on a study of the cell cycle by conducting a 'virtual' protein-protein interaction screen using seed proteins from two RNAi screens. In this study they showed how interaction data

can help analyse data from large-scale RNAi screens producing higher quality hits, removing a significant proportion of false positives and false negatives (Guest et al. 2011a). In another study, Chen *et al* used network-based analysis and identified cliques in the mitotic spindle network (closely related to the system in our study) that were members of kinetochore-associated complexes (T.-C. Chen et al. 2009).

There are other studies close to both our model system (mitosis) and model organism (*Drosophila*). For example, Ohta *et al* analysed the composition of mitotic chromosomes using a machine learning method known as Random Forest analysis. Random Forest is a classification method that constructs decision trees using random subsets of the training data that are then used to predict output values for the test data. They used features that were derived from a series of quantitative proteomics experiments, which involved isotope labelling (Ohta et al. 2010). Yan *et al* also used a Random Forest algorithm to predict gene function in *Drosophila*, on a larger genome-wide level (Yan et al. 2010). They used their model to predict GO and Kyoto Encyclopaedia of Genes and Genomes or KEGG memberships (Consortium 2000; Kanehisa et al. 2012).

A more recent study by Rojas *et al*, used several layers of omics data to identify proteins associated with the human mitotic spindle (Rojas et al. 2012). This was a comprehensive study, which used separate classifiers for each type of data and then an integrated model combining all classifiers to predict proteins related to the mitotic spindle within the human proteome. They report a 75% hit-rate after experimental validation.

2.1.2 Our Study in Drosophila

In this project we develop an integrative network-based model with an aim to find new potential microtubule-associated proteins, or MAPs, that might have a role in mitosis.

By using prior knowledge about microtubule-associated proteins and their role in mitosis and using a network-based strategy to narrow our focus, we enhance the signal-to-noise ratio that is inherent to large-scale genome-wide studies (e.g. (Yan et al. 2010) and (Rojas et al. 2012)). This, at least in part, explains why the model produced in this study performs better than the above-mentioned studies, even using a simpler method, logistic regression, rather than Random Forest algorithms.

In the first part a network of MAPs, called the MAP interactome, was created. This network included two classes of proteins, experimentally determined MAPs found in different organisms including *Drosophila*, and other proteins referred to as indirect interactors that connect two or more of these MAPs in the network. The techniques to transfer nodes and edges from other organisms into the MAP network and the addition of indirect interactors were adopted from the work of (Fisher 2009).

In the second step, different types of publicly available high-throughput datasets were integrated to produce a high performance prediction model, which ranked all the proteins in the MAP network according to the likelihood of possessing a putative role in mitosis. The second part will not only guide experimental work

in narrowing down on key candidates, but will also help answer the broader question of whether we have enough data to predict protein function.

2.2 Data

Protein interaction data is routinely stored in a pair-wise format representing the interaction between two proteins. This data can be assembled together into a network where each node represents an individual protein and each edge (line) between two nodes represents the interaction between two proteins.

2.2.1 MAP datasets

The original dataset

The seed proteins which form the basis of the MAP interactome are based on the MAP dataset gathered by Hughes *et al* in the Wakefield Lab (Hughes et al. 2008). Using a proteomic analysis, which employed a microtubule co-sedimentation assay coupled with 2D-PAGE and Mass Spectroscopy, they isolated 269 microtubule-associated proteins from *Drosophila* embryos (Table 2.1).

MAP datasets in other organisms

Similar MAP datasets have been reported in several model organisms, including the rock cress plant (*Arabidopsis thaliana*), rat (*Rattus norvegicus*), mouse (*Mus musculus*) and humans (*Homo sapiens*) (Table 2.1). MAPs were isolated from cultured cells of these organisms using the same basic strategy involving a co-

sedimentation protocol that co-purified microtubules and associated proteins, which were then resolved by chromatographic columns (LC) and subsequently analysed using tandem mass spectroscopy (MS/MS). The human dataset was collected slightly differently, in that mitotic spindles were isolated instead of the microtubule content of cells in interphase – the non-mitotic resting phase (Figure 2.1).

These datasets provide additional nodes, which are used to expand the original MAP dataset in *Drosophila*, potentially maximizing our capture of mitotic MAPs. Table 2.1 gives the summary of all these datasets along with the number of MAPs reported.

Species	No. of MAPs	Reference	Experiment Type
Fly	269	Hughes et al., 2008	MT co-sedimentation
Human	794	Sauer et al, 2005	Purification of mitotic spindles
Rat	371	Sakamoto et al, 2008	MT co-sedimentation
Mouse	410	Patel et al, 2009	MT co-sedimentation
Arabidopsis	113	Chuong et al, 2004	Tubulin Affinity Chromatography

Table 2.1: MAP datasets from five organisms (*Drosophila*, Human, Rat, Mouse and *Arabidopsis*) used in this study. The number of proteins in each MAP dataset and their methodologies are given. All studies conduct co-sedimentation analysis of microtubules in their respective system, except the Sauer *et al* study in human cells, which analyses purified mitotic spindles.

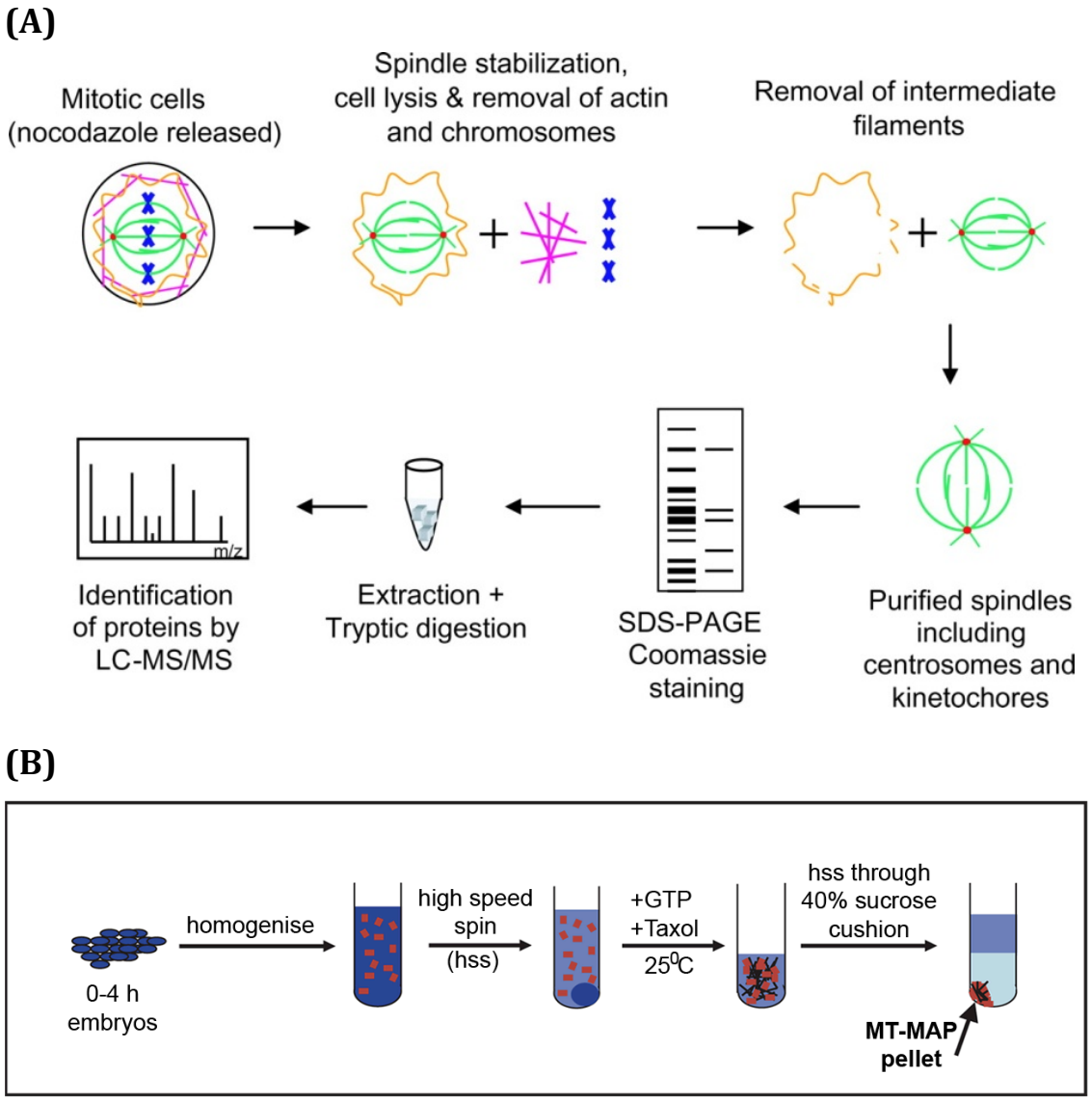


Figure 2.1: Methodologies of the MAP studies. The human dataset (Sauer *et a*l) was obtained using a spindle purification method (A), while all the other MAP datasets were similar to the MT-co-sedimentation assay used by Hughes *et al (*B). Diagrams adapted from (Sauer et al. 2005) and (Hughes et al. 2008).

The human dataset was reported by Sauer *et al* in a study which enriched for mitotic cells in HeLa S3 cell lines (Sauer et al. 2005). This was done by synchronizing the cells at G_1/S phase and then incubating them in a medium with

nocodazole that blocks all cells at the metaphase stage of mitosis. Spindles were then harvested from these cells after lysis, using centrifugation. Analysis using LC-MS/MS then identified 794 proteins, 154 of which were previously uncharacterized.

MAPs in mouse were identified by Patel *et al* using a co-sedimentation assay to isolate taxol-stabilized microtubules from a macrophage cell line (Patel et al. 2009). Subsequent LC-MS/MS analysis identified 410 proteins. The study compared the MAP interactomes of inactive and activated macrophages and identified 94 proteins with differential gene regulations.

A similar proteomic analysis of microtubule-associated proteins was conducted in *Arabidopsis* by Chuong *et al* (Chuong et al. 2004). They used a tubulin affinity chromatography followed by LC-MS/MS and identified 113 microtubule-associated proteins.

Sakamoto *et al* conducted their MAP proteomic analysis in rat brain cells (Sakamoto et al. 2008). In addition to the MT co-sedimentation protocol and LC-MS/MS analysis this study also incorporated an MT affinity column chromatography to enrich and detect low abundance MAPs. They identified a total of 371 proteins, 42 of which were previously uncharacterized.

2.2.2 Interaction Data

The fly protein-protein interactions (PPI) data were gathered from BioGRID (Stark et al. 2010) and DIP (Salwinski et al. 2004). 39402 physical interactions from both databases were amalgamated and stored in a single file after removing

duplicate interactions (as of May 2010). These included interactions obtained from both binary yeast-2-hybrid and co-complex methods. BioGRID was the source of protein-protein interactions for the Human (42496 interactions), Mouse (1660 interactions), Rat (573 interactions) and *Arabidposis* (3573 interactions) proteins as well (as of May 2011). All proteins were mapped and stored as UniProt accession numbers (inclusive of all isoforms).

2.2.3 MAP features

The features used in the prediction model to identify MAPs that have a possible role in mitosis came from a variety of high-throughput datasets, including gene expression, protein domains, genome-wide gene knockdown screens, cross-species conservation of MAPs based on sequence homology and, genetic and protein-protein interaction data. Different types of features were derived from these raw datasets and used in different combinations before reaching the final predictive set of features with the highest performance (Table 2.6). The final features gathered for each protein of the MAP network, their sources and biological relevance are as follows. The coverage of each feature in the training set, MAP network and the entire fly proteome are given in Table 2.2. The table clearly shows the coverage is far from complete with the existing data available. In the next few sections, the source of each of these features is discussed in more detail.

Features	Data Source	% Training Set	% MAPs	% Proteome
PmitoG	BioGRID+DIP	71.54	67.11	17.85
GmitoG	BioGRID+DIP	45.09	20.46	7.30
mitodom	CDD	36.52	13.23	19.36
cevi	BLAST	36.78	19.64	1.62
exp2cut	Graveley et al	66.25	51.59	80.79
m24	Gauher et al	78.34	76.37	57.20
e15	Gauher et al	78.34	76.37	57.20
RNAi	12 GW studies	25.44	8.91	6.06
Xeno-evi	BLAST	4.28	2.73	1.51
Rat-evi	BLAST	12.59	7.32	0.49

Table 2.2: The data sources and the coverage of the final MAP features in the training set, the MAP network and the entire *Drosophila* proteome. The features include mitotic neighbours (PmitoG, GmitoG), gene expression (e15, m24 and exp2cut), protein domains (mitodom), genome-wide RNAi screens (RNAi) and cross species conservation (cevi, xeno-evi, rat-evi).

Mitotic Neighbours (PmitoG, GmitoG)

The first type of feature is the number of mitotic neighbours of each protein. This is based on the widely accepted assumption of guilt-by-association, i.e. interacting proteins are likely to have common function (Coulombe et al. 2008).

Two features fall under this class, one based on protein interaction data and the other on genetic interactions (PmitoG and GmitoG). Each feature included the number of direct mitotic neighbours for each protein in the network. Mitotic

proteins were defined as proteins, which had at least one mitosis-related term in their ascribed Gene Ontology annotation (Table 2.3).

Protein interactions were obtained from BioGRID and DIP (as of May 2010). Genetic interactions (GI) were gathered from FlyBase (Consortium 2003). All proteins were mapped to their corresponding FlyBase accessions (FBgn numbers) before their neighbours were counted.

Gene Expression (e15, m24 and exp2cut)

Two features were compiled from a genome-wide gene expression dataset present in FlyBase. These two features were selected after testing raw and different scaled versions of expression data, and expression data from different time points for the best predictive power.

'mitotic' GO terms
Mitosis
Meiosis
Chromosome
Chromatid
Centromere
Centriole
Centrosome
Kinetochore
Spindle
Microtubule
Cell division

Table 2.3: Gene Ontology (GO) terms used to gather mitosis-related proteins. Proteins with GO annotations that contain at least one of these terms were included in the list of mitotic proteins.

The first gene expression dataset used as a feature was reported by Gauher *et al* in 2008 (Flybase communication). Raw levels of expression were used for each protein from two different developmental levels as reported by the study, i.e. two hour-old embryos and 24 hour-old pupae (e15 and m24). Since greater mitotic activity occurs in early embryonic stage, mitotic proteins were assumed to have a higher expression at this stage, compared to later larval stage, which undergoes little cell division.

A second and more recent gene expression dataset (Graveley et al. 2010), used to obtain the expression of 2-hour embryos (exp2cut). The raw data was scaled and used as a feature, i.e. '0' for relative expression values below zero, '1' for values between 0 and 2000, and '2' for values above 2000.

Protein Domain Composition (mitodom)

Domain compositions of each protein in the network were analysed for any overlap with domains found in 'mitotic' proteins, which could imply similar function. Domains were gathered for each protein by using the Conserved Domain Database (CDD) Search tool from NCBI (Marchler-Bauer et al. 2010). A BLASTp (Altschul SF, Gish W, Miller W, Myers EW 1990) with an e-value cut-off of 10^{-10} was used to search for domains from CDD. The domains present in all 'mitotic proteins' were first gathered in a list. Then the number of 'mitotic' domains from that list, present in each of our protein was counted and used as a feature (mitodom).

Genome-wide RNAi screens (RNAi)

Several large-scale gene-knockdown screens using RNA interference have been published which have reported hundreds of mitotic proteins. The proteins, which have mutant mitotic phenotypes reported, were compiled in a list. Table 2.4 shows the details of the studies used to gather RNAi hits. In this feature (RNAi), a binary score was assigned to each protein in our network with a '1' for reported hits in any of these screens, and a '0' for those ones with no reported mutant phenotype after knock-downs.

Reference	Focus of screen	Hits
O'Farell *et al,* 2004	Cytokinesis	36
Field *et al,* 2004	Cytokinesis	214
Battencourt *et al,* 2004	Kinases in mitosis	80
Goshima *et al,* 2007	Mitotic spindle assembly	200
Glover *et al,* 2007	Phosphatases in mitosis	22
Gatti *et al,* 2008	Mitosis	155
Pellman *et al,* 2008	Centrosome clustering	133
Straight *et al,* 2008	Centromere propagation	4
Raff *et al,* 2008	PCM recruitment	32

Table 2.4: Genome-wide RNAi screens in *Drosophila* that focused study mitosis-related functions. The focus of each screen and the number of mitotic hits reported in each of them are given.

Cross-species conservation based on sequence homology (cevi)

Finally, data on proteins with homologs in MAP datasets of five other organisms were used. Lists of fly homologs were obtained using the reciprocal-best-hit

method. This included the Human, Rat, Mouse and *Arabidposis* experimental MAP datasets and a recently reported dataset in *Xenopus* (Gache et al. 2010). The number of homologs for each protein in the MAP network was computed from the datasets of each organism. After testing, homologue numbers for the *Xenopus* (xeno-evi) and Rat (rat-evi) datasets were used for their predictive power. A cross-species MAP evidence (cevi) was also used which was the number of MAP datasets (total of the five datasets in five species) in which each fly MAP from the network had at least one homolog.

Training data

The positive and negative training datasets used for the prediction model were based on the Gene Ontology annotation (Consortium 2000). The positive training set with 209 proteins (out of 1324 proteins in the MAP network) contained proteins that had well-characterized mitotic functions and possessed at least one 'mitotic' term in their GO annotations as described earlier (Table 2.3). The negative training set included 209 of the most well-annotated proteins, i.e. proteins with the highest number of GO terms ascribed to them, and which did not contain any of the selected 'mitotic' term (Table 2.3).

2.3 Methods

2.3.1 Building the MAP network

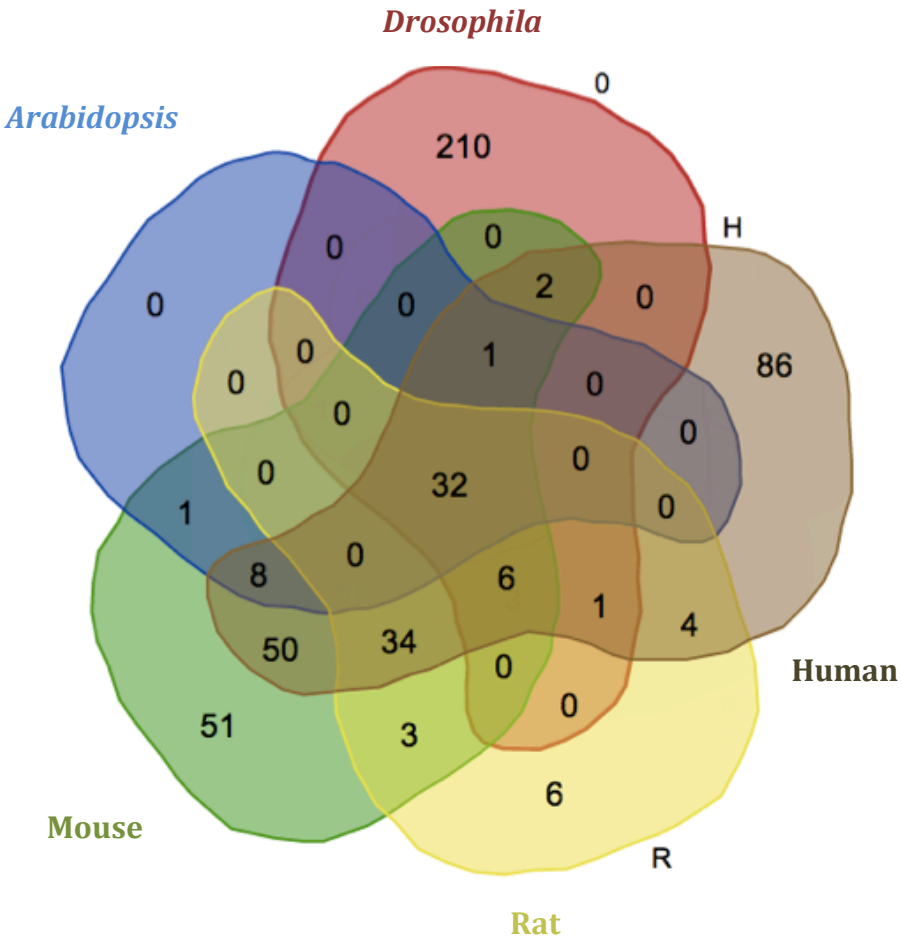
Fly seed proteins

The original MAPs dataset from *Drosophila* (Hughes et al. 2008) with 269 proteins was used as the seed proteins for building our extended MAP interactome.

Transferring Homologs

The *Drosophila* homologs of MAPs from the other four MAP datasets in Human, Mouse, Rat and *Arabidopsis* were used to extend the dataset used as the seed proteins. These were gathered by using a reciprocal best hits method. This method involved a BLAST analysis of these MAPs against the *Drosophila* protein database from the UniProt database (Magrane & Consortium 2011), using a strict e-value cut-off of 10^{-10} (Oliynyk et al. 2007). *Drosophila* hits from these BLAST results for each protein were then subjected to another BLAST, against the original species' database, also from UniProt. Proteins, which did not give the original query protein as a 'hit' were discarded, and the remaining were saved as *Drosophila* homologs from each dataset. The addition of homologs to the original set of 269 proteins brought the total to 495 *Drosophila* MAPs.

The overlap between *Drosophila* MAPs and *Drosophila* homologs of all other MAP datasets is shown in detail in Figure 2.2. Thirty-two proteins are found across all datasets. The highest overlap is observed between the human and mouse MAP datasets with 50 proteins in common. Most *Drosophila* proteins (210 out of 269) have no overlap with any other dataset.



Datasets	# Proteins
D	210
H	86
M	51
H M	50
H M R	34
A D H M R	32
A H M	8
D H M R	6
R	6
H R	4
M R	3
D H M	2
A D H M	1
D H R	1
A M	1

Figure 2.2: Venn diagram showing the overlap between *Drosophila* MAPs (D) and *Drosophila* homologs of Human (H), Mouse (M), Rat (R) and *Arabidopsis (A)* MAP datasets. The table gives the number of unique proteins in each category.

Transferring Interologs

After gathering homologs, interologous interactions were also transferred from Human, Mouse, Rat and *Arabidopsis* binary protein interaction data. The concept of 'interologs' was first proposed by Walhout *et al* (Walhout 2000). In their study they used it to analyse new interactions between proteins in *C. elegans*. This method predicts interactions in one organism based on interactions between protein orthologs in other organisms, i.e. if proteins A and B have orthologs A' and B' in another organism which are known to interact, then a predicted interaction is assigned between proteins A and B, which is called an interologues (Figure 2.3).

Matthews *et al* (Matthews et al. 2001) later used the same concept on a large-scale with a more systematic approach to predict interactions in different species based on their protein orthology. Recently, interologues have been used to predict new interactions in a variety of different species (Wiles et al. 2010) The addition of homologs and interologues for the four organisms produced a tightly connected network when compared to the original dataset.

Due to the inherent limitations of high-throughput interaction data like high error rates and incompleteness, several studies have examined the evidence for homology based transfer of protein-protein interactions and have concluded that with the data in hand very limited inferences are correct (Mika & Rost 2006; Lewis et al. 2012). Lewis *et al*, therefore call for a strict definition of homology when transferring interactions across species, i.e. a low e-value cut-off when using BLASTp (Altschul SF, Gish W, Miller W, Myers EW 1990) for identifying homologous proteins (Lewis et al. 2012).

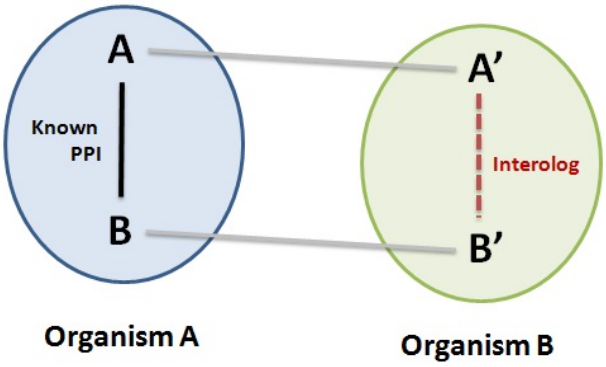


Figure 2.3: A schematic of interologs. Proteins A' and B' are homologs of proteins A and B, respectively. The red line shows the interologues transferred to organism B.

Interologues were transferred from each organism using the same homology tables and the BioGRID database for interaction data of all the species. The very conservative reciprocal best hits approach and a fairly strict e-value threshold of 10^{-10} was used to maximize correct inferences of interologouss interactions. This completed our list of experimentally determined MAPs in *Drosophila* and homologs from other species.

Indirect interactors

Additional proteins were added to the network on the basis of a strict criterion, i.e. they bind to at least two experimentally determined MAPs (Figure 2.4). These proteins, referred to as indirect interactors, were valuable additions to our list of potential MAPs, as they are connected to known MAPs and are hence more likely to share a common function. They also increase the connectivity of our network and allow us to re-capture 128 MAPs, which had no reported direct interactions to any other MAP. It essentially mimics one aspect of co-complex experiments, which unravel both direct and indirect interactions between proteins.

Addition of indirect interactors gave a fully connected network with 1324 nodes all together in a single giant component compared to the networks based on the two previous sets in Table 2.5. Figure 2.6 shows the network visualization in Cytoscape.

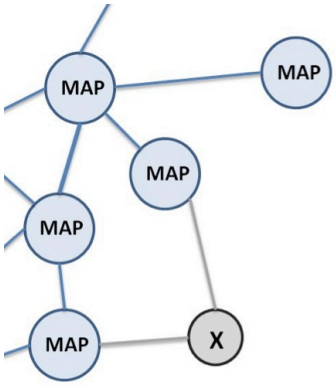


Figure 2.4: Indirect interactors are any proteins from the *Drosophila* proteome that are attached to two or more MAPs. Note that the indirect interactor, protein 'X', is part of an open triplet (incomplete triangle).

Calculating Summary Statistics

The following summary statistics were computed using NetworkAnalyzer on Cytoscape (Smoot et al. 2010), or custom scripts using Perl (Appendix I).

- Degrees correspond to the number of neighbours a node (protein) has in the network. The average degree for a network is the average of the degrees of all the proteins in the network.
- Clustering coefficient is the number of edges between a protein and its neighbours divided by the total number of possible edges between the protein and its neighbours. The network clustering coefficient is the average of the clustering coefficients of all the nodes in the network.

- Shortest path length refers to the shortest path between to nodes. The characteristic path length or average shortest path length for a network is the average of all the shortest path lengths in the largest component of the network.
- Network diameter is the maximum of all shortest path lengths.

2.3.2 Prediction model

All MAP features were compiled in a matrix and fitted using logistic regression based on the ordinary least squares method using MATLAB.

Preliminary Data analysis

Before fitting a regression model using all the features, a basic preliminary data analysis was conducted to check if the key assumption of regression models, i.e. whether each feature had any linear correlation with the outputs in our training set or not, is satisfied. Figure 2.5 shows the plots and correlation coefficients for our final subset of predictive features. The coverage for each feature was also computed in our training set, the proteins in the entire MAP interactome and the entire fly proteome for comparison (Table 2.2).

Accuracy of the model and Feature Selection

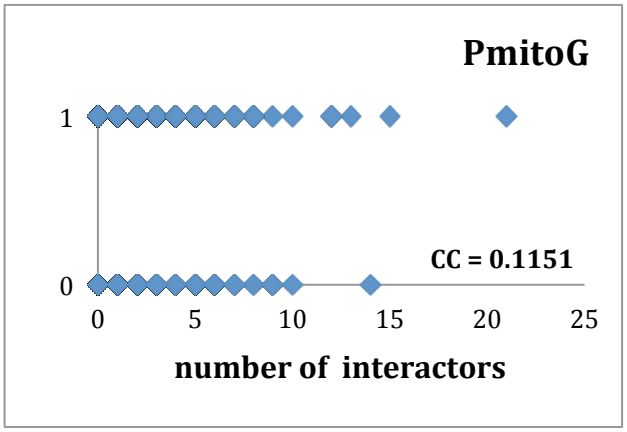
Several features were used in the initial run of the model and a predictive subset was selected from them on the basis of the significant p-values of their coefficients in the fitted equation ($p<0.05$). After feature selection, the accuracy of the model was tested by doing a 10-fold cross validation. Receiver operating characteristic, or ROC, curves were used to analyse the performance of each test.

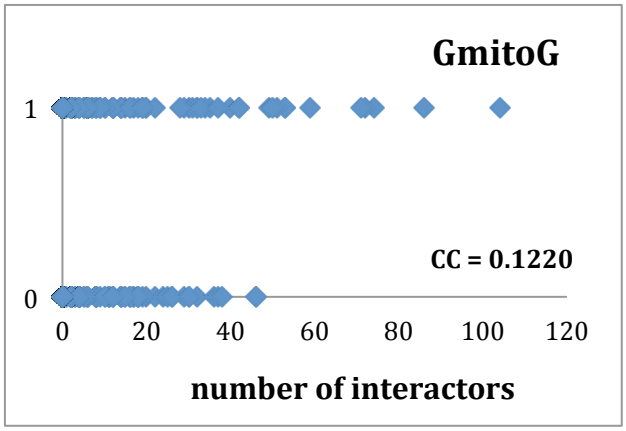
The ROC curve is a plot that is used to illustrate the performance of a classifier system. A ROC curve is obtained by plotting the true positive rate (fraction of true positives in the positives) and the false positive rate (fraction of false positives out of the negatives). A common statistic derived from the ROC curve is area under the curve or AUC. The higher the area under the ROC curve, the better the performance of the classifier system.

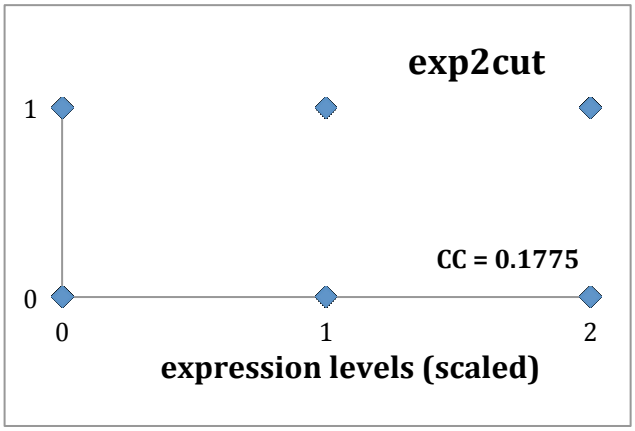
Another test for model selection, the Akaike Information Criterion (AIC) was also computed to compare different models produced in the test run (Akaike 1974). AIC was calculated by the following formula:

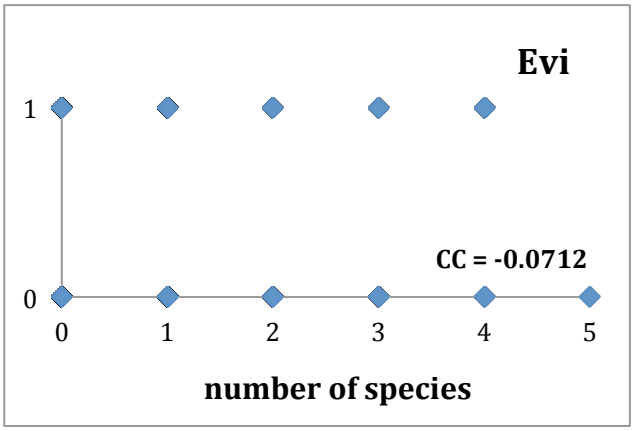
$$AIC = n \log (RSS/n) + 2k$$

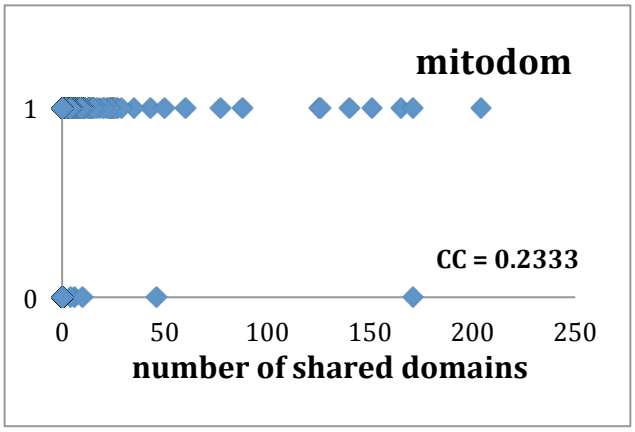
Where n is the number of observations, k is the number of features and RSS is the residual sums of squares. AIC was used to measure the relative quality of the model based on the trade-off between goodness-of-fit and the complexity of the model.

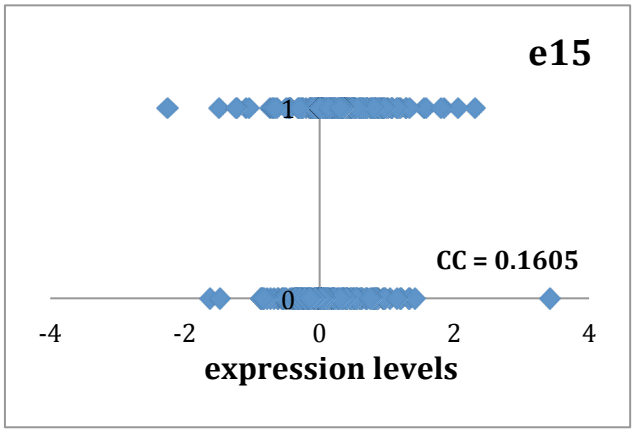

PmitoG
1
0
CC = 0.1151
0 5 10 15 20 25
number of interactors

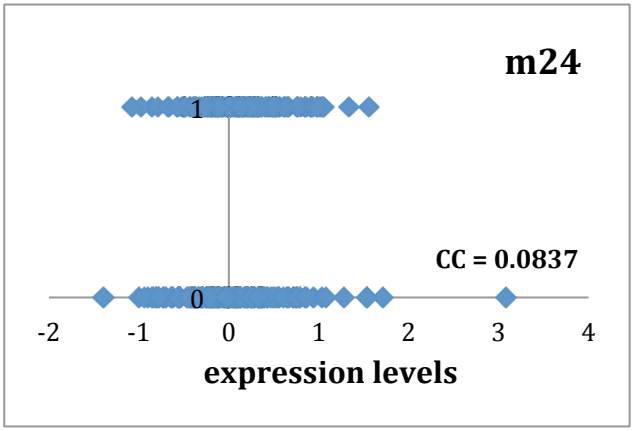

GmitoG
1
0
CC = 0.1220
0 20 40 60 80 100 120
number of interactors

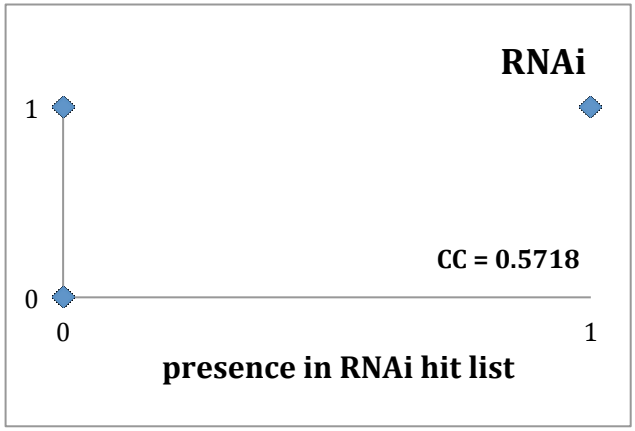

exp2cut
1
0
CC = 0.1775
0 1 2
expression levels (scaled)


Evi
1
0
CC = -0.0712
0 1 2 3 4 5
number of species


mitodom
1
0
CC = 0.2333
0 50 100 150 200 250
number of shared domains


e15
1
0
CC = 0.1605
-4 -2 0 2 4
expression levels


m24
1
0
CC = 0.0837
-2 -1 0 1 2 3 4
expression levels


RNAi
1
0
CC = 0.5718
0 1
presence in RNAi hit list

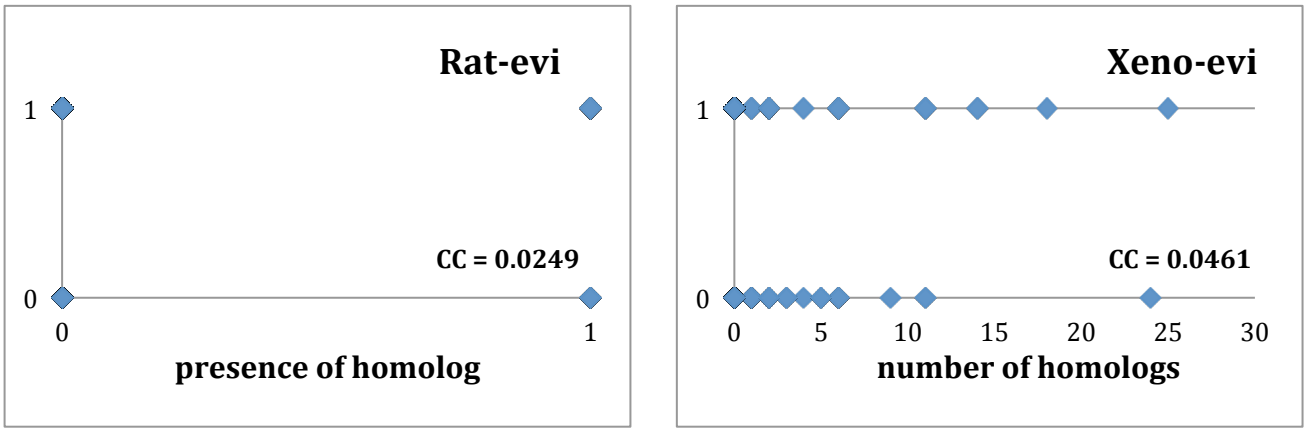


Figure 2.5: Plots showing the correlation of the features of the selected subset with their outputs along with their correlation coefficients (CC). Each data point represents a protein from the training set. The y-axis shows the output (1 – mitotic, 0- non mitotic). The x-axis shows each feature - protein-protein interactions (PmitoG), genetic interactions (GmitoG), cross-species conservation (evi), presence of 'mitotic' domains (mitodom), presence of reported RNAi phenotype (RNAi), scaled expression from Graveley *et al* (exp2cut), pupal (m24) and embryonic (e15) expression from Gauher *et al*, and numbers of orthologs in the rat (Rat-evi) and the presence in *Xenopus* (Xeno-evi) MAP datasets.

2.4 Results

2.4.1 MAP network

Starting off with the seed proteins from the original dataset only 50 of 269 proteins had experimentally-determined interactions available. Expanding this dataset with homologs and interologues from the four other MAP datasets, the number of proteins in our extended dataset became 495 of which 275 were connected to each other with 1356 interactions (Table 2.5, Figure 2.6). This not only substantially increased the average degree and average clustering coefficient of the network, but also recaptured 48 more proteins from the fly MAPs dataset (Figure 2.6b).

This extended and tightly connected network was then complemented with indirect interactors i.e. proteins that bound at least two experimentally-determined MAPs. As well as adding 851 novel proteins to the network it also helped us recapture 128 proteins from the extended MAPs dataset, which did not have direct interactions to other MAPs (Figure 2.6c). In moving from one set to the next, the average degree is slightly increased, but interestingly the average clustering coefficient decreases to 0.102 from 0.336 (Table 2.5). A possible underlying reason for this can be the contribution of indirect interactors to 'incomplete triangles' or open triplets in the network. Since the global clustering coefficient can also be defined as the number of closed triplets divided by the total number of triplets in a network, an increase in the number of open triplets will decrease the clustering coefficient. Table 2.5 also shows the diameter and

the average shortest path length for the largest component of the network, both of which increase each time the datasets are expanded.

Our 'complete' MAP network covers 8.1% of the entire fly genome. It contains 188 proteins out of the original 269 fly MAPs along with 215 homologs transferred from other organisms, the rest of experimentally–determined MAPs are lost since they do not possess any interactions with any other MAP in the network. The remaining 921 proteins in the network are indirect interactors.

Network	Proteins in dataset	Total Nodes	Total Edges	Number of components	Nodes in largest comp.	Edges in largest comp.	Average Degree	Clustering Coefficient	Network Diameter	Shortest Path Length
Original Fly MAPs	269	50	32	18	8	7	1.28	0	3	1.609
MAPs with homologs and interologues	495	275	1356	11	254	1345	5.48	0.336	9	3.128
Extended Network with indirect interactors	1416	1324	4289	1	1324	4289	6.49	0.102	8	3.836
Entire Fly Proteome	17530	9653	39402	67	9498	39301	8.162	0.022	11	4.304

Table 2.5: Summary statistics of all three networks and the entire fly proteome. This includes the network of original *Drosophila* MAPs, the extended network of fly MAPs with homologs and interologs from other species, and finally the extended network with indirect interactors. Total number of nodes in the network for each dataset may differ from the original number in the dataset as PPI links have not been reported for all proteins.

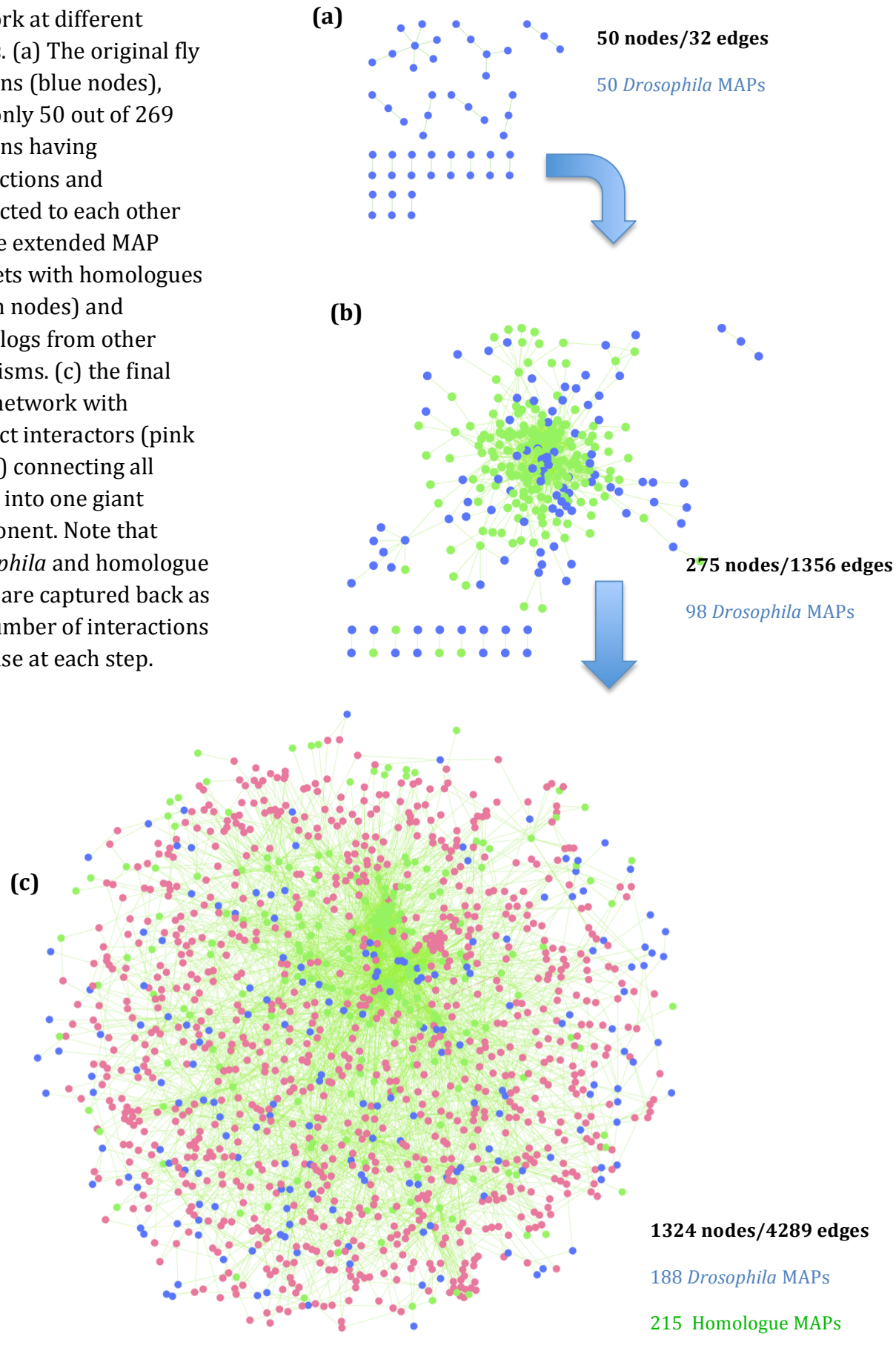


Figure 2.6: The MAP network at different stages. (a) The original fly proteins (blue nodes), with only 50 out of 269 proteins having interactions and connected to each other (b) the extended MAP datasets with homologues (green nodes) and interologs from other organisms. (c) the final MAP network with indirect interactors (pink nodes) connecting all nodes into one giant component. Note that *Drosophila* and homologue MAPs are captured back as the number of interactions increase at each step.

2.4.2 MAP prediction model

The final MAP prediction model gave a high performance as illustrated by Table 2.6. Four models were developed which differed in the number, combinations and types of features derived from the raw data in Section 2.2.3. Out of the four different models, the fourth model stands out with its AUC and AIC values. The highest AUC and lowest AIC values indicate the best quality model with the highest accuracy.

Individual features produce a smaller AUC compared to the use of all features together as shown by the ROC curves in Figure 2.7. The highest performing individual features were protein domains and RNAi while the remaining features have a similar level of performance. Cross-species conservation in MAP datasets performs poorly which is partially explained by the very low coverage. Mitotic neighbours, both using protein and genetic interaction data, perform poorly as individual features, regardless of greater coverage, but when used collectively enhance the performance of the model as shown by the exhaustive test (Figure 2.7).

All possible combinations of these features were tested in order to narrow down on the smallest predictive subset that produces the highest performance. Table 2.7 gives us the list of top performing sets. The 4^{th} set was chosen which had seven features, three less than the original ten, and gave an average AUC of 0.958 which was very close to the full set. The three excluded features were the scaled expression levels from 2-hour old embryos (Gauher *et al* dataset), the gene expression feature from 24-hour old pupa (Graveley *et al* dataset), and the cross

species conservation in the Rat MAP dataset. The final set of features was used to predict scores for all the 1324 proteins in the MAP interactome.

Model	Average AUC	Std. Dev.	AIC
1	0.722	0.062	-389.378
2	0.735	0.078	-391.64
3	0.883	0.043	-355.814
4	**0.961**	**0.096**	**-447.854**

Table 2.6: Initial models and their performances as average AUCs with standard deviations and their AIC statistics. The model with the highest AUC and the lowest AIC indicates the best model. Model 4 performs the best in this table.

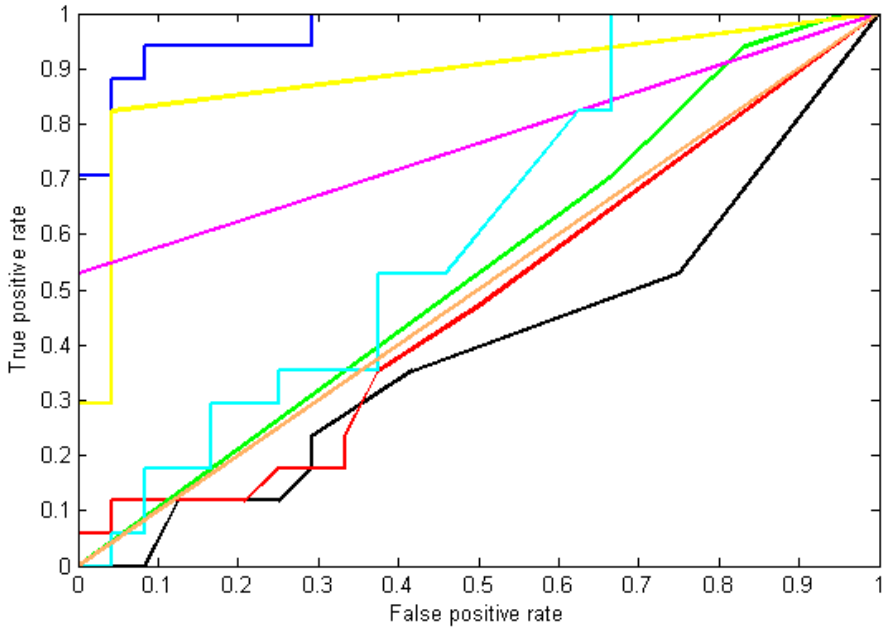


Figure 2.7: ROC curves for the individual and overall performances of all features. The blue curve represents all features used together, giving the highest area under curve. The rest are protein interactions (black), genetic interactions (red), cross-species conservation (green), domain compositions (yellow), RNAi, (magenta) and embryonic expression (cyan).

No.	Feature subset	avg. AUC
1	A C E F G I J	0.964
2	A C D E F G H I J	0.962
3	A B C E F G H J	0.961
4	A B C E F G J	0.960
5	A C D E F I J	0.960
6	C D E F 8 9 J	0.959
7	A C E F G 8 J	0.959
8	C D E F G 8 J	0.959
9	A B C D E F G I J	0.959
10	A C D E F G I J	0.959

Legend	
A	Pmito
B	Gmito
C	cevi
D	exp2cut
E	mitodom
F	RNAi
G	e15
H	m24
I	R
J	X

Table 2.7: Top 10 models obtained after an exhaustive test using all possible combinations of the selected subset of features. The 4th model was chosen and used, which minimized the set of features with little compromise on the average AUC. The legend outlines each feature and its assigned letter.

2.4.3 Mitotic MAP predictions

The ranked list based on the scores produced by the model gave very interesting biological insights. Most of the heat-shock proteins and actin binding proteins, which are often considered noise in interaction datasets, were found at the bottom of the list when ordered according to score, showing that our model could differentiate between more generic 'sticky' proteins and highly interactive mitotic proteins. Figure 2.8 shows the interactome with nodes coloured according to their scores. Almost 38% of the negative training set rank at the bottom of the list, while more than 53% of the positive set rank at the top. The number of proteins scoring zero are partially explained by the incompleteness of data for those proteins.

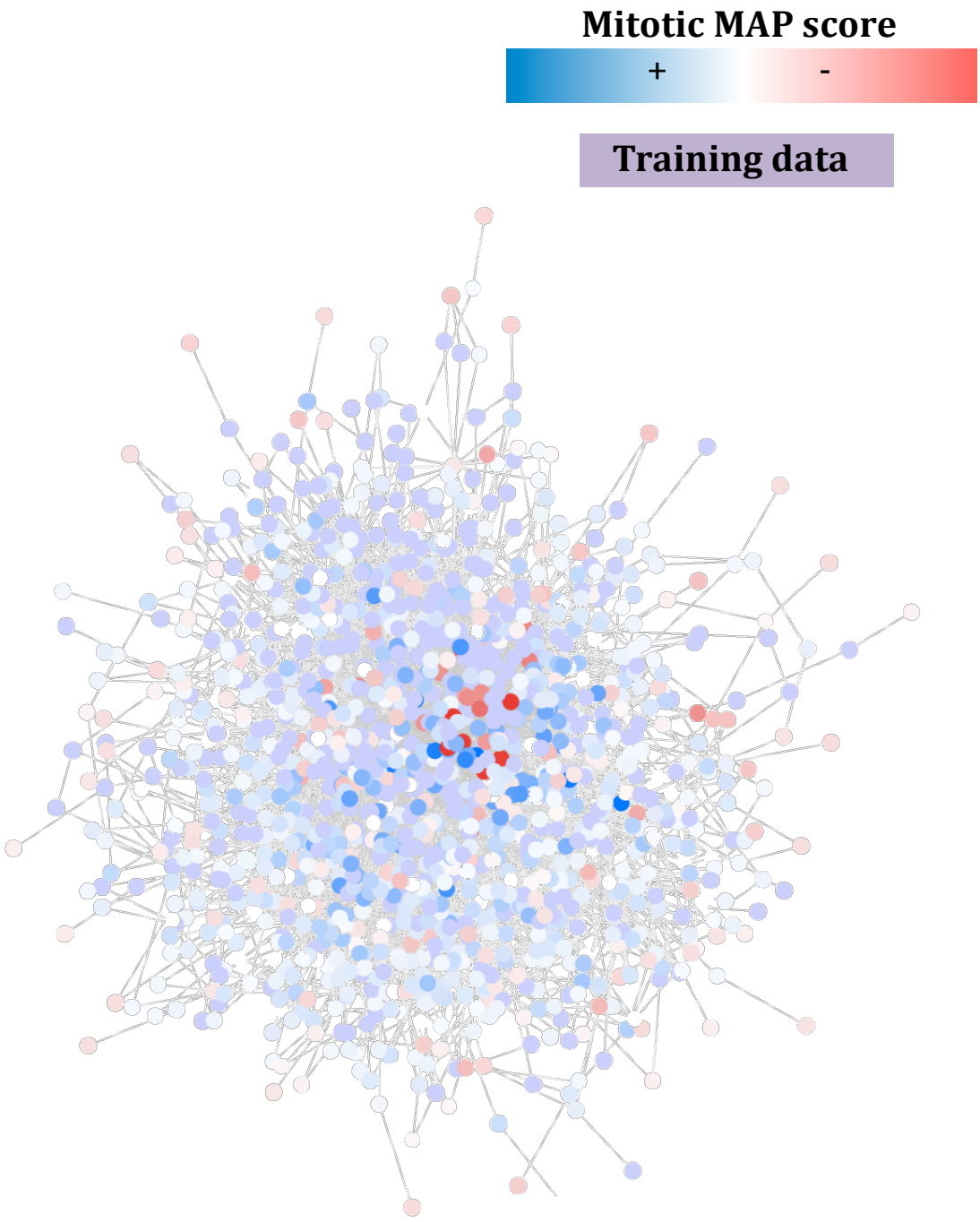


Figure 2.8: The MAP interactome, with 1324 proteins and 4289 interactions. Nodes are coloured on the basis of their predicted scores in the model. Dark blue nodes indicate scores closer to 1, which means a high likelihood of involvement in mitosis. Dark red scores indicate lowest scores closer to 0 and therefore very low likelihood of mitotic function.

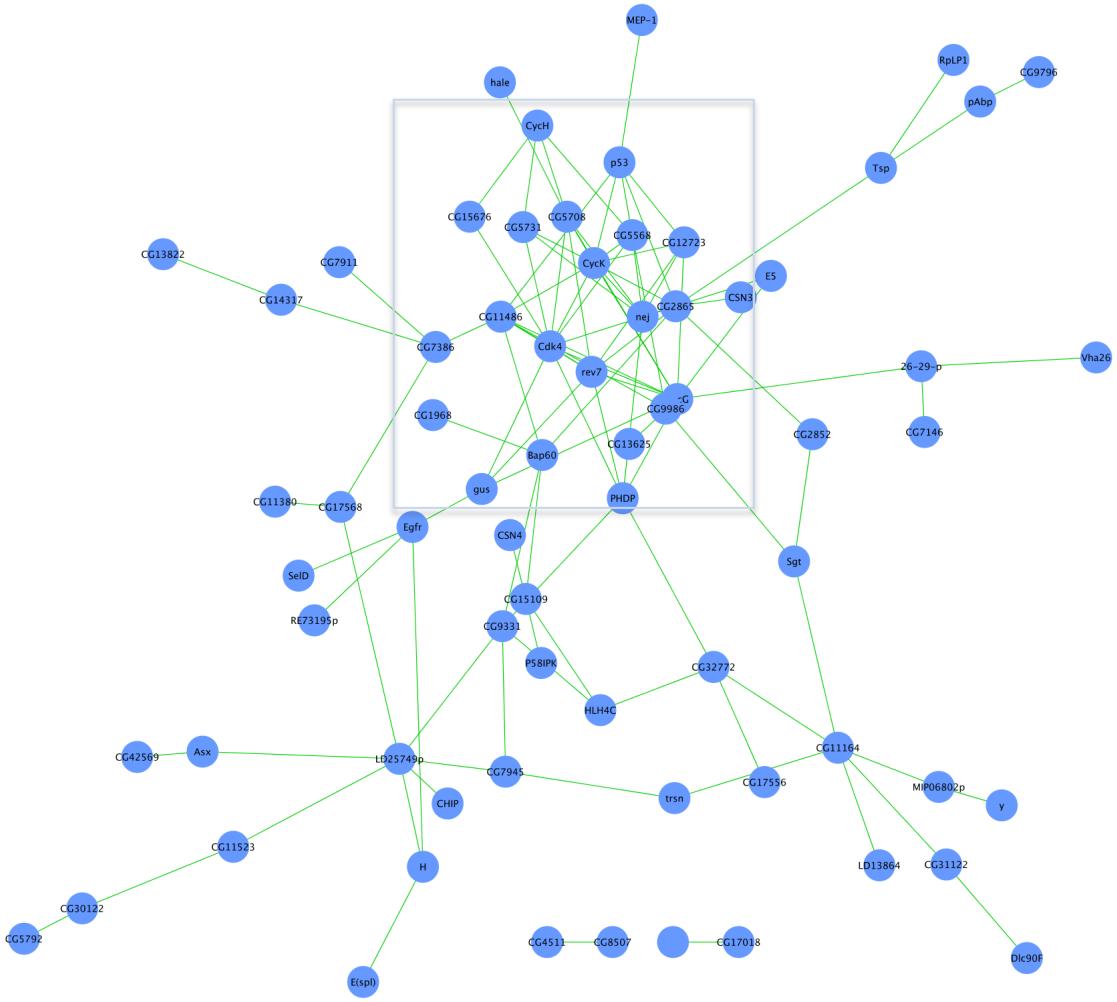


Figure 2.9: The PPI network for the top 100 proteins (based on scores in the MAP prediction model) with in the MAP interactome. A highly connected subcluster can be observed (box). This intra-MAP clustering is significant if compared to a random set of proteins. The network contains 71 proteins (out of the top 100 that had direct interactions amongst them) and 110 interactions.

A look at the screenshot of the colour coded version of the MAP interactome shows proteins that score high on the ranked list in blue while those with a low score in red (Figure 2.8).

The top 100 hits

For further analysis, the top 100 hits from the test data of our model (ranked list minus the training set) were pursued, presuming they have the highest likelihood of involvement in mitosis based on their high scores (Appendix II). A network was created to analyse the connectivity within these 100 proteins. Out of these 100 proteins, 71 had direct interactions between them and appeared on the top100 network (Figure 2.9). Although not apparent from 2D representation, roughly a third of these proteins had significantly higher connectivity between them. Presuming this implied relevant molecular function in a common biological process, the subcluster was analysed in greater detail.

In order to obtain a global view of the subcluster within the *Drosophila* interactome and to identify any known mitotic proteins in its neighbourhood, a network of this subcluster (31 out of the 71 proteins with degrees>2) was recreated using BioGrid. Analysis of the direct interactors of these proteins showed several key mitotic proteins and complexes, giving interesting insights into the different mitotic role these proteins might have within the cell. The subcluster also contained 16 previously uncharacterized proteins.

With several known mitotic proteins and numerous previously uncharacterized proteins highly connected to each other, this subcluster made an interesting candidate for our experimental validation.

The following section further analyses the members of this subcluster, termed subcluster-16, and the biological significance of its known mitotic members.

Subcluster-16

After a thorough study of all the members of subcluster-16 several uncharacterized proteins were observed that were behaving as indirect interactors of well-characterized proteins with known functions. The known mitotic/cell cycle proteins included the p53, the PNG complex, SAK and rev7/mad2. For simplicity in analysis, all 15 Cyclins and CDKs, which are known cell cycle/cell division regulators, were collapsed into a single meta-node (Figure 2.10). This leaves behind a total of thirty-three proteins in subcluster-16. Table 2.8 outlines available data about each member.

To understand the importance of these proteins and their function within the cell, a brief discussion is given in the following paragraphs.

p53 in humans is one of the most prominent of all genes involved in cancer. This transcription factor plays a pivotal role in tumour suppression and its abnormal function has been involved in almost all human cancers. The *Drosophila* version of p53 (also dmp53) shares a similar DNA binding domain, which allows it to bind human p53 response elements suggesting it might also act as a transcriptional regulator. Furthermore like the human p53, it can induce apoptosis upon over-expression. Other possible roles like transactivation and oligomerization, and the regulation of dmp53 itself remains unclear (Mesquita et al. 2010).

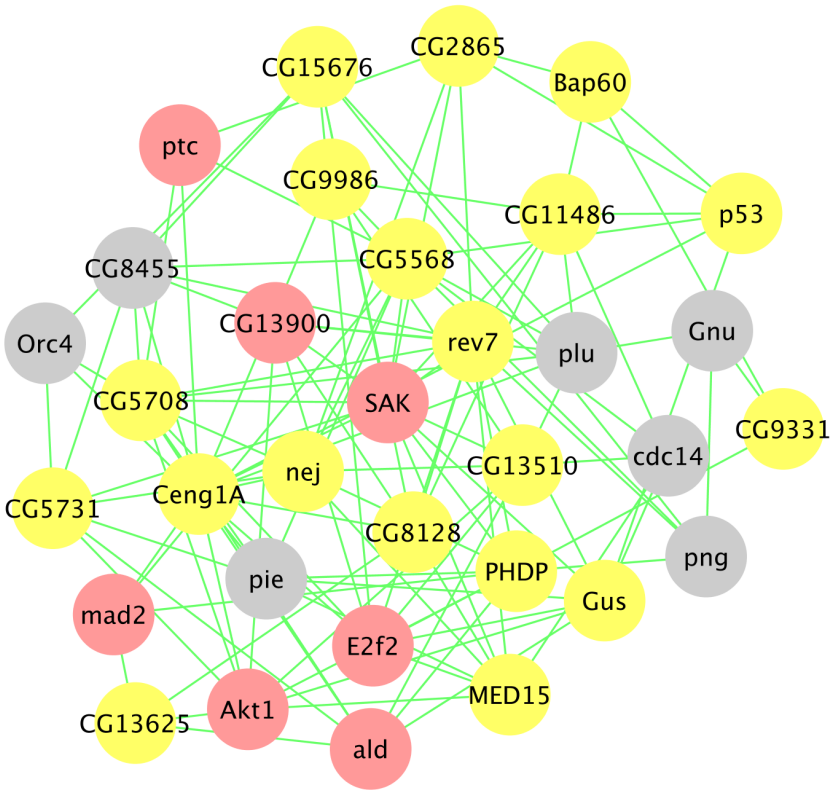


Figure 2.10: Subcluster-16 with all it members, excluding the CDKs and cyclins meta-node. The subcluster includes members of the test data (yellow), training set (red) and 6 proteins from the global neighbourhood of the proteins outside the MAP interactome.

Shamanski *et al* first reported the PNG complex in 2004 (Shamanski & Orr-Weaver 1991). They showed that Png as the kinase subunit binds two regulatory sub-units, Plu and Gnu, forming a complex, which helps in the translational regulation of Cyclin B and therefore the coordination of early S-M cycles in *Drosophila* embryos. Like CDK1/Cyclin complexes, they promote entry to mitosis and block DNA replication. Other studies have also shown the complex as an antagonist of PUMILIO-dependent translational repression, suggesting an alternative PNG-mediated route of de-repression of translation in embryonic cell

cycles (Vardy & Orr-Weaver 2007). Little is known about regulation of mitotic Cyclin translation and mechanisms of translational regulation during development in *Drosophila*.

Sak kinase (SAK) belongs to the family of mitotic regulators called Polo-like Kinases (Plks). These are serine-threonine protein kinases, which are known for their roles in centrosomal functions. The other member of the Plk family in *Drosophila* is Polo, the founding member of the family. Mammalian counterparts of SAK (Plk4) and Polo (Plk1-3) are known to be involved in cancers. In *Drosophila*, SAK has been reported to be involved in centriole duplication, making it an essential player in centrosome integrity and hence mitotic fidelity (Bettencourt-Dias et al. 2005; Swallow et al. 2005).

Cyclins and Cyclin-dependent Kinases (CDKs) make up the complex regulatory system that controls cell-cycle in the eukaryotic cell. Cyclins are stage-specific activators of the catalytic subunits of CDK proteins, which are then phosphorylated by the Cdk-activating kinase, CAK/Cdk7. In *Drosophila* CDK1, CDK2 and CDK4 regulate M, G_1/S and S, and the G_1 phase respectively, while Cyclins D, E, A and B regulate the G_1, G_1/S, S and M phases respectively. Cyclins oscillate in the cell cycle to generate different Cyclin-CDK complexes that abruptly switch on different stages of the cell cycle. They have a very crucial role in mitosis where they activate secondary proteins that affect chromosome duplication and spindle assembly. Another key player, which regulates entry into anaphase, is the anaphase-promoting complex (APC) which ubiquinates and

Gene	Flybase Accession	s16 source	MAP source	MAP scores	Embryonic Expression	Gene Ontology
nej	FBgn0261617	P	I	1.000	high	Acetyltransferase, cell cycle
CG5708	FBgn0032196	P	I	0.818	moderate	Zn-binding, LIM-type domain
CG9986	FBgn0039589	P	I	0.808	moderate	None
CG2865	FBgn0023526	P	I	0.763	moderate	Regulation of cell cycle
Bap60	FBgn0025463	P	I	0.722	v high	RNApolII TF, reg. of GE, heterochromatin assembly
CG15676	FBgn0034651	P	I	0.652	none	Chaperone binding, protein folding
CG5568	FBgn0035641	P	I	0.613	high	Ligase activity, metabolic process
CG11486	FBgn0035397	P	I	0.569	high	Protein binding, kinase-like domain
MED15	FBgn0027592	P	I	0.552	high	RNApolII TF
rev7	FBgn0037345	P	I	0.495	low	Regulation of cell cycle
p53	FBgn0039044	P	I	0.491	moderate	TF binding, apoptosis, Cell Cycle, response to UV
Gus	FBgn0026238	P	I	0.468	v high	protein binding, ant/post axis specification
CG13625	FBgn0039210	P	I	0.462	moderate	mRNA splicing
PHDP	FBgn0025334	P	I	0.452	zero-low	Transcription factor
CG5731	FBgn0032192	P	I	0.423	v low	a-N-acetylgalactosaminidase
CG9331	FBgn0032889	P	I	0.385	v high in late E	NAD-binding. Oxidoreductase activity
CG8128	FBgn0030668	P	I	0.333	moderate	Hydrolase activity, regulation of cell cycle
CG13510	FBgn0034758	P	I	0.263	high in late E/L	Cold acclimation
Ceng1A	FBgn0028509	P	R	0.000	moderate	Small GTPase mediated signal transduction
SAK	FBgn0026371	T	I	1.000	moderate	Protein ser/thr kinase, centriole replication
mad2	FBgn0035640	T	D	1.000	v high	Mitotic spindle assembly checkpoint
E2f2	FBgn0024371	T	I	0.835	high	TF, regulation of S-phase
CG13900	FBgn0035162	T	I	0.723	v high	Damaged DNA binding, mitotic spindle organisation
ptc	FBgn0003892	T	I	0.677	v low	Reg. of cell cycle/mitosis, biosynthesis
ald	FBgn0000063	T	MR	0.000	moderate	Cell cycle kinase
Akt1	FBgn0010379	T	HMR	0.000	moderate	Protein ser/thr kinase, signal transduction
plu	FBgn0003114	O	-	-	high	Regulation of DNA replication; DNA-binding
cdc14	FBgn0031952	O	-	-	low	Phosphatase; reg. of centrosome sep. & cytokinesis
pie	FBgn0005683	O	-	-	high in ov	Zn-binding, comp'd eye dev., neurogenesis
CG8455	FBgn0031997	O	-	-	moderate	None
Gnu	FBgn0001120	O	-	-	v high	Regulation of cell cycle
Orc4	FBgn0023181	O	-	-	moderate/high	DNA-binding, ATP-binding, DNA replication initiation
png	FBgn0000826	O	-	-	v high	Cell Cycle kinase

Table 2.8: Member of subcluster-16. The table shows their source in the MAP interactome and subcluster-16, along with MAP scores from the prediction model, embryonic expression data and gene ontology annotations. MAP sources include proteins from the ***D****rosophila*, **H**uman, **M**ouse, **R**at and ***A****rabidopsis* MAP datasets and **I**ndirect interactors. Subcluster sources includes proteins from the **T**raining set, **P**rediction set and **O**utside the MAP interactome.

degrades S- and M-phase Cyclins, deactivating the relevant CDKs and bringing M-phase to an end. Other highly connected Cyclins in subcluster-16 were Cyclin-J activity), Cyclin-G (role in development and meiotic recombination repair), Cyclin-C (role in snRNA 3'-end formation) and Cyclin-H (role in Cdk activation).

Mad2 has a role in the Spindle-Assembly checkpoint (SAC), a group of proteins that delays the onset of anaphase if chromosomes are not aligned properly. Mad2 binds Cdc20, which leaves APC inactive and thus delays the progression to anaphase. Chromosome segregation is a critical step in mitosis as any aberrant function and mis-segregation can lead to aneuploidy.

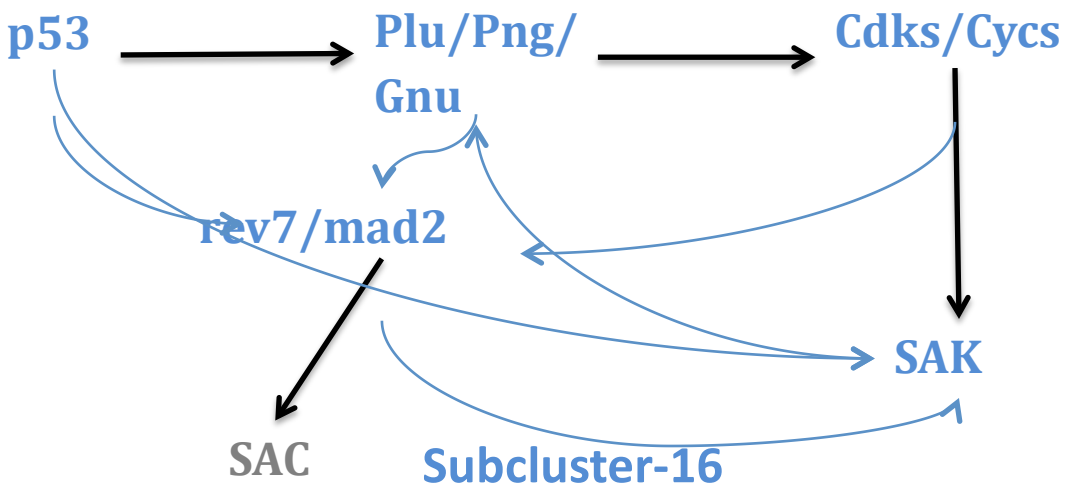


Figure 2.11: A schematic diagram showing the 5 major protein sets that exist in a sequential pathway from the G_1/S phase (black arrows). Subcluster16 suggests hypothetical cross-linking routes between these important proteins (blue arrows), via high-scoring putative MAPs from our interactome. SAK is a protein kinase with a role in centriole biogenesis. SAC refers to the Spindle Assembly Checkpoint.

Several members of subcluster-16 function as indirect interactors connecting two or more of the known mitotic proteins, suggesting the possibility of an intermediate function for these putative mitotic MAPs. Figure 2.11 shows a simplified schematic diagram highlighting the known mitotic components of subcluster-16 which are connected by dark arrows and ordered according to the stage of cell cycle/mitosis they are involved in. the lighter arrows indicate hypothetical links between these mitotic proteins through one or more of the putative mitotic MAPs which are either uncharacterized or have no reported mitotic function. These interconnecting proteins along with the subcluster make good candidates for experimental validation and functional characterization.

2.5 Conclusions

The MAP interactome was created starting off with a set of *seed proteins* from an experimental MAP dataset in the model organism of our interest, i.e. *Drosophila melanogaster*. Using a strict e-value threshold and the conservative reciprocal best hits method, homologs and interologues were transferred from biochemical MAP datasets in four other organisms, i.e. Human, Mouse, Rat and the rockcress plant, *Arabidopsis*.

The addition of indirect interactors not only maximized the retention of experimentally determined MAPs despite the unavailability of interactions, but through a 'virtual pull-down' strategy, also gathered the most likely mitotic MAPs from the fly proteome. In fact, a predominant part of the test data and the MAP network is made of this class of protein.

In order to score the nodes in the MAP network to further narrow down on the most likely mitotic MAPs, different biochemical, bioinformatics and functional datasets were integrated and fitted using logistic regression. The final prediction model ranked the proteins in the network with a very high accuracy (96%).

The top 100 hits in the test data were further analysed as candidates for experimental validation. These hits produced a highly connected cluster when observed in the fly interactome, with an interesting composition of well-characterized protein complexes involved in mitosis and a fair number of previously uncharacterized proteins (30%) that scored high on the ranked list, connecting them.

In the next Chapter the subcluster within the top 100 hits, called subcluster-16, is subject to functional validation using a thorough gene-knockdown screen. Each gene is knocked down in *Drosophila* S2 cells using an RNA interference protocol and analysed using immunofluorescence for mutant phenotypes during mitosis. A subset of the uncharacterized interconnecting proteins in subcluster-16 will then be selected for biochemical validation in the subsequent Chapter.

Chapter 3
Experimental *in vitro* validation – the RNAi screen

3.1 Introduction

With the exponential increase in genotypic data generated over the last decade, there has been a great demand of a quick, scalable and efficient way of analysing the function of sequenced genes.

Traditionally, different loss-of-function or gene knockout systems have been developed for these studies of individual genes in different model systems. In *Drosophila* for example, these can be generated using defined chromosomal deletions, p-element mediated mutagenesis or mutagenesis through the use of chemical reagents (Spradling et al. 1999). But these methods have been very slow, cumbersome, expensive, and therefore, not easily scalable. Other methods have been developed lately in functional genomics, which are faster, cheaper and easily scalable for the study of gene function. One example is RNAi, or RNA interference technology, used in this study for advantages that are explained in the following sections.

In this Chapter, I analyse the subcluster obtained from our top 100 hits in Chapter 2. I first look into extant knowledge available about members of this cluster followed by an introduction to RNAi and *Drosophila* S2 cells. Then I

describe the RNAi-based functional screen conducted in order to validate the function of our predicted mitotic MAPs. *Drosophila* S2 cells were used as the *in vitro* system for studying the function of these genes.

3.1.1 Subcluster-16 - analysis of extant data

Subcluster-16 is comprised of 33 proteins, which fall into three classes. The predominant class is the predicted set, which have been ranked by the MAP model as mitotic proteins. The next class contains proteins, which are members of the training set used to fit the prediction model. All of these belong to the positive training set, which were known mitotic proteins based on their gene ontology (GO) annotation. The final class is the group of proteins that do not belong to our MAP interactome, but are proteins from the *Drosophila* interactome that cluster with members of subcluster-16. Out of the 33 proteins, 19 are predicted mitotic proteins, 7 belong to the positive training set and 7 come from outside the MAP interactome (Figure 3.1).

Twenty-three out of the thirty-three proteins have human homologs attributed to them on the FlyBase database (Table 3.1). Analysing the gene ontologies of the human homologs provides a picture of the biological processes they are currently known to be involved in and their exact molecular function. Conducting an enrichment analysis of the Human homologs shows mitotic GO terms for Biological Process as highly enriched in these 23 proteins (top 9 are all mitotic). The molecular function annotation shows protein kinase activity, protein binding activity and transcription activation activity as highly enriched. In terms of cellular component ontologies, no significant enrichment can be seen.

The top hits include SWI/SNF complex, small ribonucleoprotein complex, spindle and chromatin remodelling complex (Figure 3.2A). Some mitotic terms are retained in a similar analysis of homologs for proteins excluding the training set (known mitotic proteins) along with histone acetylation and some metabolic processes. Molecular function annotations are predominantly composed of regulatory activities (Figure 3.2B).

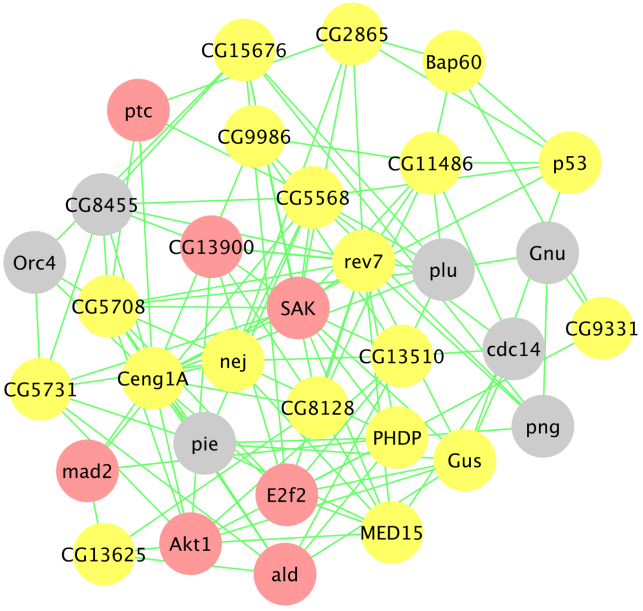
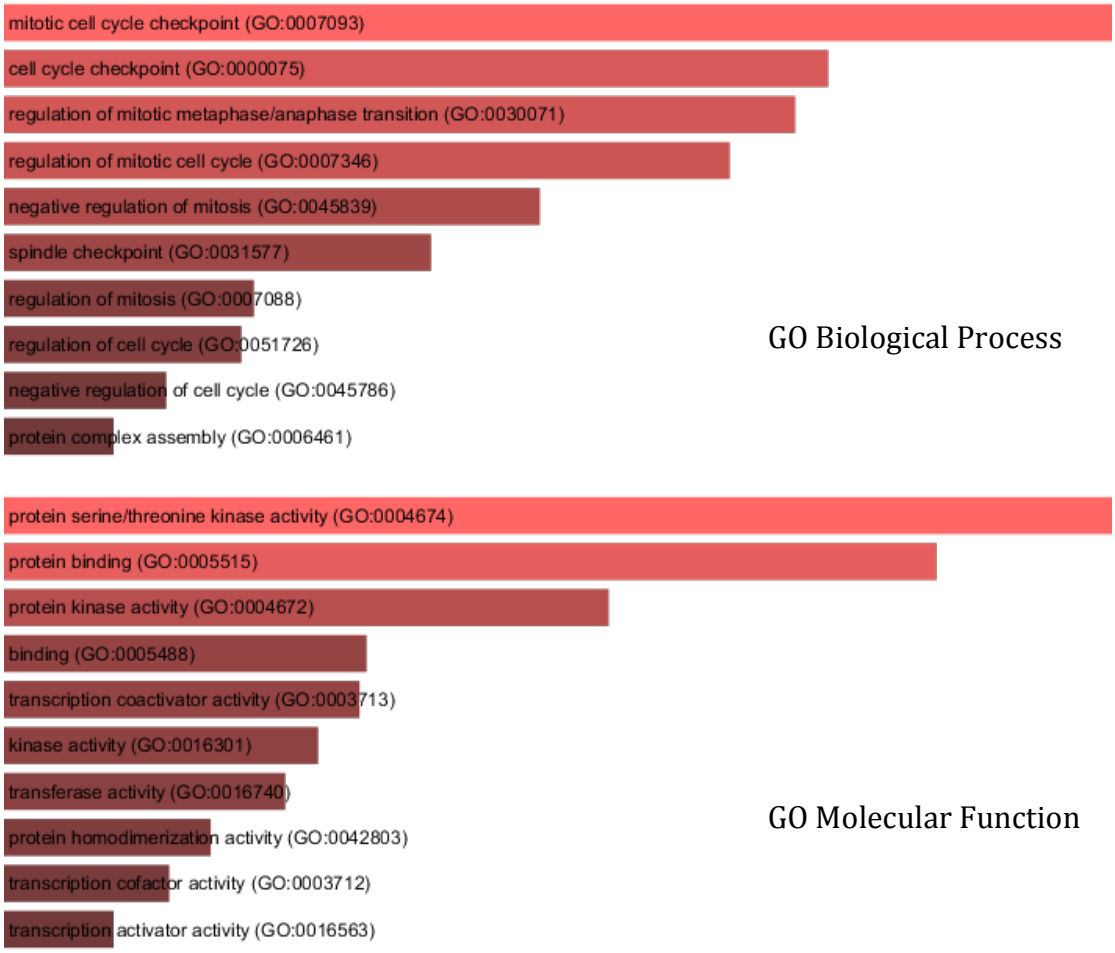


Figure 3.1: Subcluster-16 with its member proteins colour-coded according to their type, i.e. proteins from the predicted set of mitotic proteins (yellow), the training set (red) or proteins from outside the MAP interactome that tightly bind the subcluster (grey).


mitotic cell cycle checkpoint (GO:0007093)
cell cycle checkpoint (GO:0000075)
regulation of mitotic metaphase/anaphase transition (GO:0030071)
regulation of mitotic cell cycle (GO:0007346)
negative regulation of mitosis (GO:0045839)
spindle checkpoint (GO:0031577)
regulation of mitosis (GO:0007088)
regulation of cell cycle (GO:0051726)
negative regulation of cell cycle (GO:0045786)
protein complex assembly (GO:0006461)
GO Biological Process
protein serine/threonine kinase activity (GO:0004674)
protein binding (GO:0005515)
protein kinase activity (GO:0004672)
binding (GO:0005488)
transcription coactivator activity (GO:0003713)
kinase activity (GO:0016301)
transferase activity (GO:0016740)
protein homodimerization activity (GO:0042803)
transcription cofactor activity (GO:0003712)
transcription activator activity (GO:0016563)
GO Molecular Function
SWI/SNF complex (GO:0016514)
small nuclear ribonucleoprotein complex (GO:0030532)
spindle (GO:0005819)
chromatin remodeling complex (GO:0016585)
GO Cellular Component

(A)

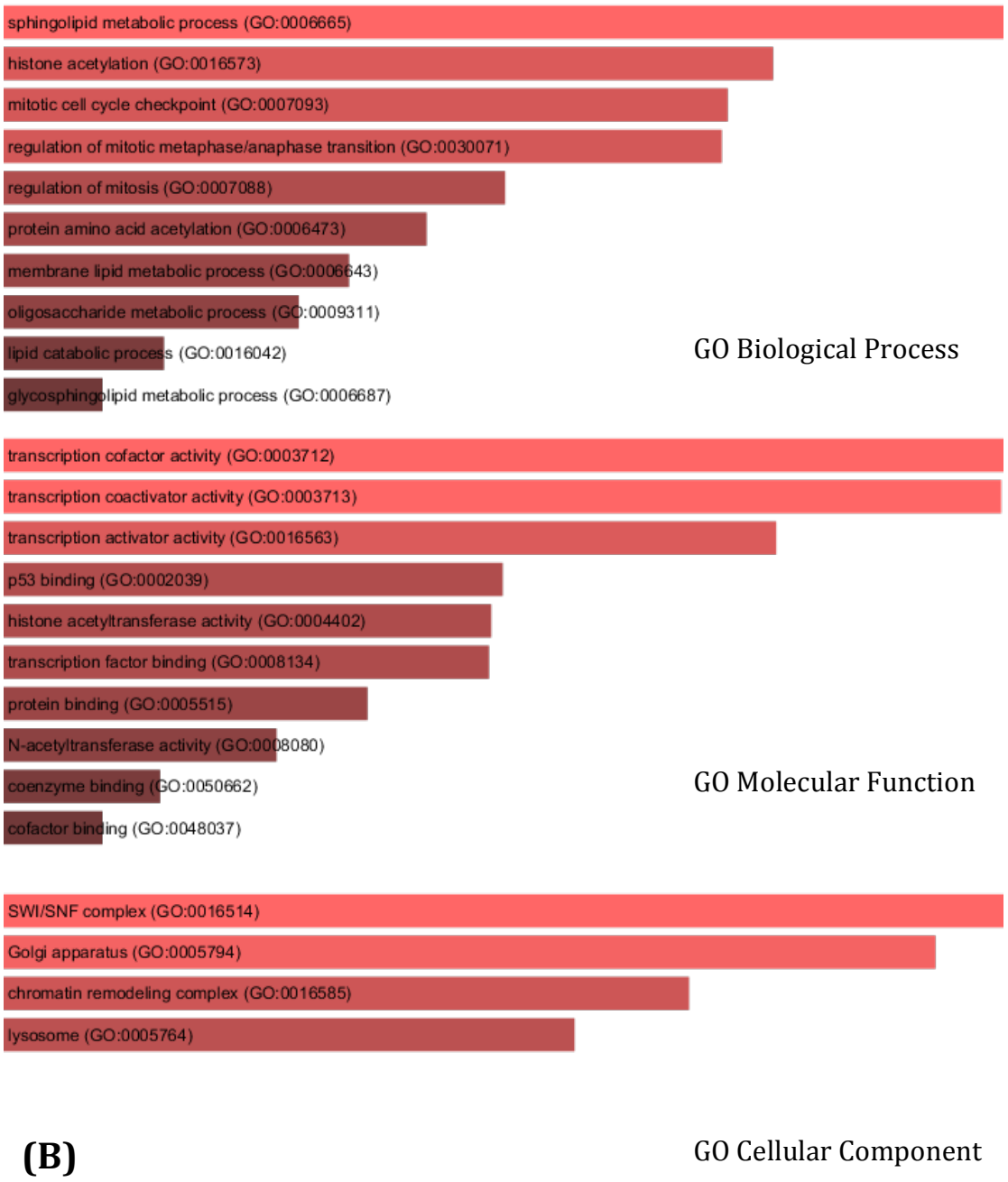


Figure 3.2: Enrichment analysis of GO annotations for the Human homologs available for (A) all 23 homolog proteins, and (B) homologs of proteins excluding the training set. The bars represent the top ten of enriched terms in the order of their statistical significance ($p<0.05$), except for the GO Cellular Component annotations, which only have four terms. Enrichr (Chen et al. 2013) was used to conduct the analysis.

Gene	FB number	Human	Uniprot ID	Gene Name
png	FBgn0000826	-	-	-
PHDP	FBgn0025334	PHX2A	O14813	Paired mesoderm homeobox protein 2A
mad2	FBgn0035640	MAD2L1	Q13257	Mitotic spindle assembly checkpoint protein MAD2A
E2f2	FBgn0024371	-	-	-
CG9986	FBgn0039589	C12orf4	Q9NQ89	Uncharacterized protein C12orf4
CG13900	FBgn0035162	SF3B3	Q15393	Splicing factor 3B subunit 3
CG13510	FBgn0034758	-	-	-
CG11486	FBgn0035397	PAN3	Q58A45	PAB-dependent poly(A)-specific ribonuclease subunit 3
ald	FBgn0000063	TTK	P33981	Dual specificity protein kinase TTK
plu	FBgn0003114	-	-	-
pie	FBgn0005683	-	-	-
MED15	FBgn0027592	MED15	Q96RN5	Mediator of RNA polymerase II transcription subunit 15
CG9331	FBgn0032889	GRHPR	Q9UBQ7	Glyoxylate reductase/hydroxypyruvate reductase
CG5708	FBgn0032196	LMO4	P61968	LIM domain only protein 4 (LMO-4)
CG13625	FBgn0039210	BUD13	Q9BRD0	BUD13 homolog
CenG1A	FBgn0028509	AGAP3	Q96P47	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3 (Centaurin-gamma-like family
cdc14	FBgn0031952	CDC14A	Q9UNH5	Dual specificity protein phosphatase CDC14A
ptc	FBgn0003892	PTCH2	Q9Y6C5	Protein patched homologue 2
nej	FBgn0261617	CREBBP	Q92793	CREB-binding protein
SAK	FBgn0026371	PLK4	O00444	Serine/threonine-protein kinase PLK4
p53	FBgn0039044	-	-	-
gnu	FBgn0001120	-	-	-
CG5731	FBgn0032192	GLA	P06280	Alpha-galactosidase A
CG2865	FBgn0023526	-	-	-
CG15676	FBgn0034651	-	-	-
Bap60	FBgn0025463	SMARCD2	Q92925	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2
rev7	FBgn0037345	MAD2L2	Q9UI95	Mitotic spindle assembly checkpoint protein MAD2B (Mitotic arrest deficient 2-like protein 2)
Orc4	FBgn0023181	ORC4	O43929	Origin recognition complex subunit 4
gus	FBgn0026238	SPSB1	Q96BD6	SPRY domain-containing SOCS box protein 1 (SSB-1)
CG8455	FBgn0031997	MPPE1	Q53F39	Metallophosphoesterase 1
CG8128	FBgn0030668	NUDT6	P53370	Nucleoside diphosphate-linked moiety X motif 6
CG5568	FBgn0035641			
Akt1	FBgn0010379	AKT2	P31751	RAC-beta serine/threonine-protein kinase

Table 3.1: All 33 genes from the subcluster and their human homologs.

3.1.2 RNAi – a brief introduction

Ribonucleic Acid (RNA) molecules have traditionally been known for their messenger, catalytic and structural role only, until the landmark experiment by Fire and Mello (Fire et al. 1998), which established a regulatory role of mRNA. Several previously reported studies laid the foundation for this experiment.

The earliest experiment was conducted by Jorgensen *et al* (Napoli et al. 1990), when they attempted to produce Petunias with a darker violet colour, by over expressing the gene encoding Anthocyanin. This produced the unintuitive result of a 50 times decrease in expression, a phenomenon they termed co-suppression.

In a separate experiment, Guo *et al* (Guo & Kemphues 1995), attempted to knockdown the *par-I* gene in the nematode, *C. elegans*, by treating with the antisense strand of its RNA (hence termed, antisense RNA technology). The results were not just a knockdown in the experiments treated with antisense strands, but also the control experiments, which were treated with the sense strand of the RNA. Finally, in an experiment using the fungus, *N. crassa*, was conducted in an attempt to enhance its orange colour by expressing multiple copies of the gene responsible (Romano & Macino 1992). The result was the suppression of the orange colour in one-third of the mould, a property referred to at that time as 'quelling'.

Fire *et al*, in their experiment, treated *C. elegans* with the unc22 gene RNA (Fire et al. 1998). The treatment included the sense, antisense and the double strands in separate experiments. The double-strand treatment produced a 10 times greater suppression of the gene compared to the other two experiments, establishing the dsRNA-mediated silencing of gene expression for the first time.

In the same year another group (Kennerdell & Carthew 1998) demonstrated a similar process in *Drosophila*, demonstrating the process was not restricted to certain taxa only. This well conserved post-transcriptional double-stranded RNA-mediated regulation of gene expression is called RNA interference, or RNAi.

3.1.3 RNAi in *Drosophila*

RNAi in *Drosophila* is a two-step process. First is the initiation phase whereby the dsRNA is cleaved into a population of different smaller RNA products, termed small interfering RNA or siRNA molecules. This step is carried out by the enzyme called Dicer (Tijsterman & Plasterk 2004). These siRNA species have a 19bp duplex region with 2 nucleotide overhangs on each 3'-end.

In the second 'effector' phase, the siRNA molecule undergoes unwinding and one strand is selectively integrated into the protein complex, RISC or the RNA-induced silencing complex. Here the siRNA programmes the complex and acts as the specificity co-factor, while the protein complex contains proteins which carryout the catalytic activity. The siRNA finds matching target mRNA molecules and stably binds them which allows RISC to undergo the cleavage of the target molecule. The complex is then free to interact with other mRNA molecule (Figure 3.3). This two-step process was first discovered in *Drosophila* and later found to be conserved in most animals including humans (Tuschl et al. 1999).

With the discovery of RNAi and its potential for use in functional genomics, fruit flies proved to be an efficient model system. Several experiments in the labs of Hannon and Dixon demonstrated that *Drosophila* tissue culture cells could

readily take up long dsRNA molecules and suppress the expression of target proteins (Hammond et al. 2000; Clemens et al. 2000). This not only bypassed the challenges of dsRNA or siRNA delivery into cells, but also reduced the cost of these experiments. This sparked the generation of multiple genome-wide cDNA libraries for large-scale RNAi screens by direct treatment of cultured cells,

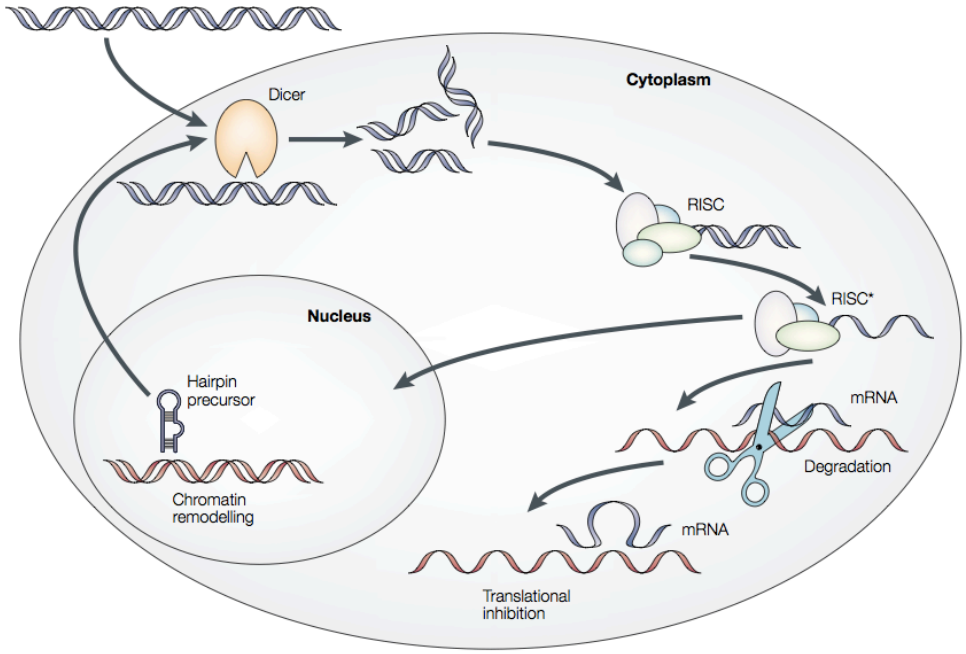


Figure 3.3: A schematic representation of the two-step RNAi process in *Drosophila*. Figure taken from (Wasi 2003).

generating results in a matter of days. In the next few years the first genome-wide screens were published followed by dozens of large-scale screens studying different biological processes (Mohr et al. 2010).

There are several other reasons, which enable *Drosophila* to stand out amongst other model organisms for use in RNAi screen.

- **Ease and simplicity**: there are well-established protocols available with very few steps, which allow the amplification of long dsRNAs through *in vitro* transcription. No trypsinization required as cells grow in monolayers that are easily detached by mild agitation.
- **Low-cost**: Lower cost compared to mammalian systems – no transfection reagents needed, wide pH range for growth, no CO_2 required for culture, grows at a reasonable range of 25-27^{o}C.
- **Effective**: Multiple different 21-mer species of siRNAs are produced by the endogenous Dicer ensuring suppression (unlike the single siRNA molecule used in mammalian cell lines). Treatment of *Drosophila* cells does not trigger any inflammatory responses like human cells.
- **Ideal for studying human genes**: Fly genome is well conserved (functions can be taken across species) and is three-fold less redundant than humans (no masking of knockdown phenotypes by duplicate genes). 75% human diseases genes have fly homologs (Reiter et al. 2001).

3.1.4 *Drosophila* S2 cell lines

Several *Drosophila* cell lines have been maintained, which have been derived from various tissues and developmental stages of the organism (DGRC 2013). For the experimental *in vitro* analysis of subcluster-16, I used the *Drosophila* S2 cell line, which is one of the most well characterized and widely used cell lines,

both in focused experiments and genome-wide studies. The S2 cell line was derived from late-stage embryos, one of a few different types of immortalized cell lines by Imogene Schneider (Schneider 1972). The gene expression patterns and cellular behaviour later indicated that these cells were derived from hemocytes with macrophage-like phagocytic properties (Schneider 1972). This property of S2 (and several other *Drosophila*) cell lines enables them to readily take up long dsRNA fragments making them highly susceptible to RNAi (Hammond et al. 2000; Clemens et al. 2000). Although there is variability in knockdown and cell viability, this property is a key advantage of the *Drosophila* system over mammalian systems as described in detail in the previous section. S2 cell lines have two subtypes based on different sets of membrane receptors and therefore varying strengths of adherence (the original S2 cell line has weaker adherence to substrates as compared to the S2R+ subtype). Although S2 cells are roughly spherical in shape, they can actively spread on surfaces, for example slides coated with the lectin, concanavalin A (ConA), which helps in high resolution microscopy (Rogers & Rogers 2008).

Loss-of-function studies are one of the most informative experimental strategies to study the function of genes. In this Chapter, I conducted an RNA interference screen and validated the mitotic function of several members of subcluster-16.

3.2 Materials and Methods

3.2.1 Amplifying genes

The genes for each protein were amplified from cDNA templates procured from the BDGP Gold clone cDNA collection at the *Drosophila* Genome Resource Centre at Indiana University (Bloomington, US). Annealing temperatures were optimized using PCR reactions with Taq polymerase (New England Biolabs). Final products were then amplified using high-fidelity Phusion polymerase (New England Biolabs) at the optimized annealing temperature.

3.2.2 Design of dsRNA

The PCR primers were designed using the web-based tool, SnapDragon (Kulkarni et al. 2006), which enforces a strict base-pair identity threshold to limit off-target effects (Ma et al. 2006; Kulkarni et al. 2006) and is recommended by the *Drosophila* RNAi screening centre. The threshold of predicted off-targets was set to the recommended 19bp, with a product size limited between 250-500bp. The number of primer pairs to call was set to 1. The full list of primer pairs is given in Appendix III. For each pair, the T7 promoter sequence. Each primer (TAATACGACTCACTATAGGG) was added to the 5' end of both the primers.

3.2.3 Production of dsRNA

Double stranded RNA was produced via i*n vitro* transcription of each cDNA product using the Ambion® MEGAscript® T7 kit (Life Technologies). The protocol provided by the manufacturer was followed for a 10-16 hour reaction.

This followed by treatment with the TURBO DNase provided with the kit. The concentration of dsRNA was quantified by running a 1/10 dilution on a 1% agarose gel.

3.2.4 Cultivating S2 cells

The S2 cell line was passaged in Schneider's *Drosophila* S2 Medium (Lonza) with 10% Fetal Bovine Serum (Gibco) and no anti-fungal agents. Cells were kept at 25^{o}C in volumes of 10ml, passaged (1:5) into fresh medium every 4 days.

3.2.5 dsRNA transfection

Cells were distributed at a concentration of 1.0x10^{6} in standard six-well plates in 2ml of serum-free S2 medium. 15 µg of dsRNA was added per reaction directly into the cells and incubated at 25^{o}C for 3-4 hours allowing for the uptake by cells. 1ml of 30% fetal bovine serum was added (10% final concentration), the plate was sealed with parafilm and incubated for 5 days.

3.2.6 Preparing ConA coverslips

Standard 20x20mm coverslips were immersed in concentrated hydrochloric acid for 2 hours and then washed and air-dried. The coverslips were then sprayed with ethanol and air-dried. After placing them on Parafilm on a flat surface in the laminar air-flow hood, 20µl of 0.5mM concanavalin A was pipetted onto the surface of the coverslip and spread over using a clean 100µl tip. They were left to air-dry before being stored in a suitable box at 25^{o}C and used within a week.

3.2.7 Fixing of RNAi-treated S2 cells

After 5 days of dsRNA treatment, the cells were agitated by pipetting (mild enough not to create bubbles). The concentration was measured using a hemocytometer. ConA slips were placed at the bottom of each well (with the treated side facing up) in a new 6-well plate and $1.0x10^6$ cells were pipetted onto each coverslip. The 6-well plate was then covered with the lid and placed in the hood for 2 hours for the cells to settle and spread on the conA surface. The medium was discarded and the cells washed (outside the laminar hood) using Dulbecco's PBS. 3ml of pre-chilled (-20°C) 100% methanol was then poured directly from above each coverslip and the plate was incubated at -20°C (on dry ice, or in the -20°C freezer) for 20 minutes. This was followed by 20 minute warm-up incubation at room temperature before the methanol was removed and the coverslips with fixed S2 cells were washed to remove any remaining methanol.

3.2.8 Immunofluorescence

The coverslips were prepared for antibody treatment by incubating in 0.5ml block solution (4% bovine serum albumin in PBS with 0.1% Triton) for 30 minutes. The block was removed and the coverslips were dried using a 100µl tip and an aspirator. 100µl of primary antibodies were applied to each coverslip i.e. mouse DM1a anti-tubulin antibody (Sigma) and rabbit anti-Asl antibody (gift from Jordan Raff) (1:1000 in block solution). The cells were incubated overnight at 4°C. The primary antibody removed, and the cells were washed three times (10 minutes each on a rocker) with 1ml of PBST solution. The wash was removed

with an aspirator and a 100µl of secondary antibody solution was pipetted on top of each coverslip, i.e. Alexa Fluor® 488 anti-mouse and Alexa Fluor® 555 anti-rabbit secondary antibodies (1:1000 in blocking solution). The secondary antibody removed, and the cells were washed three times (10 minutes each on a rocker) with 1ml of PBST solution.

3.2.9 Mounting the cells for microscopy

The VectorShield® mounting medium (Vector Labs) was used which contains DAPI (4'6-dimamidino-2-phenylindole) as the stain for DNA and glycerol for mounting. A 5µl drop of VectorShield was placed in the middle of standard slide and the coverslip was inverted on top of it (fixed cells facing the solution) with care to avoid trapping air bubbles. The slide was now ready for inverted microscopy using an oil immersion lens. DAPI produces a blue fluorescence when bound to DNA with an excitation at 360nm and emission at 460nm.

3.2.10 Microscopy and analysis

The cells were imaged under oil at 25°C, using a Nikon Eclipse TE2000-U inverted microscope with a Nikon Plan APO CV 60X 1.4N/A objective and a Hamamatsu c8484-056 camera. The IPLab software was used to capture and save the images from each channel. Images were analysed and processed using NCBI's open-source ImageJ software (Schneider et al. 2012).

3.3 Results and Discussion

3.3.1 Designing and optimizing the RNAi screen

This RNAi-based knockdown screen was devised to study the morphology of the mitotic spindle, centrosome and mitotic chromosomes.

I produced an array of double-stranded RNA (dsRNA) molecules using *in vitro* transcription for all thirty-three of our test genes. These dsRNAs were amplified using a strict base-pair identity threshold to minimize off-target effects.

Cells were passaged regularly to maintain proper confluency and any flasks with poor growth were discarded. dsRNA aliquots were stored in sterile tubes to avoid any contamination and minimize the chances of infection in wells with poor growth or any sign of infection were discarded.

RNAi protocols have described condensation of culture media in wells around the edges of plates due to evaporation (Goshima 2010). To avoid these effects each plate was sealed with Parafilm to minimize any evaporation.

3.3.2 Results of the RNAi screen

RNAi-treated cells for all genes were fixed and processed for immunofluorescence microscopy using antibodies. The standard DM1a anti-tubulin antibody was used to stain the microtubules. A mouse antibody against the Asl protein (a gift from J Raff) was used to stain the centrioles. DAPI in the mounting medium was used to stain the DNA/chromosomes in the fixed cells.

Each well of RNAi-treated cells was analysed manually using fluorescence microscopy. Each gene was analysed by two independent RNAi experiments with an average of 60 spindles each, followed by blind analysis (the slides not labelled with names). Each spindle was analysed and imaged through three channels (green, red and blue) for spindle, chromosome and centrosome fluorescence. dsRNA against the bacterial Beta-lactamase (Blac) gene was used as a negative control. SAK, a kinase with an established role in centrosome duplication and a characteristic monoastral bipolar spindle (which is also part of our subcluster) was used as the positive control.

For spindle phenotypes, the proportion of abnormal spindles, i.e. any deviation from the normal diamond-shaped spindle (related to shape, length and tubulin intensity), was quantified. For centrosomes, the number of Asl dots per mitotic cell was the primary readout. And for chromosomes, the presence of mitotic cells with unaligned metaphase chromosomes was quantified.

After comparing the percentage of abnormal spindles, chromosomes and centrosome numbers to the control experiments, the statistical significance was measured using a two-tailed t-test ($p<0.05$).

Based on statistical significance, 26 out of the 33 test genes (79%) produced mutant phenotypes (Appendix IV). 9 genes produced spindle defects, 8 genes produced centrosome number defects, 9 genes produced spindle and centrosome defects while 7 genes did not produce any significant defect upon RNAi treatment (Figure 3.4, Appendix V).

The mutant phenotypes were further scored for chromosome phenotypes (the proportions of mitotic cells with unaligned metaphase chromosomes), though the analysis was based on trends (values were not statistically significant). Fifteen genes produced strong abnormal phenotypes (4 with 70% or more, and 11 with 50% or more) compared to the rest, which had weak phenotypes or no phenotypes at all.

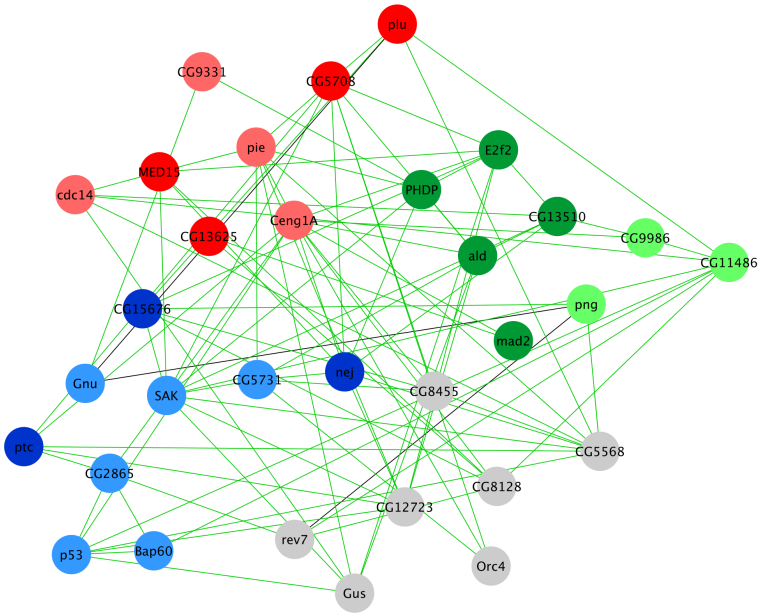


Figure 3.4: The subcluster with nodes color-coded according to the phenotypic class, i.e. Class I with spindle defects (green), Class II with centrosome defects (red), Class III with spindle and centrosome defects (blue) and Class IV with no mutant phenotypes (grey). The darker shade in each class represents genes that have stronger chromosome defects.

The genes, whose knockdown produced spindle phenotypes, were further classified into different classes based on, their spindle length (measured using ImageJ (Schneider et al. 2012)) and the type of mutant phenotype based on

visual analysis i.e. *mashed*, *spiky*, *dim* or *moustache* (based on trends, not statistically significant) (Appendix VI).

In our analysis, all genes have been compared with the genome-wide RNAi screen published by the Vale Lab (Goshima et al. 2007b). Only two genes from our subcluster have been reported as hits in the Vale screen, but the database still has valuable data publicly available which includes computer-generated image galleries of spindles, and details of whether any gene has been reported as a hit in the primary (manual and computer) screens. The MitoCheck database which stores data from a genome-wide mitotic RNAi screen in human cells (Neumann et al. 2010) has been analysed for genes that have human homologs. Gene Ontology (GO) annotations have been gathered and analysed from FlyBase (Consortium 2003). All three of these data sources have been used for comparison with the results of the RNAi screen in this Chapter (Appendix VII). A complete table of the raw results from the RNAi screen and the averages with corresponding statistical analysis are given in Appendices I and II.

3.3.3 Class I: Genes with spindle defects

From the total of thirty-three genes, nine genes produced statistically significant proportions of abnormal spindle phenotypes in the screen when compared to the control experiment (Figure 3.5A). No significant abnormality was observed in the number of mitotic cells with abnormal centrosome counts. Figure 3.5B shows the distribution of centrosome numbers in mitotic cells, both in control and RNAi experiments.

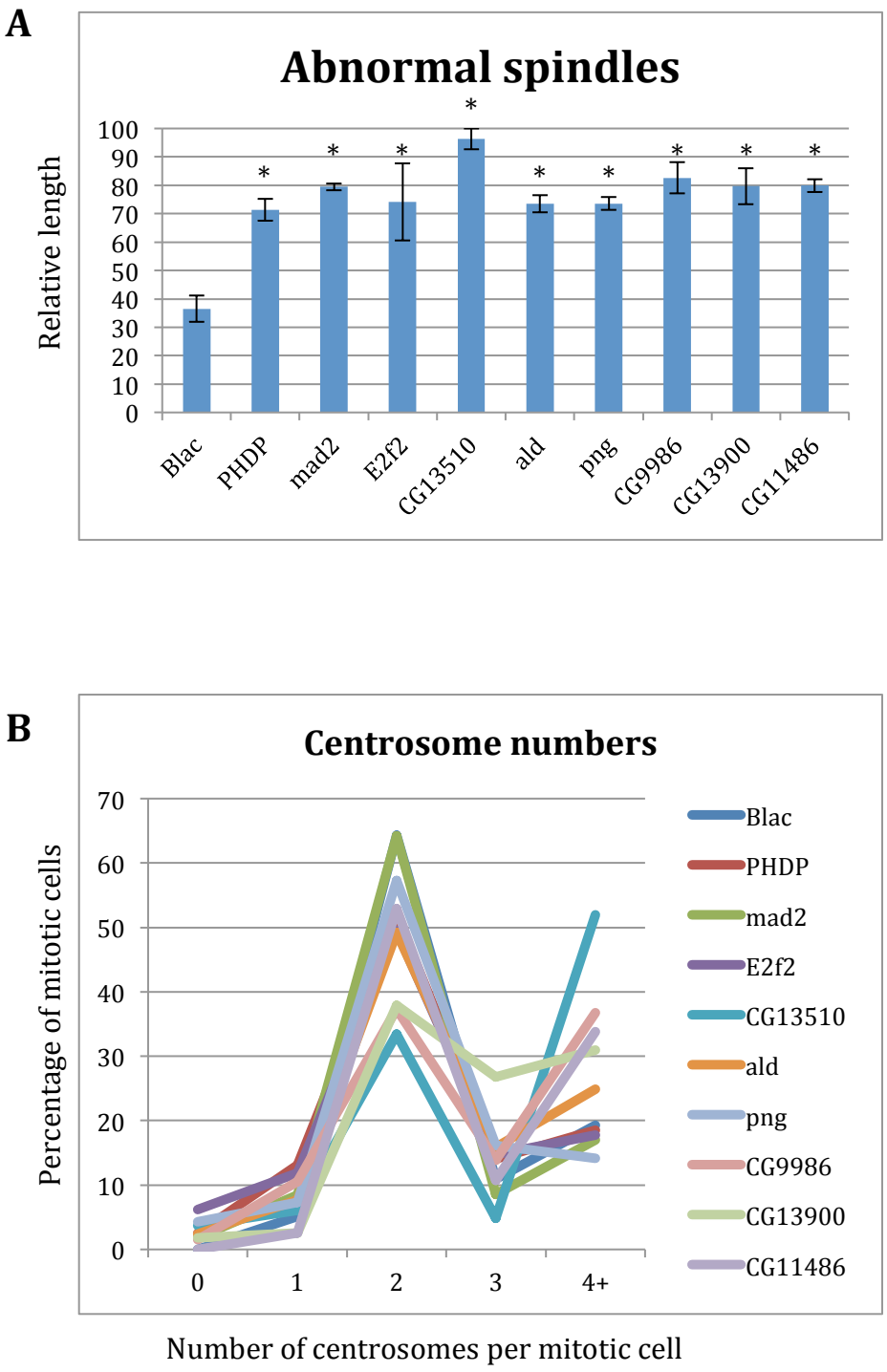
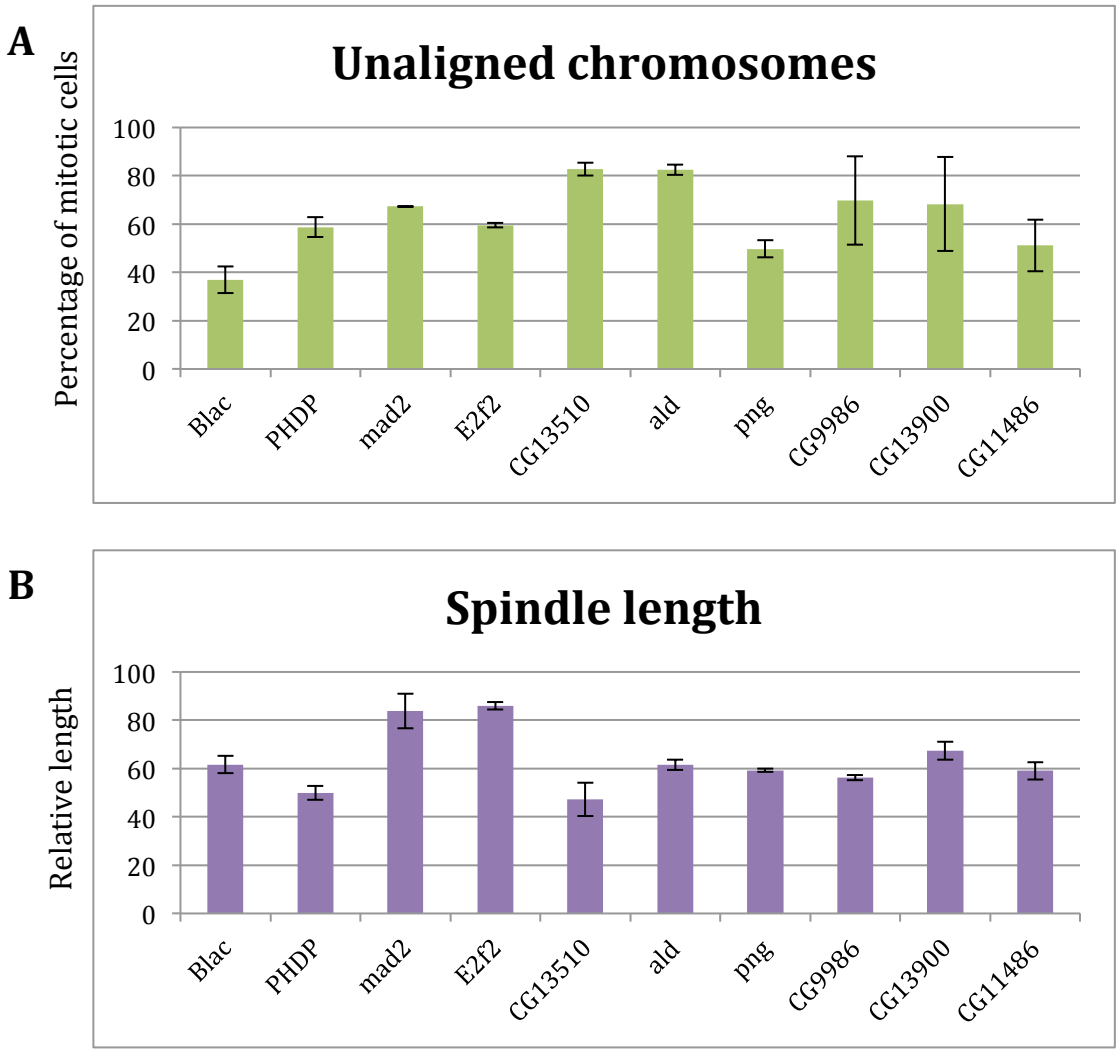


Figure 3.5: Spindle and centrosome defects for members of Class I. All spindle phenotypes are significantly different (paired t-test, $p<0.05$) from the control (Blac) experiment. The proportions of abnormal centrosome numbers per mitotic cell are not significant in this class. (A) Bar chart showing the proportions of abnormal spindle phenotypes in the percentage of mitotic cells analysed. Each bar represents the average of two independent RNAi experiments (±SEM). (B) Line plots showing the centrosome number distribution as percentages of mitotic cells in each RNAi experiment. Each point is the average of two independent



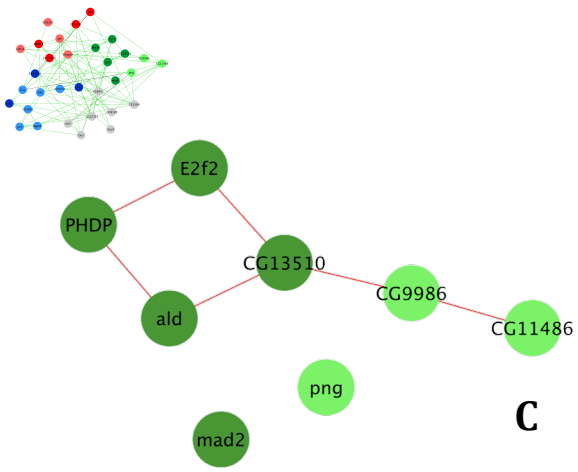


Figure 3.6: (A) Proportions of unaligned chromosomes, with each bar the average of two independent experiments (±SEM). (B) Average spindle length from two independent RNAi experiments for each gene (±SEM). (C) The connectivity of Class I proteins in the subcluster. Darker coloured nodes represent genes with stronger chromosome alignment defects.

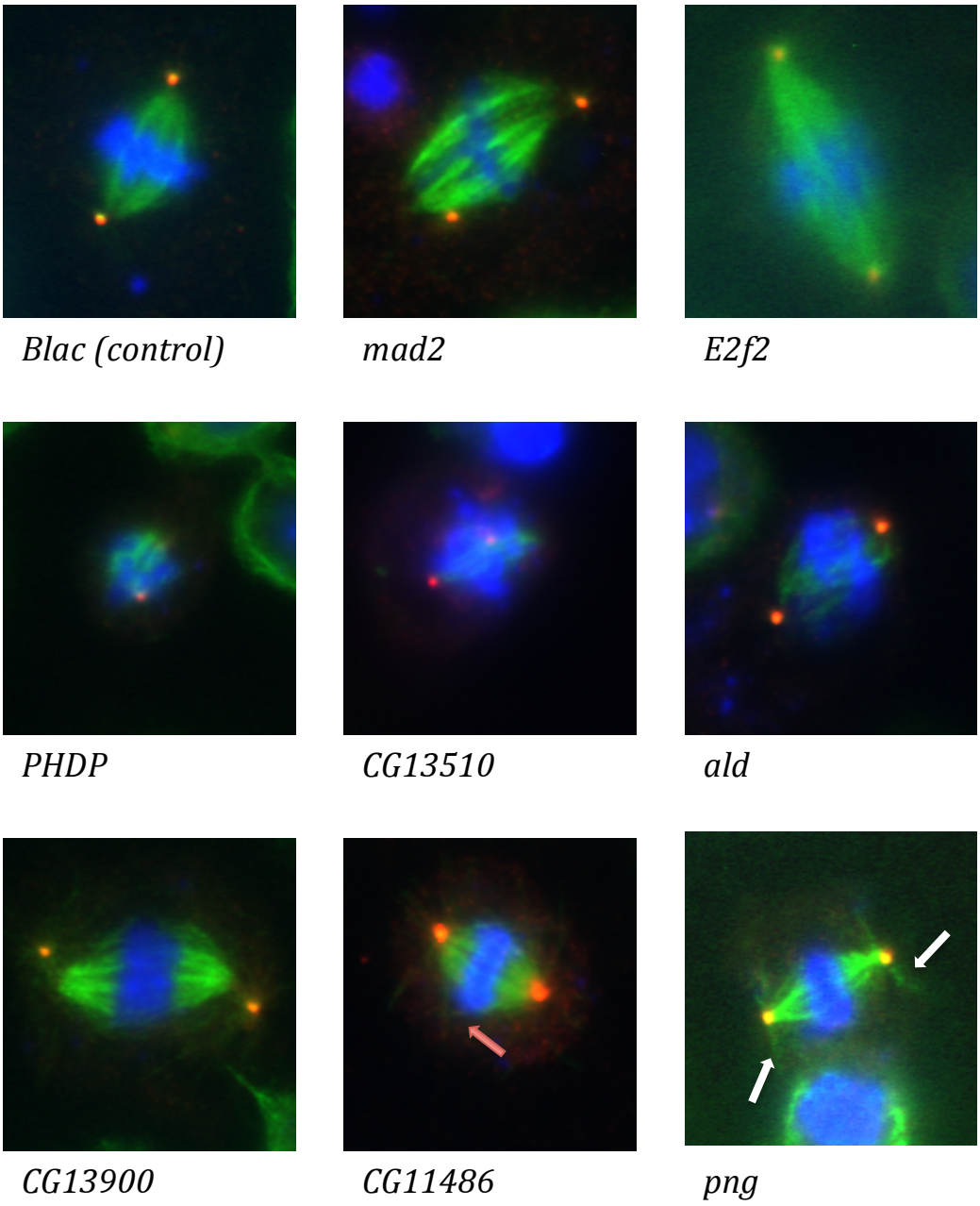


Figure 3.7: Representative images from each RNAi experiment of Class I. Each image is a merge of the three channels, showing tubulin/spindle (green), DNA (blue), and centrosomes (red). Blac is the negative control experiment with producing a wild-type diamond-shaped bipolar spindle. *CG11486* shows a characteristic spikey phenotype (red arrow) while *png* has the moustache spindle phenotype with bundles of astral MTs (white arrows).

Out of these nine genes, five showed stronger chromosome alignment defects (*mad2*, *e2f2*, *PHDP*, *CG13510* and *ald*) in metaphase compared to the remaining four (*CG13900*, *CG11486*, *png* and *CG9986*) (Figure 3.6A). Figure 3.7 shows representative images for the RNAi experiment of each gene.

These nine proteins can be further classified into sub-classes based on the type of spindle phenotype. Two genes (*CG9986* and *CG11486*) produced *spiky* spindles with microtubule spiking out and distorting the characteristic diamond shape, *ald* showed a *mashed* spindle phenotype and *png* showing a characteristic *moustache* phenotype with bundles of astral MTs on either side of the centrosomes. Based on trends in spindle length (Figure 3.6B), two genes, *CG13510* and *PHDP*, can be observed with *short* spindles while three genes (*mad2*, *e2f2* and *CG13900*) produced *long* spindles compared to spindles in the control experiment.

Analysing the connectivity of this phenotypic class in the subcluster reveals interactions amongst the genes with *long* spindles, the three genes with *spiky* spindles (Figure 3.6C). In the subcluster, *CG13900*, *e2f2*, *mad2* and *ald* belong to the training set, which means they have mitotic and annotations in the Gene Ontology (GO) database and are known to play roles in cell division.

E2f2 is a transcription factor and is annotated with GO terms related to the negative regulation of DNA replication and the transition to S-phase of the cell cycle. The *long* spindle phenotype might be an indication of a secondary regulatory role in mitotic spindle organization.

Mad2 is a well characterized and conserved member of the spindle assembly checkpoint (Buffin et al. 2007) and *ald*, also known as *mps1* is a also a well-conserved cell cycle-related protein kinase (Althoff et al. 2012).

CG13900 has GO annotation related to mitosis and mRNA splicing, which is comparable to its human homolog, *SF3B3*, which is a subunit of splicing factor B3. *CG13900* is the only member of this phenotypic class, which has been reported as a hit in the Vale screen (Goshima et al. 2007c), with a weak 'long spindles' phenotypic. This phenotype is reproduced in our screen as well.

png or *pan gu*, is the only gene which is from outside the MAP interactome. *png* encodes a serine kinase and is the key protein in the PNG complex responsible for the regulation of the development translation of Cyclin B (Vardy & Orr-Weaver 2007) and cyclin A (Vardy et al. 2009) in early embryos. It is only reported as a computer hit in the Vale screen with a 'dim tubulin at the poles' phenotype. A visual analysis of the spindle gallery shows an overall *dim* spindle, consistent with our result.

Previously uncharacterized proteins include *CG13510*, *CG11486*, *CG9986* and *PHDP*. *CG13510* and *CG9986* have no known molecular functions and do not have human homologs. They have not been reported as hits in the Vale screen, but a look at the spindle galleries show *short* spindles and *shredded* chromosomes, respectively. This is consistent with our results.

CG11486 has a kinase-like domain and a very generic GO annotation (protein binding). Its human homolog, *PAN3* is known as a ribonuclease with a role in poly-A tail formation and decapping. No hits are reported in the Vale screen.

PHDP is a small protein with a very low expression in early embryos. It has a molecular function annotation related to transcription factor activity, which is predicted from its human homolog, *PHX2A*, a transcription factor with a DNA-dependent activity and a role in organ development.

3.3.4 Class II: Genes with centrosome defects

This phenotypic cluster contains eight genes with statistically significant proportions of mitotic cells with abnormal number of centrosomes. The distribution of centrosome numbers in mitotic cells is given in Figure 3.8B. *CG5708*, *CG9331* and *CG13625* show the highest proportions of mitotic cells with 4+ centrosomes, compared to the control experiments.

Amongst the genes in this class, *med15* shows a relatively higher proportion of mitotic cells with abnormal spindles, but in comparison to the control does not meet our cut-off for statistical significance ($p>0.05$), and is therefore classified as Class II (Figure 3.8A). Similarly, *CG13625*, *plu* and *CG5708* also present subtle spindle phenotypes that can partially explain the relatively stronger chromosome alignment defects during metaphase compared to the remaining four genes (*CG9331*, *cenG1A*, *cdc14* and *pie*) (Figure 3.9B). Figure 3.10 shows representative images for the RNAi experiment of each gene.

In the subcluster, all genes of this phenotypic class interact with each other in a more or less linear fashion, with the exception of *CG9331* and *cenG1A*, which lie unconnected (Figure 3.9C).

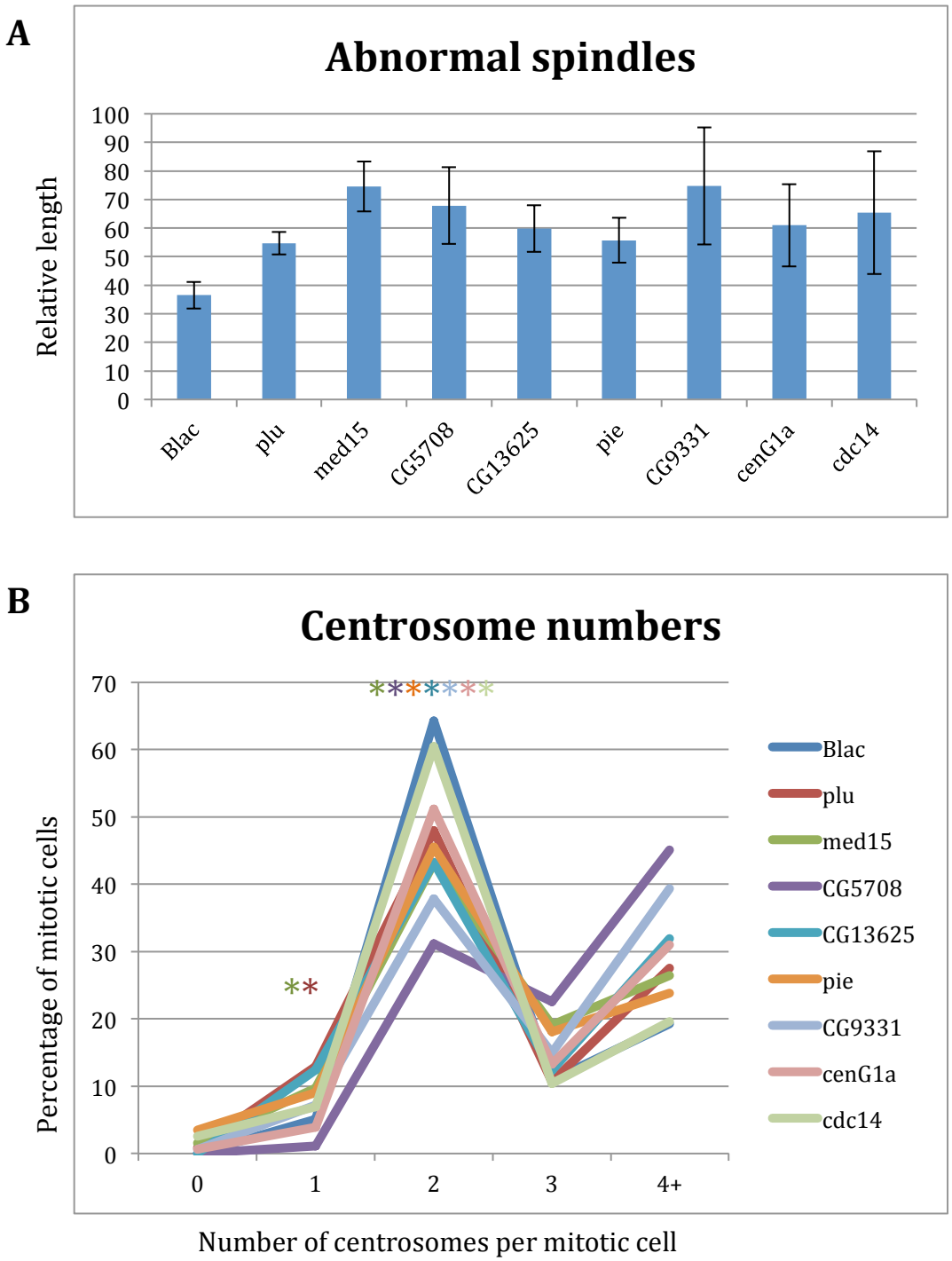


Figure 3.8: Spindle and centrosome defects for members of Class II. All centrosome phenotypes are significantly different (paired t-test, $p<0.05$) from the control (Blac) experiment. The proportions of abnormal spindle phenotypes are not significant in this class. (A) Bar chart showing the proportions of abnormal spindle phenotypes in the percentage of mitotic cells analysed. Each bar represents the average of two independent RNAi experiments (±SEM). (B) Line plots showing the centrosome number distribution as percentages of mitotic cells in each RNAi experiment. Each point is the average of two independent experiments. The control (Blac) has roughly

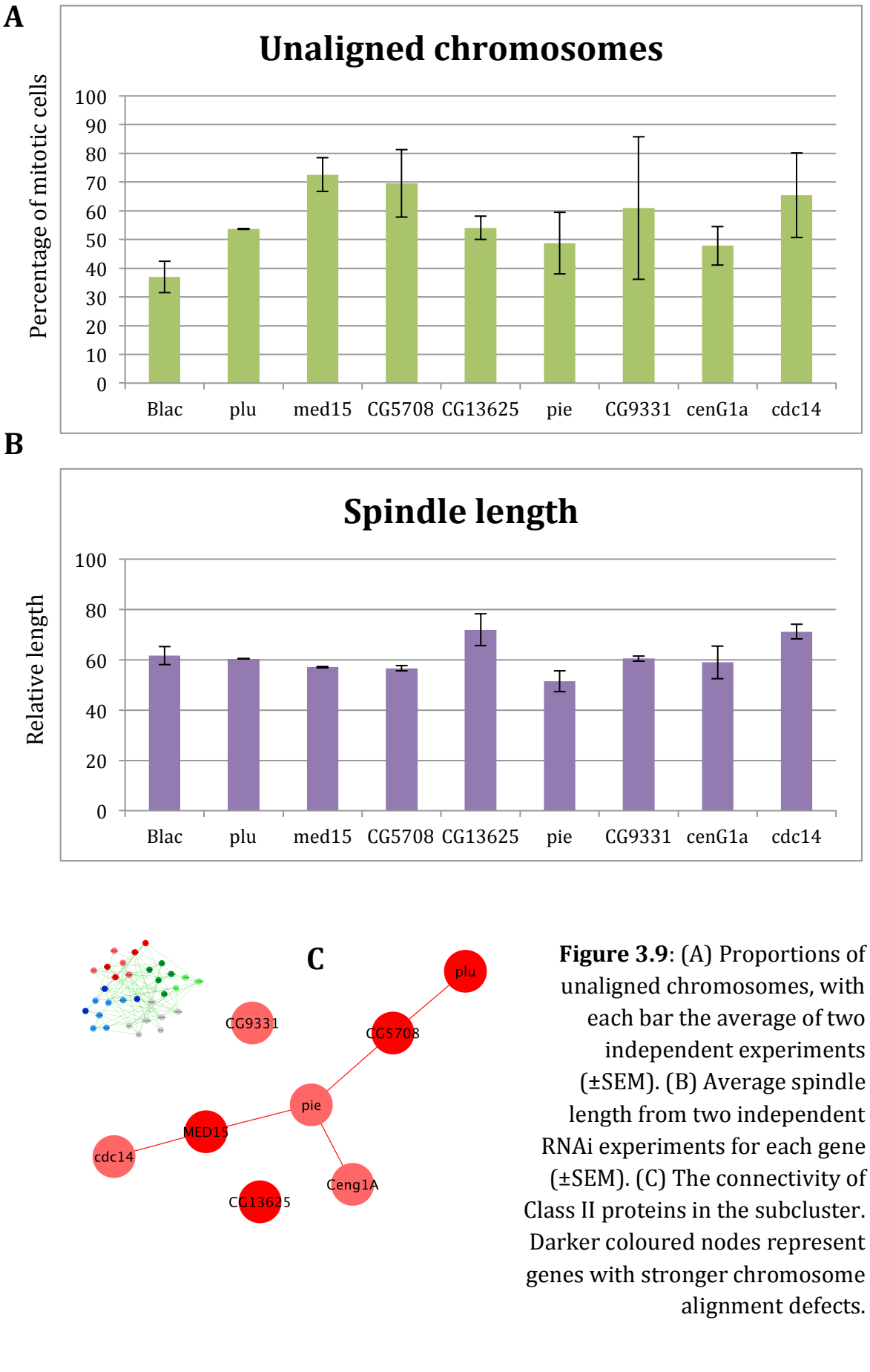


Figure 3.9: (A) Proportions of unaligned chromosomes, with each bar the average of two independent experiments (±SEM). (B) Average spindle length from two independent RNAi experiments for each gene (±SEM). (C) The connectivity of Class II proteins in the subcluster. Darker coloured nodes represent genes with stronger chromosome alignment defects.

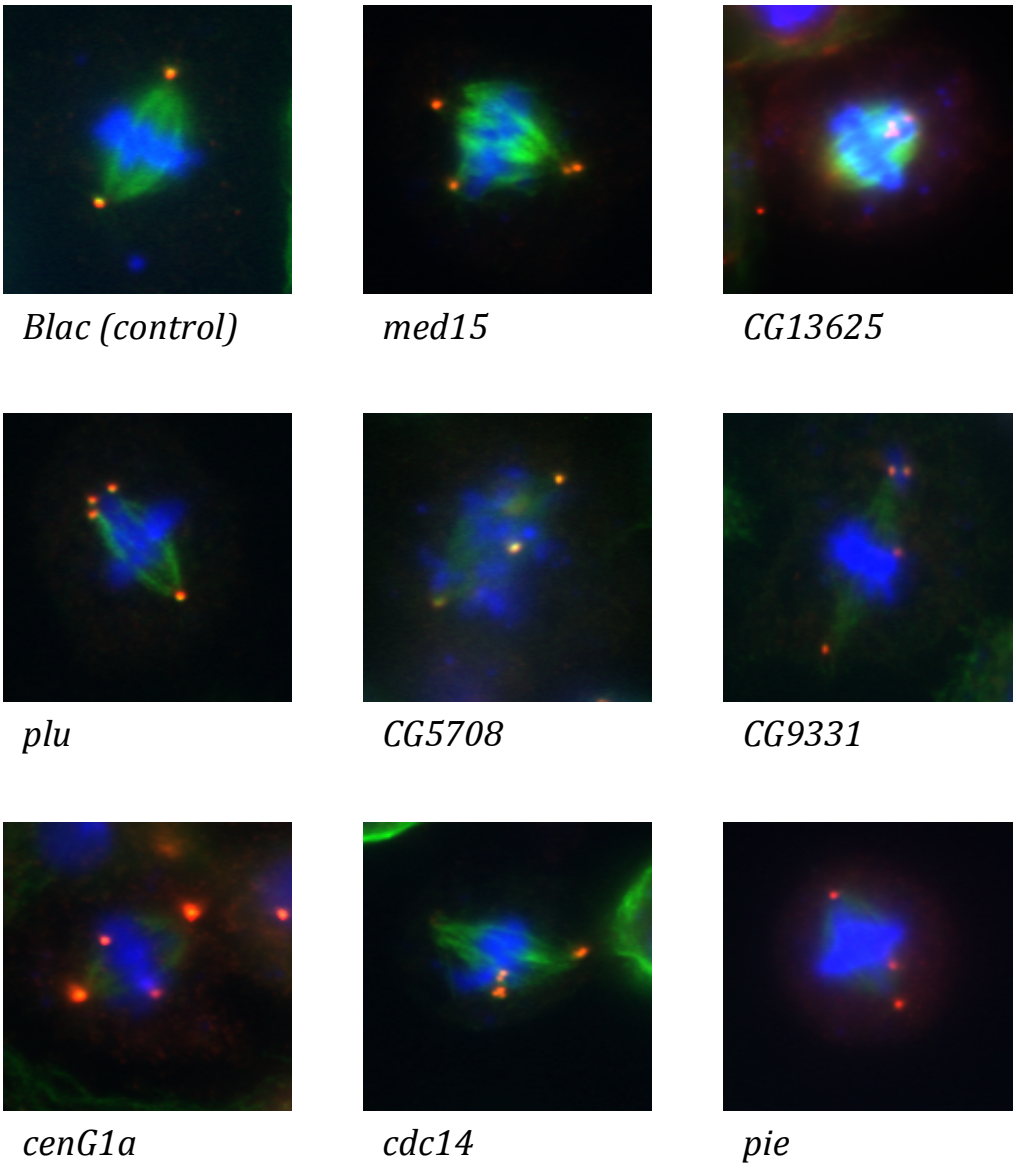


Figure 3.10: Representative images from each RNAi experiment of Class II. Each image is a merge of the three channels, showing tubulin/spindle (green), DNA (blue), and centrosomes (red). Blac is the negative control experiment with producing a wild-type diamond-shaped bipolar spindle.

Most of the genes in this class are predicted mitotic proteins and do not have any known mitotic functions. Three genes, *plu*, *cdc14* and *pie*, come from outside the MAP interactome.

plu is a regulatory subunit of the PNG kinase complex, which is known to have a role in regulating the translation of cyclins in early embryos (Elfring et al. 1997). *plu* is annotated with GO terms related to regulation of DNA replication and DNA-binding function. It has no human homologs.

pie is reported to have Zn-ion binding functions with roles in neurogenesis and compound eye development. It is reported as a manual hit in the Vale screen with a weak 'chromosome misalign' phenotype. A look at the spindle galleries for *pie* show spindles with large/dense centrosomal staining potentially caused by higher number of centrioles. This supports the main phenotype of this class.

cdc14 is a known and conserved phosphatase that has GO annotations related to the regulation of centrosome separation and cytokinesis. It has a computer hit with high nuclear number in the Vale screen and the human homologue has a 'dynamic changes' phenotype in the MitoCheck screen.

The remaining five proteins belong to the predicted set of proteins which have no mitotic function reported.

CG13625 has mRNA splicing annotations and has a human homologue called *BUD13* which is coiled-coil containing protein involved in mRNA transport and splicing. Based on spindle length, *CG13625* appears to have a weak *long* spindle phenotype along the significantly higher centrosome numbers per mitotic cell.

CG5708 is a Zn-binding protein with LIM-type domain and has no biological process annotations. It interacts directly with *pie*, another Zn-binding protein of this class suggesting a common biological process. The human homologue of *CG5708* is *LMO4* or the LIM-domain only protein involved in the transcriptional regulation and has a supernumerary centrosome phenotype reported (Montañez-Wiscovich et al. 2010).

med15, like its human homolog, is a transcription co-factor of RNA polymerase. *cenG1A*, is the Centaurin gamma 1A protein, involved in GTPase-mediated signal transduction.

The Vale screen reports no hits for *CG5708*, *med15* and *cenG1A*, but 'dense' centrosomes can be noted in the image galleries when compared to other RNAi experiments, suggesting the possibility of higher number of centrosomes.

CG9331 also produced a *dim* spindle phenotype. It encodes a NAD-binding oxidoreductase and has several metabolic activities.

3.3.5 Class III: Genes with spindle and centrosome defects

This phenotypic class has nine genes that show both spindle and centrosome abnormalities that present statistically significant abnormalities compared to control experiments (Figure 3.11). The centrosome number distribution is given in Figure 3.11B. Almost all have higher proportions of mitotic cells with 4+ cells compared to the 20% cells in the control experiment. *SAK* is an exception, which has almost 45% cells having 1 centrosome.

Three of these nine proteins (*ptc*, *nej* and *CG15676*) demonstrate a strong and consistent unaligned chromosome phenotype compared to the other genes, which have weaker phenotypes (Figure 3.12B). Figure 3.13 shows representative images for the RNAi experiment of each gene.

Based on the spindle length comparisons (Figure 3.12A), *ptc* also shows a strong *short* spindle phenotype, whereas *CG5731* and *gnu* show weak *short* spindle phenotypes.

Based on visual analysis of the mutant spindle shapes, the genes gave a variety of *mashed*, *weak* and *spiky* phenotypes. *SAK*, a well-known protein implicated in centrosome duplication (Bettencourt-Dias et al. 2005), and also our positive control, produces a characteristic *monoastral bipolar* spindle (appearing as 'broken diamonds' during microscopy).

In the subcluster, the nine genes except for *gnu* are all linked together via direct interactions. *SAK* and *CG2865* form the centre of this network (Figure 3.12C). Out of these nine proteins *SAK* and *ptc* belong to the training set and hence are known to be implicated in mitosis, whereas *gnu*, the third member of the PNG complex, comes from outside the MAP interactome. The remaining six proteins are predicted mitotic proteins.

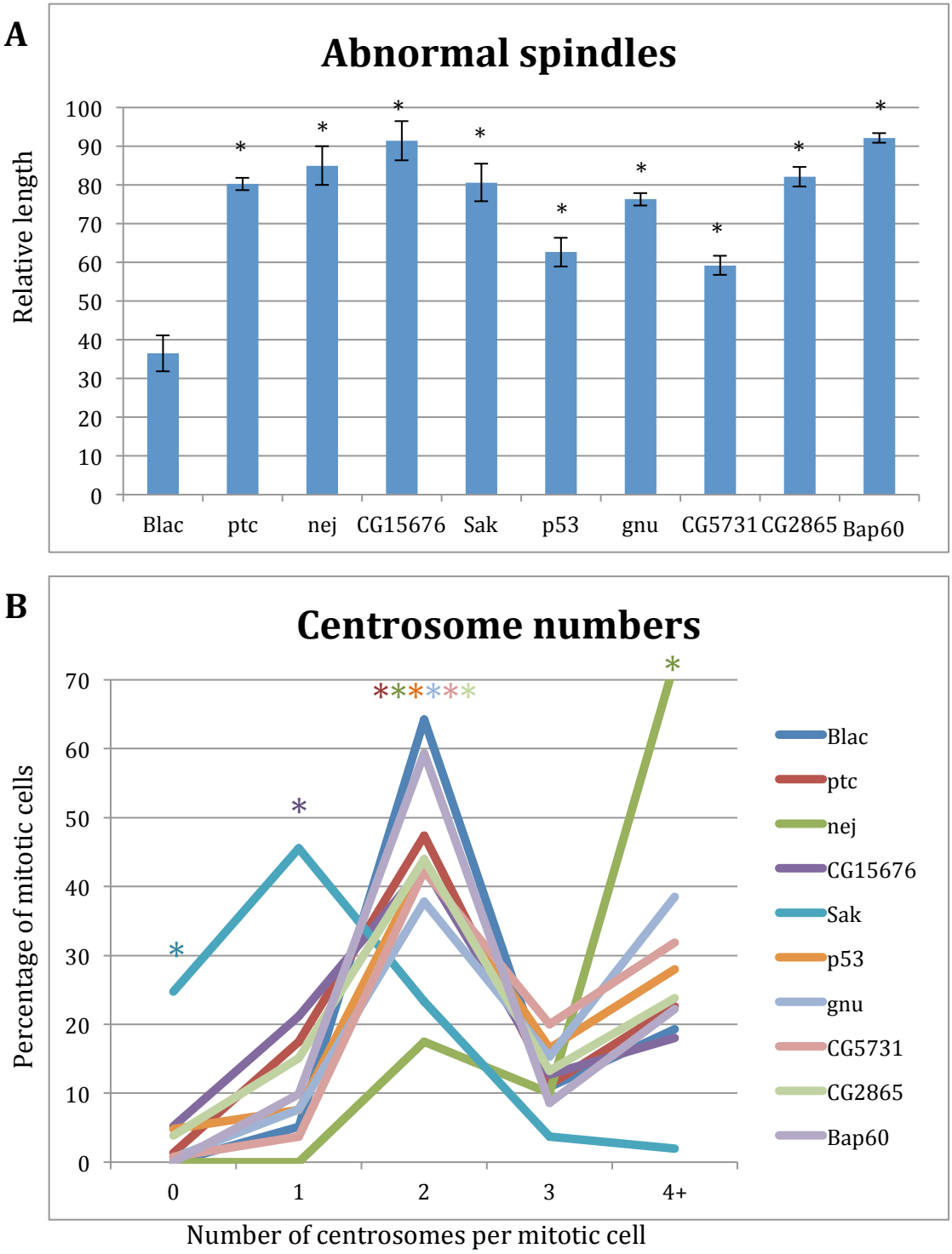
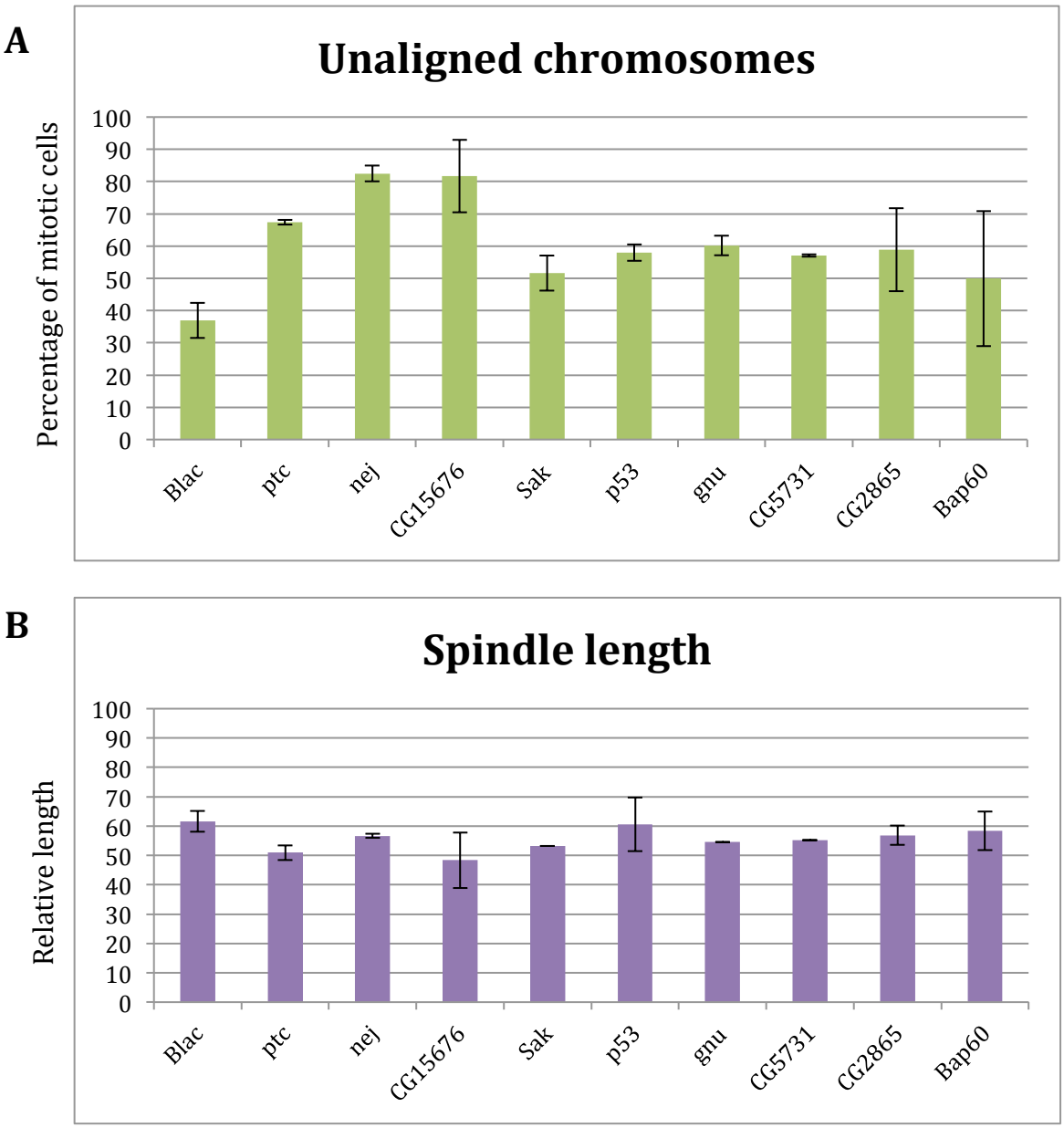


Figure 3.11: Spindle and centrosome defects for members of Class III. All spindle phenotypes and centrosome numbers are significantly different (paired t-test, $p<0.05$) from the control (Blac) experiment. (A) Bar chart showing the proportions of abnormal spindle phenotypes in the percentage of mitotic cells analysed. Each bar represents the average of two independent RNAi experiments (±SEM). (B) Line plots showing the centrosome number distribution as percentages of mitotic cells in each RNAi experiment. Each point is the average of two independent experiments. The control (Blac) has roughly 20% cells with 4+ centrosomes.



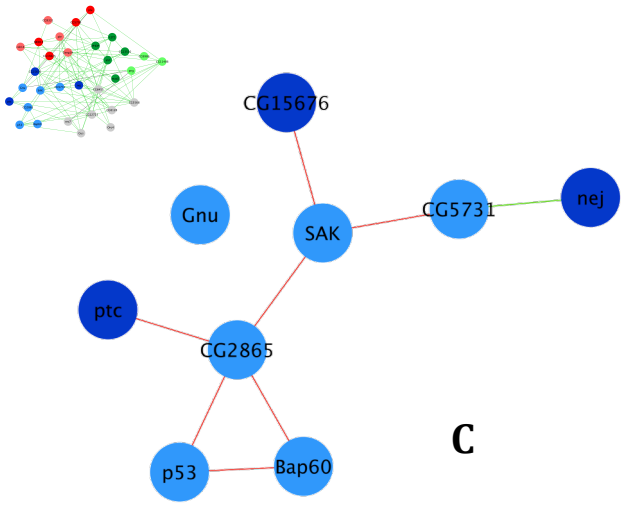


Figure 3.12: (A) Proportions of unaligned chromosomes, with each bar the average of two independent experiments (±SEM). (B) Average spindle length from two independent RNAi experiments for each gene (±SEM). (C) The connectivity of Class III proteins in the subcluster. Darker coloured nodes represent genes with stronger chromosome alignment defects.

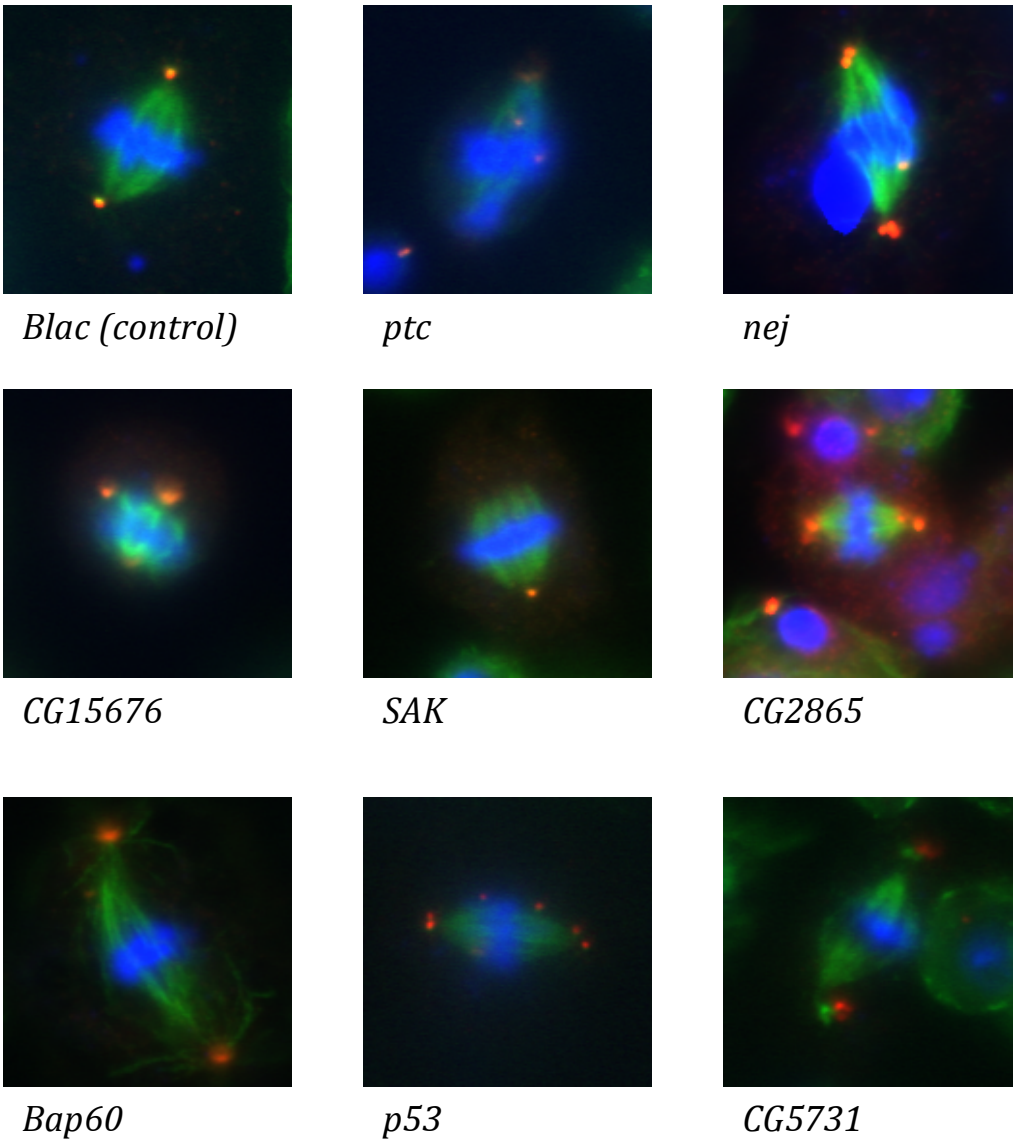


Figure 3.13: Representative images from each RNAi experiment of Class III. Each image is a merge of the three channels, showing tubulin/spindle (green), DNA (blue), and centrosomes (red). Blac is the negative control experiment with producing a wild-type diamond-shaped bipolar spindle.

SAK is a kinase with an established role in centrosome duplication. It is the homologue of the gene for the human PLK4 protein, which plays a role in centriole duplication and pericentriole formation. SAK is a reported hit in the Vale screen and has a 'monoastral bipolar' phenotype. The human homolog, *PLK4*, is one of the few in our study, which is a MitoCheck hit and has several phenotypes.

ptc or *patched* gene has several GO annotations (regulation of cell cycle, morphogenesis and biosynthesis), and has a human homolog, *PTCH2* which has a hedgehog family protein binding activity. No hits have been reported for *ptc* and its homologue in the Vale screen and MitoCheck.

p53 is one of the most well-studied proteins, a transcription factor primarily involved in cell death and apoptotic cascades along with several regulatory roles (e.g. cell cycle and response to UV). *p53* knockdown produced a *dim* spindle in our screen and although it is not reported as a hit in the Vale screen, it does show a similar *dim* phenotype in the image galleries.

Bap60 is an RNA polymerase II transcription factor with several regulatory roles. It has a human homologue (*SMARCD2*) and is reported as a manual hit only in the Vale screen with a 'long spindle' phenotype.

CG2865 is another regulatory protein with a GO annotation related to regulation of cell cycle. It does not appear to have a human homolog. *CG2865* is not a hit in the Vale screen but the analysis of the image galleries show dense centrosomes in some spindles, indicating the possibility of high centrosome numbers.

CG5731, an alpha-N-galatosaminidase, has an enzymatic molecular function like its human homolog, *GLA*. It produces the *moustache* phenotype in this study and has no mitotic phenotypes reported elsewhere.

CG15676 has GO annotations related to chaperone binding and protein folding with no specific molecular function established yet. It has no human homologs and has no hits in the Vale screen. In this study, *small* and *mashed* spindle phenotype was recorded with *shredded* chromosomes, which showed consistency with images in the Vale screen.

nej, or *nejire*, is an acetyltransferase which also produced a *mashed* spindle and *shredded* chromosome phenotype. The only comment found in the Vale screen was 'few mitotic cells', a property reproduced in our screen as well (only ~20 spindles analysed in each RNAi experiment). *nej* is also the top hit (i.e. the highest scoring protein) of our prediction set. The human homologue of *nej*, *CREBBP*, has a role in histone acetylation and other signalling and morphogenesis process. This indicates the possibility of a role for *nej* in mitosis with acetylation as the post-translational regulation mechanism, which makes it a very interesting candidate for further molecular characterization.

3.3.6 Class IV: Genes with no significant defects

From a total of thirty-three genes, seven show no significant spindle or centrosome defects (Figure 3.14). No consistent chromosome alignment defect or spindle length abnormality can be observed when compared to the control experiment (Figure 3.15). Figure 3.16 shows representative images for the RNAi experiment of each gene.

In the subcluster, one gene (*akt1*) belongs to the training set, and two genes (*CG8455* and *orc4*) come from outside the MAP interactome, which interact highly with the members of the subcluster. These genes are relatively less connected to each other when compared to previous classes (Figure 3.15C), giving credence to the relatively greater number of interactions within members of the previous classes with mutant phenotypes (Figures 3.9C and 3.12C). Most of these genes have human homologs which have several non-mitotic functions reported. None of them appear as mitotic hits in the Vale screen or the MitoCheck.

akt1 is serine/threonine kinase protein with a role in several biological process. *CG5568* has a ligase activity reported, while *CG8128* and *CG8455* both have GO annotations related to hydrolase activity in metabolic processes.

gus has annotations related to protein binding in oocyte axis specification while *orc4* has a DNA and ATP binding activity with a role in DNA replication initiation. Lastly, *rev7*, a homologue of the gene for the human MAD2L2 protein, has an unknown molecular function in the regulation of cell cycle.

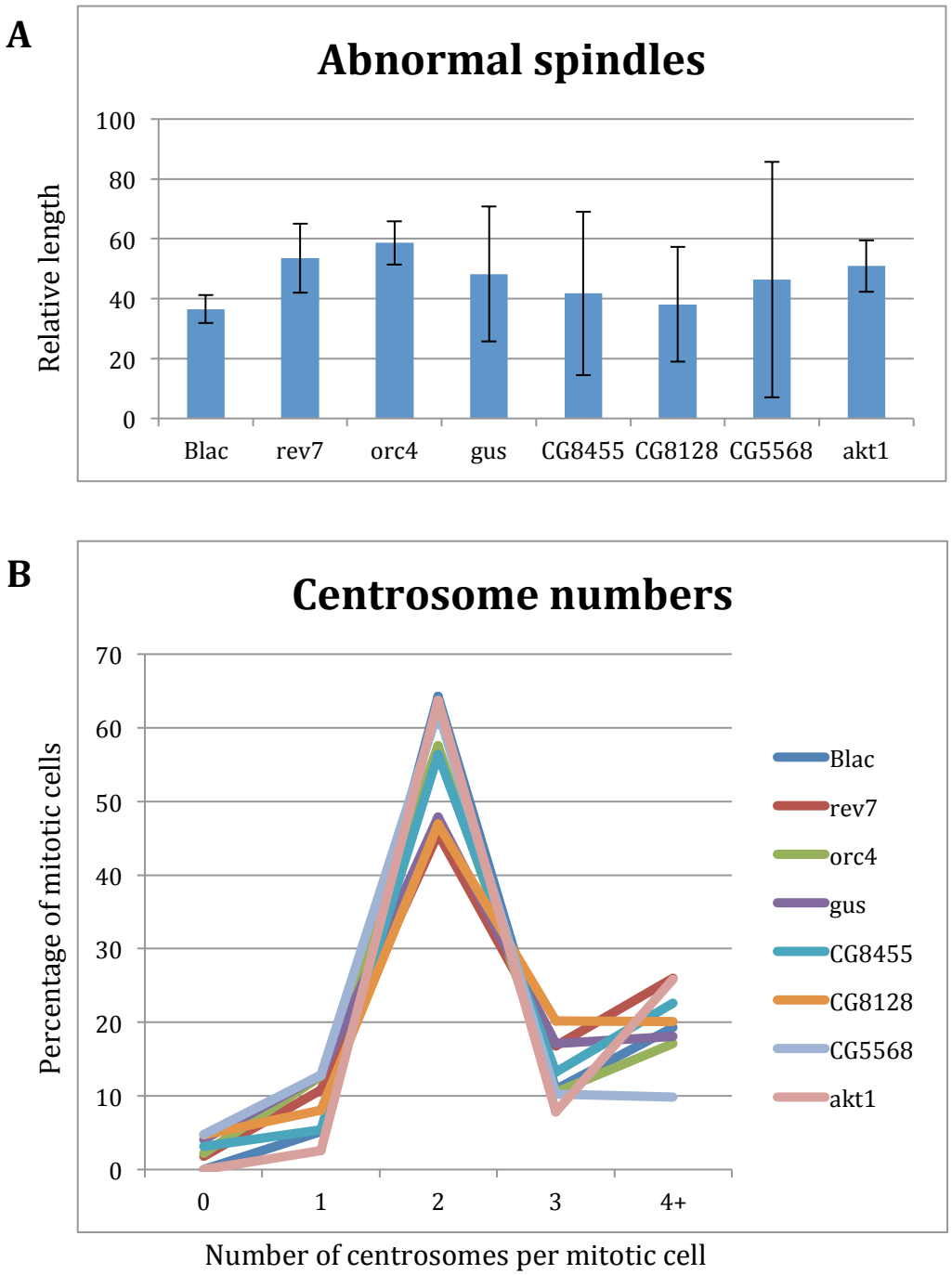


Figure 3.14: Spindle and centrosome defects for members of Class IV. No spindle phenotypes are significantly different (paired t-test, $p<0.05$) from the control (Blac) experiment. (A) Bar chart showing the proportions of abnormal spindle phenotypes in the percentage of mitotic cells analysed. Each bar represents the average of two independent RNAi experiments (±SEM). (B) Line plots showing the centrosome number distribution as percentages of mitotic cells in each RNAi experiment. Each point is the average of two independent experiments. The control (Blac) has roughly 20% cells with 4+ centrosomes.

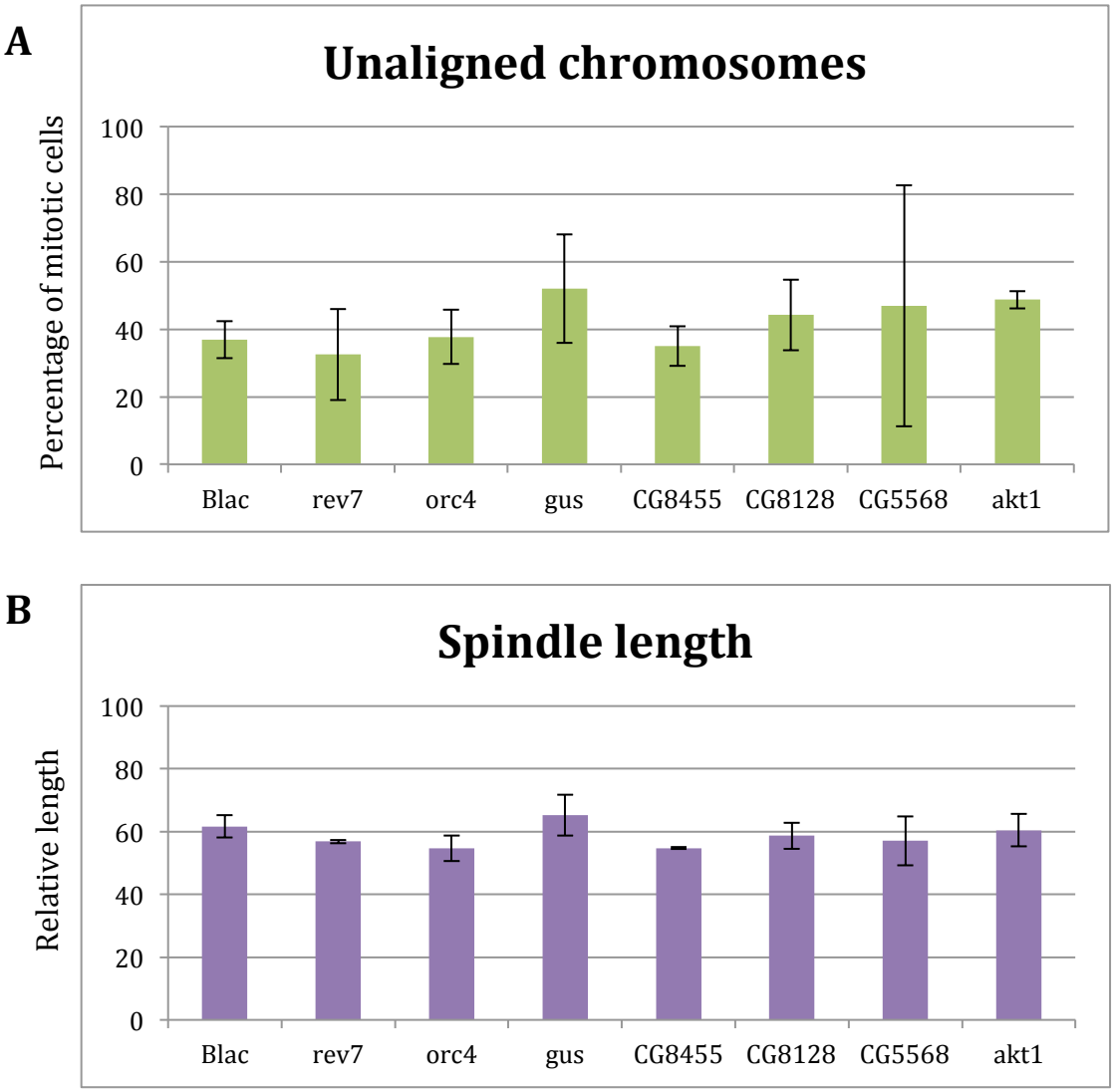


Figure 3.15: (A) Proportions of unaligned chromosomes, with each bar the average of two independent experiments (±SEM). (B) Average spindle length from two independent RNAi experiments for each gene (±SEM). (C) The connectivity of Class IV proteins in the subcluster.

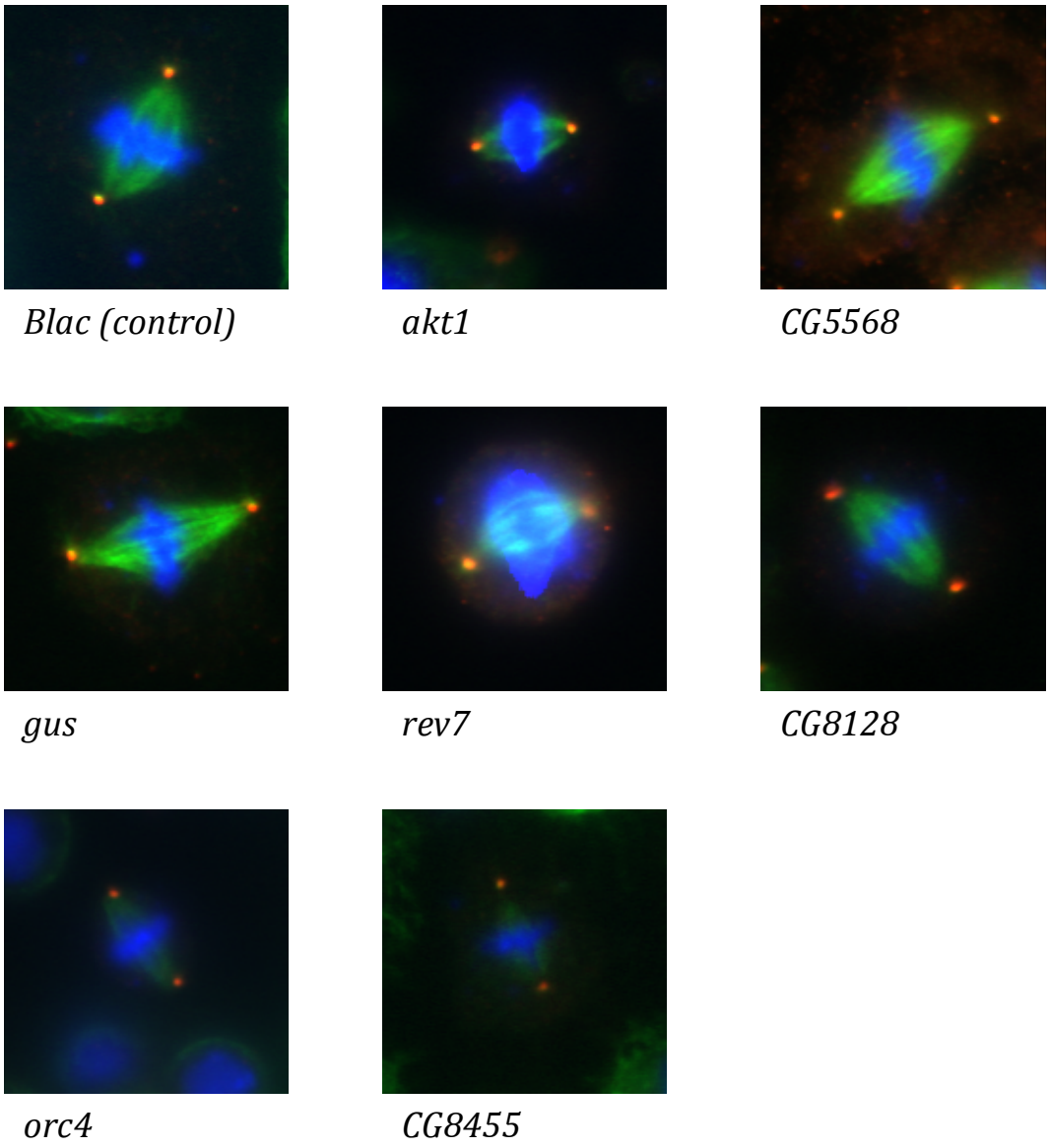


Figure 3.16: Representative images from each RNAi experiment of Class IV. Each image is a merge of the three channels, showing tubulin/spindle (green), DNA (blue), and centrosomes (red). Blac is the negative control experiment with producing a wild-type diamond-shaped bipolar spindle.

In summary, this screen suggests a mitotic function for ~80% (26 out of 33) of the predicted mitotic proteins from subcluster-16, each confirmed by two independent RNAi experiments followed by blind analysis. These RNAi experiments demonstrated that dsRNA treatment against 24 previously unreported genes disrupts normal spindle assembly and the centrosome cycle in *Drosophila* S2 cells. This includes several genes that have been missed by the Vale screen, possibly due to the high rates of false discovery, especially false negatives, that is inherent in large-scale screens (Mohr et al. 2010c).

These 26 genes break down to 6 from the training set which are known mitotic proteins (out of 7; ~86%), 15 from the predicted mitotic MAPs according to the scores from the MAP prediction model (out of 19; ~79%) and 5 from outside the MAP interactome (out of 7; ~71%), which although enriched as direct interactors of members of the MAP network, do not belong to our list of MAPs. The training set clearly has the highest hit-rate, which is expected, as all of them are known mitotic proteins. The prediction set is interesting as it has a higher hit-rate than the proteins from outside the MAP interactome, supporting the design and underlying assumptions of our MAP prediction model. These hit rates are by far higher than those encountered in undirected genome-wide screens (Guest et al. 2011). Apart from the enrichment of MAPs, the high hit-rate compared to undirected screens and the large number of centrosomal phenotypes can also be explained by another inherent bias in our study, which is the large number of interactions with known centrosomal proteins like SAK (interacts with 14 members of subcluster-16) and the large number of highly

connected Cyclins and Cdk proteins that have been ignored in our study of the subcluster (Section 2.4.3).

Most of the genes analysed in the screen have regulatory functions as documented by their GO annotations. This can also be demonstrated by the number of interactions they have with Cyclins, CDKs and complexes like the PNG kinase complex. This indicates that the product of these genes might not have structural roles in mitosis, but may have regulatory roles instead in relation to microtubules and mitosis. Many human homologs of the (24 out of 33) test genes, are regulatory in function and are not well characterized. This provides another opportunity of transferring phenotypes from the screen and studying them in human cell lines.

The connectivity between proteins in subcluster-16, in light of the RNAi screen offers interesting insights. With *SAK*, an integral protein of the centrosome cycle, as one of the hubs within the subcluster and also the large number of centrosomal phenotypes in the screen, it can be inferred that many of the members of subcluster-16 might have roles related to the centrosome cycle. This provides one possible direction for further investigation.

The most interesting gene encountered in our data was *nej*. Nej is the highest scoring protein in the MAP prediction model and has produced a strong centrosomal and chromosomal defect in the RNAi screen. Nej and its human homologue are known to have histone acetyltransferase activity and have well documented cellular functions. This suggests a possible role in mitotic regulation via acetylation making Nej an interesting case for further characterization.

The S2 cell line in *Drosophila* and the RNAi technique, provide a valuable system for the functional analysis of genes, which is cost-effective, efficient and scalable. Like any experimental system, this approach also comes with certain limitations. S2 cells are derived from embryonic cells, which have spontaneously immortalized through an unknown mechanism (Schneider 1972). Additionally, S2 cells are also known to possess high levels of variability between batches and non-textbook mitosis, which often have extra centrosomes per cell (Goshima 2010). These are two main caveats, which need to be kept in mind while interpreting results.

RNAi screens on the other hand are known to have very high false discovery rates and poor overlap between the results of different studies that focus on the same biological process (Guest et al. 2011b). Both false positives and false negatives contribute to this problem.

The main source of false positives in RNAi screens are off-target effects, or OTEs (Ma et al. 2006). OTEs are the result of the unintended silencing of multiple genes by dsRNA based on base pair complementarity. In this study, primer design for dsRNA production was carried out using SnapDragon, a software that minimizes the chances of OTEs in the RNAi experiments (Kulkarni et al. 2006; Ma et al. 2006). However, the most robust way is to conduct a second screen with different dsRNA against the same genes, which clearly was not undertaken in this thesis. This would be the first logical step to validate our hits in the screen.

The second source of error in RNAi screens are false negatives, which mainly come from the inefficient knockdown of target genes, or weak and subtle phenotypes that can be missed by the screen. Measuring transcript levels after RNAi experiments to confirm the knock down of gene transcription can control the first problem. The problem of weak phenotypes can be addressed by relaxing the statistical cut-off (at the cost of increased false discovery rate), increasing the number of replicates and/or using a sensitized background for the RNAi experiment.

In summary to address the weaknesses in this RNAi screen, re-screening the same genes with different dsRNA fragments would validate the positive hits with mutant phenotypes. This should be coupled with measuring the transcript levels of each knockdown with robust controls to minimize false negatives. The next logical step would be to try to recapitulate the phenotypes of genes of interest in a different cell line or preferably, through *in vivo* analysis in fly embryos. The Vienna Drosophila RNAi Centre (VDRC) can be a good resource for obtaining RNAi fly lines for such a study (Dietzl et al. 2007).

3.4 Conclusions

In this Chapter I set out to conduct an *in vitro* functional validation of members of subcluster-16 using an RNAi screen in S2 cells. I recapitulated mutant phenotypes of genes with known mitotic function (as reported by the Vale screen and others), which validated the proper functioning of the RNAi pipeline designed and optimized for this screen. The results also discovered that the dsRNA-treatment of 24 previously unreported genes produce aberrant spindle and centrosomal phenotypes in S2 cells. Altogether, the RNAi screen validated the mitotic function of ~80% of the subcluster member proteins (26 out of 33) and roughly the same for subset predicted mitotic proteins within subcluster-16 (15 out of 19).

The identification of mitotic phenotypes on its own does not provide sufficient knowledge about the underlying molecular function of each gene and the biological meaning of each interaction between the genes in the sub-cluster at the molecular level. Follow up experiments, especially, *in vivo* RNAi, can be conducted to recapitulate and validate the phenotypes in this study.

In the next Chapter, I conduct an *in vivo* analysis of a selected group of genes from the subcluster in *Drosophila* embryos. I analyse their subcellular localization and interaction with other proteins including microtubules.

Chapter 4
Experimental *in vivo* validation: GFP localization and proteomics

4.1 Introduction

The discovery of the green fluorescent protein (GFP) has been a landmark moment in the history of biochemistry and cell biology.

GFP is a 30 kDa protein which was first found in the jelly fish, *Aequoria victoria,* fluorescing in photocytes that are located on the margins of its umbrella. It is encoded by a single gene and is auto-activated roughly within an hour through post-translational modification, producing a bright fluorescence in the centre of its barrel-like structure. The wild-type GFP has two absorption maxima, a major peak at 395nm and a minor and 475 nm. Excitation at any of these wavelengths leads to the emission of green fluorescence at 508nm (Gerdes & Kaether 1996).

The emergence of a family of GFP proteins that consisted of similar proteins from other species and engineered variants of the natural GFP, led to the development of an array of technologies that enabled scientists to monitor spatio-temporal dynamics of cellular components at the molecular level. Members of this family of proteins have a variety of fluorescence efficiencies, absorption and emission spectra and photo-stabilities (Chudakov et al. 2005). The impact of this discovery was acknowledged by the Nobel Prize for Chemistry in 2008, which was awarded to three scientists, Osamu Shimomura, Martin

Chalfie and Roger Y. Tsien, who worked on the 'discovery, expression and development' of the green fluorescent protein.

The most popular applications of GFP in the field include monitoring gene expression (by expression under the promoter of the gene of interest) and protein localization studies (of chimeric proteins that are expressed by the gene of interest fused at one end with the GFP gene). Several techniques were soon developed to study protein dynamics that produce valuable quantitative data about cellular components over time, e.g.:

- Photo-activation, which involves activating and tracking an inactive fluorescent protein at a precise time and location in the cell.
- Fluorescence resonance energy transfer, or FRET, which is used for studying the interaction of two proteins of interest.
- Fluorescence recovery after photo-bleaching, or FRAP, is used to study the behaviour of fluorescent proteins after bleaching at a precise time and location in the cell.

As the family of GFP and GFP-like fluorescent proteins increased in size and diversity, more and more applications emerged for the structural and functional study of biological systems (Figure 4.1).

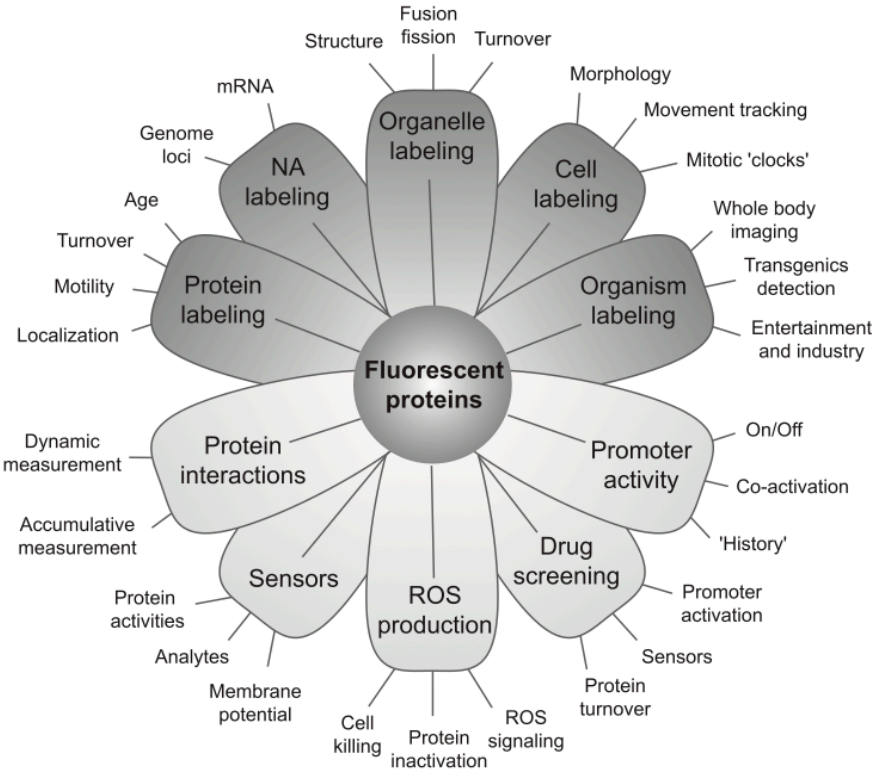


Figure 4.1: Various applications of the fluorescent proteins. Dark petals represent structural studies and light petals represent functional studies. Image taken from (Chudakov et al. 2005).

In this Chapter, two of these methodologies are exploited to study a selected cohort of proteins from subcluster-16 in fly embryos. I generated DNA plasmid constructs, each of which possessed the cDNA of a potential MAP directly downstream of the cDNA for a GFP, in order to produce transgenic flies that would express the GFP-fusion protein in the *Drosophila* early embryo. Embryos from each fly line were used to analyse the localization of each GFP fusion product and co-immuno-precipitation experiments to identify putative interacting proteins.

This *in vivo* analysis was conducted in parallel to the *in vitro* RNAi screen in Chapter 3, and therefore did not allow the results of one approach to influence the other.

4.1.1 Motivation

There are two interlinked motivations behind the *in vivo* approach pursued in this Chapter. Firstly, the identification of true interacting proteins in the early embryo can contribute substantially to our understanding of the biological processes in which the putative MAPs are involved. These experiments will allow for the exploration of the relationship between the interactions in the subcluster, which have mainly been derived from yeast-2-hybrid (Y2H) experiments, and *in vivo* interacting partners, especially given the significant levels of false positives present in Y2H datasets as discussed in Chapter 1. These insights into the real interactions occurring in the native system of *Drosophila* proteins (fly embryos), unlike yeast nuclei in the Y2H system, will help us identify which of the hypothetical interactions in subcluster-16 between known mitotic proteins that are occurring through interconnecting putative mitotic MAPs, hold true.

The second motivation was to get further insights that could complement results from the *in vitro* analysis in Chapter 3 by studying the potential localization of these proteins with in the embryo, both during interphase and mitosis. Here, the *Drosophila* embryo provides an efficient system that has relatively faster cell cycles (10-20 minutes) which means the entire cycle can be monitored in a single embryo. As mitosis progresses, dividing nuclei migrate away from the centre of embryo which leads to cycles 10-13 occurring parallel to the cortex

(Appendix VIII). This not only makes fly embryos excellent specimens for confocal microscopy, but also enables the imaging of multiple nuclei/spindles at once. Additionally due to the synchronous rounds of mitosis in a common cytoplasm along with the high levels of tubulin in the early embryo, large quantities of MAPs and mitotic regulators are presumed to be required for microtubule dynamics. This makes early embryos a good source of such proteins for biochemical isolation and pull down experiments.

Together, these approaches will help us experimentally validate the findings of the MAP prediction model (Chapter 3) and to analyse which proteins do interact with microtubules or MAPs and co-localize with mitotic components *in vivo*, or otherwise.

4.1.2 Selected test genes

From the 18 potential MAPs in subcluster-16, 15 have interactions with known proteins and protein complexes, which have regulatory functions at well-defined stages preceding and during mitosis (Figure 4.2). These are the ones that were the focus of this study, and therefore, excluded Nej, CG9331 and CG9986.

A filter of gene expression levels was applied using data from the modENCODE study (Roy et al. 2010). Expression levels at the early embryonic stage (0-3 hours) were used when the embryo is mitotically active. Four proteins (CG115676, CG12723, CG13510 and CG8128) are further excluded from the list of our test genes, based on very low to no expression during the 0-3 hour stage (Table 1).

cDNAs of all of the remaining 11 genes (Table 4.1) were amplified using PCR, in order to clone into a standard Gateway® recombination Entry Vector, prior to final recombination into available Gateway *Drosophila* Expression Vectors which were injected into fly embryos (details in section 4.2). After hatching, viable transgenic flies were successfully obtained for seven transgenes, i.e., CG5568, CG5708, CG2865, CG5731, Gus, PHDP and CenG1A.

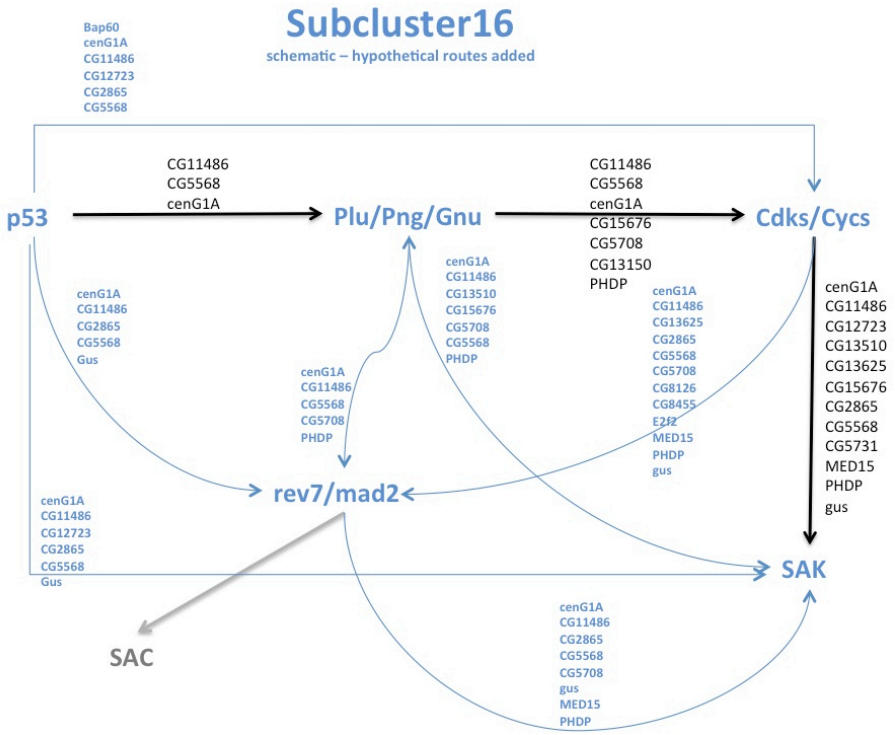


Figure 4.2: Subcluster-16 from the top 100 hits of the MAP prediction model (Chapter 2). This schematic diagram represents the five known cell division and cell cycle proteins and protein complexes in the subcluster. The dark arrows represent the temporal order of each group in the cell cycle. The light arrows represent hypothetical routes with names of proteins from the subcluster, which connect the two group on both ends of the arrow.

Protein	Rank in network	Score in network	Occurrence in subcluster-16	Expression
CG5708	4	0.818	5	Early embryos
CG2865	6	0.763	6	Early embryos
Bap60	9	0.722	1	Early embryos
CG15676	10	0.652	3	Testes only
CG5568	18	0.613	10	Early embryos
CG11486	25	0.569	10	Early embryos
MED15	31	0.552	3	Early embryos
gus	46	0.468	5	Early embryos
CG13625	47	0.462	2	Early embryos
PHDP	48	0.452	6	Early embryos
CG12723	53	0.437	3	e14-e18h only
CG5731	64	0.423	1	Early embryos
CG8128	116	0.333	1	e18-e24 only
CG13510	190	0.263	3	e24h only
cenG1A	870	0	10	Early embryos

Table 4.1: The 15 proteins out of the 18 predicted MAPs that are connected to the 5 known complexes in the subcluster. Four genes (highlight pink) are excluded, as they do not express in early embryos.

A fly line expressing a YFP-nej transgene which was generated as part of the Fly-TRAP project (Kelso et al. 2004), was available in the lab and was added to our analysis. Nej is the top-ranking hit from the MAP prediction model and a protein that directly interacts with all seven of the proteins that were cloned.

In the following section, we analyse extant data about these eight proteins and describe them in more detail.

CG5708

CG5708 is 26.6 kDa protein with a Zn-binding LIM-type domain and is thus predicted to have a zinc ion binding activity. It is expressed moderately during the early embryonic stage (Roy et al. 2010). LIM domains act as scaffolds for the assembly of multimeric protein complexes and are found in different proteins with roles in gene expression regulation, cytoskeleton organization and tumour formation. The domain itself comprises of two characteristic zinc finger motifs.

The CG5708 protein has a human homologue called LIM-domain only protein 4 (LMO4, Uniprot: P61968). The homologue protein has been reported to have a role in DNA-dependent regulation of transcription. LMO4 is also reported to be oncogenic with high expression in several breast cancer cell lines (Montañez-Wiscovich et al. 2010). The same study demonstrates that LMO4 depletion results in G_2/M arrest and amplification of centrosomes coupled with aberrant spindles, the latter reconciling with the CG5708 phenotype in Chapter 3. CG5708 has been reported to play a potential role in the stabilization of the 26S proteasome in an RNAi screen against transcription factors in *Drosophila* (Grimberg et al. 2011).

CG2865

CG2865 is one of the uncharacterized proteins with unknown molecular function. This 46.8 kDa protein has a SERTA domain and has moderate expression levels in early fly embryos. CG2865 has no reported human homolog.

The SERTA domain is the largest conserved domain in TRIP-Br proteins (Hsu et al. 2001), a new family of transcription regulators what includes the oncogene p34(SEI-1). p34 is known to have a dual role in the regulation of CDK4 and the G_1/S transition in the cell cycle (J. Li et al. 2004). The CDK4-binding region covers most of the SERTA domain. p34 (through the SERTA domain) is also known to interact with bromodomain-containing proteins like p300 (Eckner et al. 1994), which interestingly is the human homologue of Nej, the top hit in our study.

CG5731

CG5731 is an uncharacterized protein with a predicted galactosaminidase activity. It is 47 kDa and has an aldolase-type TIM barrel. Its human homolog, GLA (Uniprot: P06280), has a catabolic activity and has been annotated with functions related to oligosaccharide and small molecule biosynthesis.

PHDP

PHDP or the putative homeodomain protein is a 25 kDa protein with no certain role in any biological process. It contains a homeodomain, which are DNA binding domains that act in a sequence-specific manner and are mainly involved in the regulation of developmental processes. The human homologue of PHDP is the Paired mesoderm homeobox protein 2A (PHX2A, Uniprot: O14813), which

has similar annotations related to DNA-dependent transcriptional regulation with no clear biological process. Rychlik *et al* in their study of transcription factors show that PHX2A has a nuclear chromatin localization (Rychlik et al. 2005).

CenG1A

CenG1A is a 101 kDa member of the Centaurin gamma-1-like family of proteins. Centaurins are a family of GAPs (GTPase activating proteins) that contain an ANK repeat domain (adapter domains mediating attachment of integral membrane proteins to the spectrin-actin membrane cytoskeleton), a PH domain (found in proteins that mediate intracellular signalling and cytoskeleton organization by binding to phophatidylinositol lipids in the membrane) and the Arf-GAP domain (zinc fingers with an Arf-specific GTPase activating domain). The γ-subfamily of Centaurins has an additional GTPase-like domain (GLD) at the N-terminus.

The protein features and its human homolog, AGAP3 (Uniprot: Q96P47) provide insights into its molecular function, i.e. potential role in small GTPase-mediated signal transduction, but not much is known about the exact cellular processes where cenG1A might have a role. The protein has recently been shown to have a more generic NTPase function (Soundararajan et al. 2007). It has been shown as a possible hit in a proteomics study of the phagosome interactome (Stuart et al. 2007) and an RNAi-based study of cell adhesion (Bakal et al. 2007).

CG5568

CG5568 is a 61 kDa protein with an AMP-binding domain and is predicted to have metabolic role with a synthetase/ligase function. It has a high expression in the early fly embryos and has no reported human homolog.

Gus

Gus, or gustavus, has been reported to have a role in the oocyte anterior/posterior axis specification. This 37 kDa protein also has a high expression level in early embryos and has several domain features that include:

- B30.2/SPRY domain (found in MHC proteins/calcium release channels)
- ConA-like lectin/gluconase domain (found in cell recognition and cell adhesion proteins)
- SOCS protein domain (found in signal transduction proteins),
- SP1a/Ryanodine (SPRY) receptor (has a role in calcium release and storage from ER)

Gus has a human homolog, SPSB1 (Uniprot: Q96BD6), which has an additional protein ubiquitination role reported as well and is a potential member of the E3 ubiquitin-protein ligase complex mediating ubiquitination and subsequent proteasomal degradation of target proteins.

Nej

Nej, or nejire, is the most well studied protein in the cohort. It has been reported to have roles in several key processes ranging from sensory organ development and organelle organization to the regulation of the cell cycle and the DNA replication checkpoint. This large 350 kDa protein has high levels of expression

in early fly embryos and is comprised of several protein features, including the bromodomain, KIX domain, zinc fingers and an acetyltransferase domain. The cellular component GO annotation of Nej includes the nucleus and protein complex.

Bromodomains are found in chromatin-associated proteins and in nuclear histone acetyltransferases and interacts specifically with acetylated lysine residues, like those found in the N-terminus tails of histones. The KIX domain is a conserved domain found in CBP and p300 proteins through which they bind CREB proteins.

Nej has two reported homologs in the human proteome, which are relatively well-studied proteins, i.e. p300/EP300 and CREB-binding protein or CREBBP (Uniprot: Q7Z6C1 and Q92793, respectively). Both contain bromodomains and two types of zinc fingers each (ZZ-type and TAZ-type). The proteins play a role in several pathways (like Nej) and in addition to those in virus-host interaction (Yang et al. 1996), the HIF pathway (Kung et al. 2004), the SNF2/SWI2 pathway (Johnston 1999) and the SHH pathway (Villavicencio et al. 2000). CREBBP shows cellular localization in nucleus, the cytoplasm, histone acetyltransferase complexes, nuclear chromatin, and also the condensed chromosome outer kinetochore (based on Cellular Component annotation in Gene Ontology). Both Nej homologs have been widely studied for their histone acetyltransferase activity in different processes like the G_1 arrest (Tsuchiya et al. 2011; Smolik & Jones 2007), complex formation with APC/C (Turnell et al. 2005; Nath et al. 2011), and the G_2/M checkpoint (Y.-J. Chen et al. 2009; Wasner et al. 2003)

Ha *et al*, studied the depletion of three representative Histone Acetyltransferases (p300, CBP and PCAF) on chromosome condensation/decondensation. Several mitotic phenotypes have been reported including multipolar spindles, mitotic arrest and catastrophe and reduced condensin subunit levels. They also show aneuploidy in cancerous cell lines, which adapted to HAT depletion (Ha et al. 2009).

Analysing the connectivity of these selected proteins in subcluster-16 reveals interesting insights (Figure 4.3). Nej, the top hit of our prediction set is not connected directly to any of the known cell division/cycle proteins, but interacts with 6 of the 8 MAPs chosen for *in vivo* analysis. SAK emerges as a hub protein that interacts with all the selected genes except Nej. The PNG complex connects only through CG5708 and CG5568, while Mad2 interacts with CenG1A only (Figure 4.3B). Removing the known proteins and complexes, leaves behind the selected transgenes. Five of these proteins are connected to each other through nej, while two of them (CG2865 and Gus) remain unconnected (Figure 4.3C).

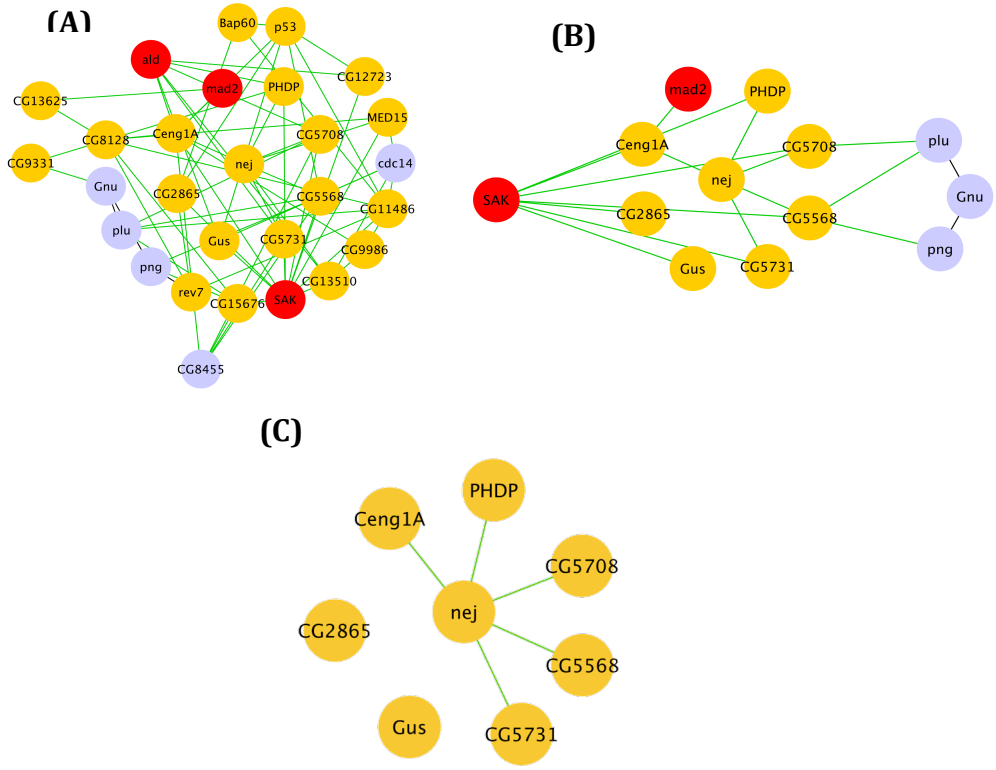


Figure 4.3: The connectivity of the members of subcluster-16 – The complete subcluster (A), the transgenes along with the known proteins, SAK, mad2 and the PNG complex (B) and the connectivity within the transgenes only.

4.2 Materials and methods

4.2.1 Cloning – TOPO/LR Gateway Vectors

Each gene was amplified from BDGP Gold cDNA templates (DGRC, Indiana University) using Phusion polymerases (NEB) and primers designed for use in the D-TOPO® reaction (Appendix IX). Each amplicon was cloned in the pENTR™ vector using the pENTR™/D-TOPO® kit (Invitrogen, Life Technologies). Half reaction volumes were followed according to the manufacturer's protocol. Each construct was transformed into competent OneShot® TOP10 cells (Invitrogen/Life technologies) and plated onto Kanamycin-LB Agar plates (50µg/ml, Sigma). Positive clones were confirmed using restriction digests (NEB enzymes) followed by DNA sequencing (Source Biosciences PLC).

The LR reaction was carried out using the LR Clonase® II kit (Invitrogen/Life technologies). Again, half reactions were carried out according to the manufacturer's protocol with an appropriate molar ratio of the each entry vector and the pPGW destination vectors with the UASp promoter and the GFP tag at the N-terminus (*Drosophila* Gateway Vector Collection (DGRC 2013)). The constructs were transformed into TOP10 cells and plated on Ampicillin-LB Agar (50 µg/ml). Successful clones were cultured in larger volumes of Amp-LB broth and purified using Midi prep kits (QIAGEN).

4.2.2 Embryo injections

The DNA constructs for each transgene (1µg/µl) were sent to BestGene Inc. for injection into w^{1118} embryos. Embryos which survive and hatch are the crossed

to w^{1118} male or female virgin flies. Flies with the pPGW construct were selected for their red eye colour. Several lines for each transgene were appropriately balanced (outsourced and carried out by BestGene Inc.) that identified the chromosomes and were shipped back.

The lines, which had the darkest red eye colour, were selected for each transgene and were crossed with homozygous flies expressing GAL4 under the maternal α-tubulin promoter. Embryos laid by the offspring of these crosses were collected for further experiments.

4.2.4 Embryo collection and storage

Embryos were collected on Apple Juice Agar plates with a blob of yeast (Tesco) paste in the middle. 3-hour collections were then dechorionated for three minutes in thin bleach (Tesco) and washed briefly in PBS with 0.1% Triton-X (Sigma). They were transferred into eppendorf tubes, weighed and flash frozen in liquid nitrogen before storage at -80^{o}C.

4.2.5 Live imaging of embryos

Embryos (1-2hour) were manually dechorionated by gentle manipulation on a double-sided sticky tape (Scotch). 10-12 embryos were lined on a coverslip with a thin layer of glue (made by soaking double-sided tape in heptane), and covered in Halocarbon oil solution (1:1 solution of Halocarbon Oil 27 and 70S; Sigma). The embryos were then visualized using a Zeiss® LSM 510 confocal microscope and imaged on a single plane, every 10 seconds. The images were converted into movies using Zeiss LSM software and processed using ImageJ.

4.2.7 GFP pull-down experiments and proteomics analysis

Embryo extracts for each co-immunoprecipitation (pull down) experiment were made from 0.6-0.9g of frozen 0-3hour old embryos homogenized in 4x volume of C buffer (50 mM HEPES at pH 7.4, 50 mM KCl, 1 mM MgCl2, 1 mM EGTA, 1mM NaF) with protease inhibitors (Roche). The homogenate was centrifuged at 14,000 rpm at 4^{o}C for 10 minutes (low-speed spin). The pellet was discarded and the supernatant was centrifuged for 30 minutes at 100,000 rpm and 4^{o}C (high-speed spin).

For the pull-down, the GFP Trap-A beads were used which are made of llama anti-GFP nanobodies covalently coupled to Protein A agarose beads (Chromotek). In some repeat experiments the GFP Trap-M beads were used which have the antibody immobilized onto magnetic beads. In both cases, the beads were prepared by washing three times with C-buffer.

For each experiment, 60µl of 1:1 bead slurry was centrifuged briefly to let the beads settle and the C buffer was removed. To these beads 1.5ml of the high-speed supernatant was added, and the tube was incubated at 4^{o}C for 3 hours on a rotator. After removing the depleted sample, the beads were again washed three times with C-buffer, before being stored at -80^{o}C. The beads were sent to the mass spectrometry facility at the University of Bristol where the beads are digested with Trypsin and processed for analysis through liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS).

4.3 Results and Discussion

4.3.1 GFP Localization in embryos

GFP-fused constructs for all eight genes were expressed in fly embryos. These were constructs with N-terminal GFP tags (except Nej, which was tagged with YFP) and were driven by the maternal α-tubulin promoter and imaged during the 1-2 hours after being laid.

The objective here was to observe the localization of the products of these transgenes in relation to microtubules or any part of the mitotic machinery (centrosome, spindle, condensed chromosomes) in the cell, both during mitosis and in interphase.

Confocal microscopy revealed a variety of localization for these proteins from specific ones, like chromatin, microtubules and spindles to more general cytoplasmic localizations. Images at interphase and different mitotic phases from representative localization movies for each fusion protein are shown in Figures 4.4-4.12. The following section describes each experiment in detail.

GFP-CG2865, the SERTA-domain containing protein produced a cytoplasmic localization, excluded from the nuclear region during interphase (Figure 4.4). GFP-CG5731 showed a similar cytoplasmic phenotype with exclusion from the nuclear region during interphase when the nuclear membrane is intact. Representative images for the GFP-CG5731 localizations are shown in Figure 4.5.

GFP-CenG1A produced a nuclear localization during interphase, that during mitosis also appeared to extend to the mitotic spindle. However, no individual microtubule localization (such as kinetochore or astral microtubule) is resolved (Figure 4.6).

CG5568 is another metabolic protein with a synthetase/ligase activity. GFP-CG5588 had a nuclear localization similar to GFP-CenG1A. The microtubule staining extended to the spindle during mitosis. However, the GFP-CG5568 localization observed is slightly weaker (Figure 4.7).

GFP-Gus has an interesting cytoplasmic localization during interphase, excluded from the nuclear region, which has an intact nuclear envelope at that stage. The staining fades during entry into mitosis and prophase, but clearly reappears on microtubules during metaphase and into the anaphase. It re-diffuses towards the exit from mitosis and regains a cytoplasmic localization. This indicates the possible interaction of Gus with microtubules during metaphase (Figure 4.8).

The GFP-PHDP fusion protein localized to chromatin during interphase and upon exit from mitosis. Figure 4.9 shows an array of six representative images from one complete interphase. The chromatin co-localization clearly intensifies during the second half of interphase with intense regions (dots in Figure 4.10D-F). The staining was lost immediately upon entry into mitosis (Figure 4.10G).

GFP-CG5708 localized on the chromatin during interphase, and similar to GFP-PHDP produced more intense dots in the chromatin region before the onset of mitosis (Figure 4.11). YFP-Nej had a nuclear localization in the embryo during interphase, which fades away upon entry into mitosis (Figure 4.12).

These results show that the subcluster proteins have a variety of subcellular localization in the fly embryo during interphase. GFP-Gus and GFP-PHDP should specific microtubule and chromatin localizations, respectively. GFP-CG5568 and GFP-CenG1A localize peculiarly to the nucleus during interphase and the mitotic spindle during mitosis. Ceng1A is a member of the Centaurin family of GAPs that have several membrane binding domains. This protein is predicted to have a role in membrane integrity through interactions with Actin and cytoskeletal organization (through the PH domain). These predicted functions can explain the localization pattern found in fly embryos in this experiment.

The GFP-CG2865 and GFP-CG5731 proteins produced more general cytoplasmic localizations and were excluded from the nuclear region during interphase when the nuclear membrane is intact. CG2865 has the SERTA domain as mentioned in the previous section. These domains are found in TRIP-Br proteins like p34 – a G_1 checkpoint and CDK4 regulator. p34 also binds the p300 protein through the SERTA domain. The presence of this domain suggests a regulatory role for CG2865, which can partially explain the cytoplasmic localization with no direct interaction to any mitotic component of the cell. The cytoplasmic localization of GFP-CG5731 can be explained in part by the fact that it is predicted to be a metabolic enzyme with a glactosaminidase/aldolase activity.

CG5708 has Zn-finger motifs and DNA-dependent transcription regulation role, which can explain the chromatin localization of GFP-CG5708 in the fly embryo. YFP-Nej also had nuclear localization, which is consistent with the fact that Nej has zing finger motifs and is a known histone acetyltransferase.

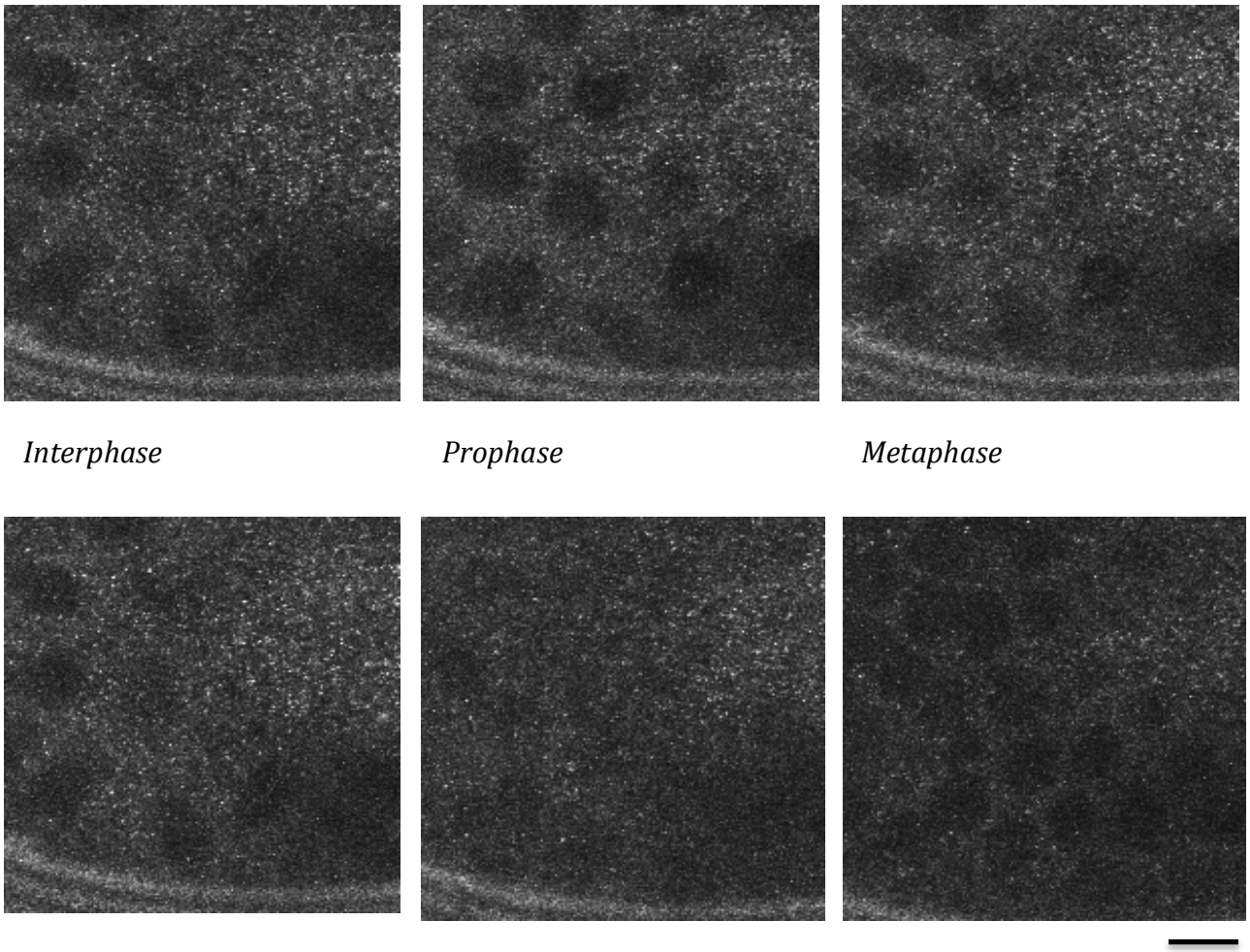


Figure 4.4: Representative images for interphase and different stages of mitosis of the GFP-CG2865 localization movie. The fusion protein seems to have a cytoplasmic localization and is excluded from the nuclear region during interphase and exit from mitosis when the nuclear membrane is intact. Scale bar = 10μm.

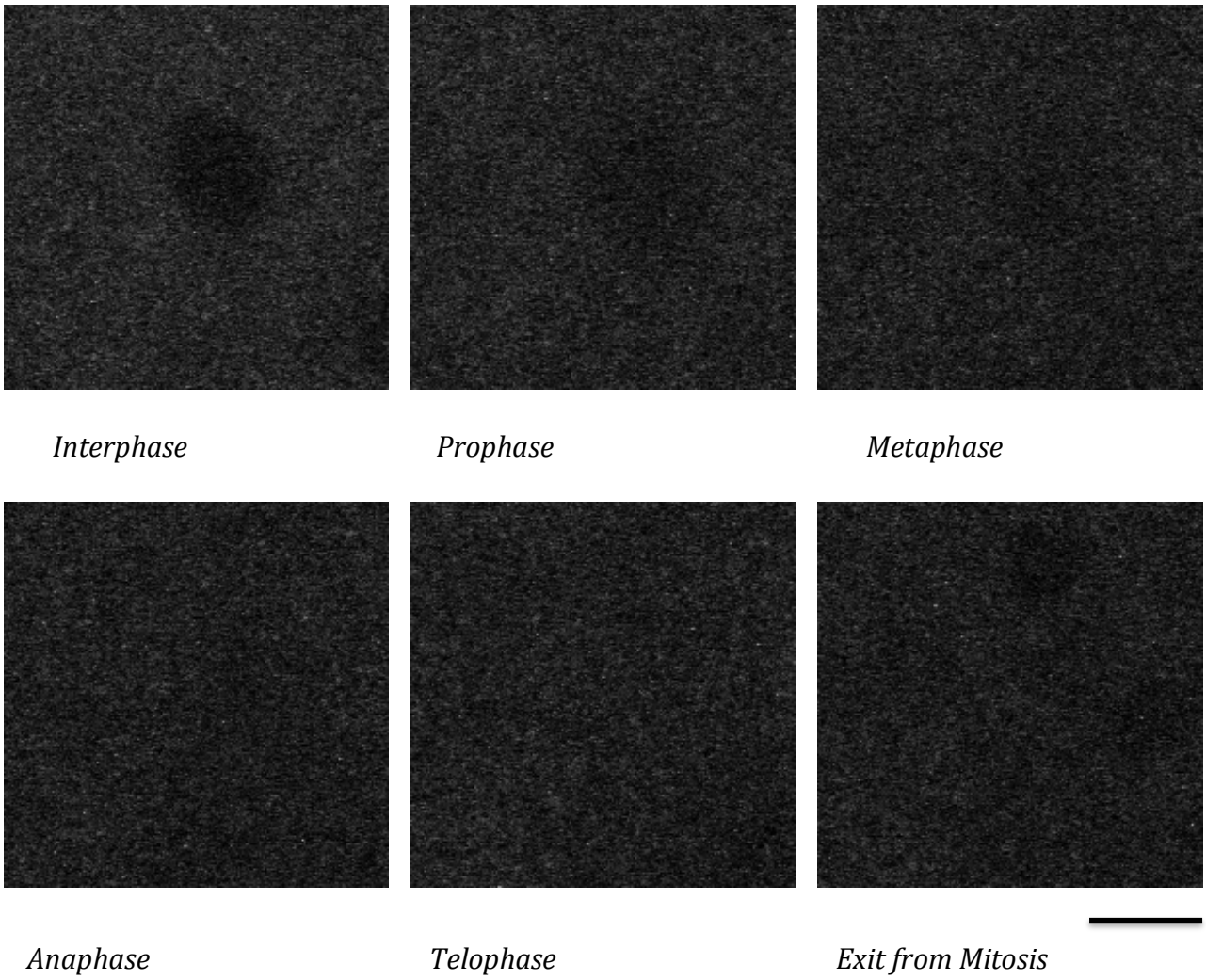


Figure 4.5: Representative images for interphase and different stages of mitosis of the GFP-CG5731 localization movie. The fusion protein seems to have a cytoplasmic localization and is excluded from the nuclear region during interphase and exit from mitosis when the nuclear membrane is intact. Scale bar = 10μm.

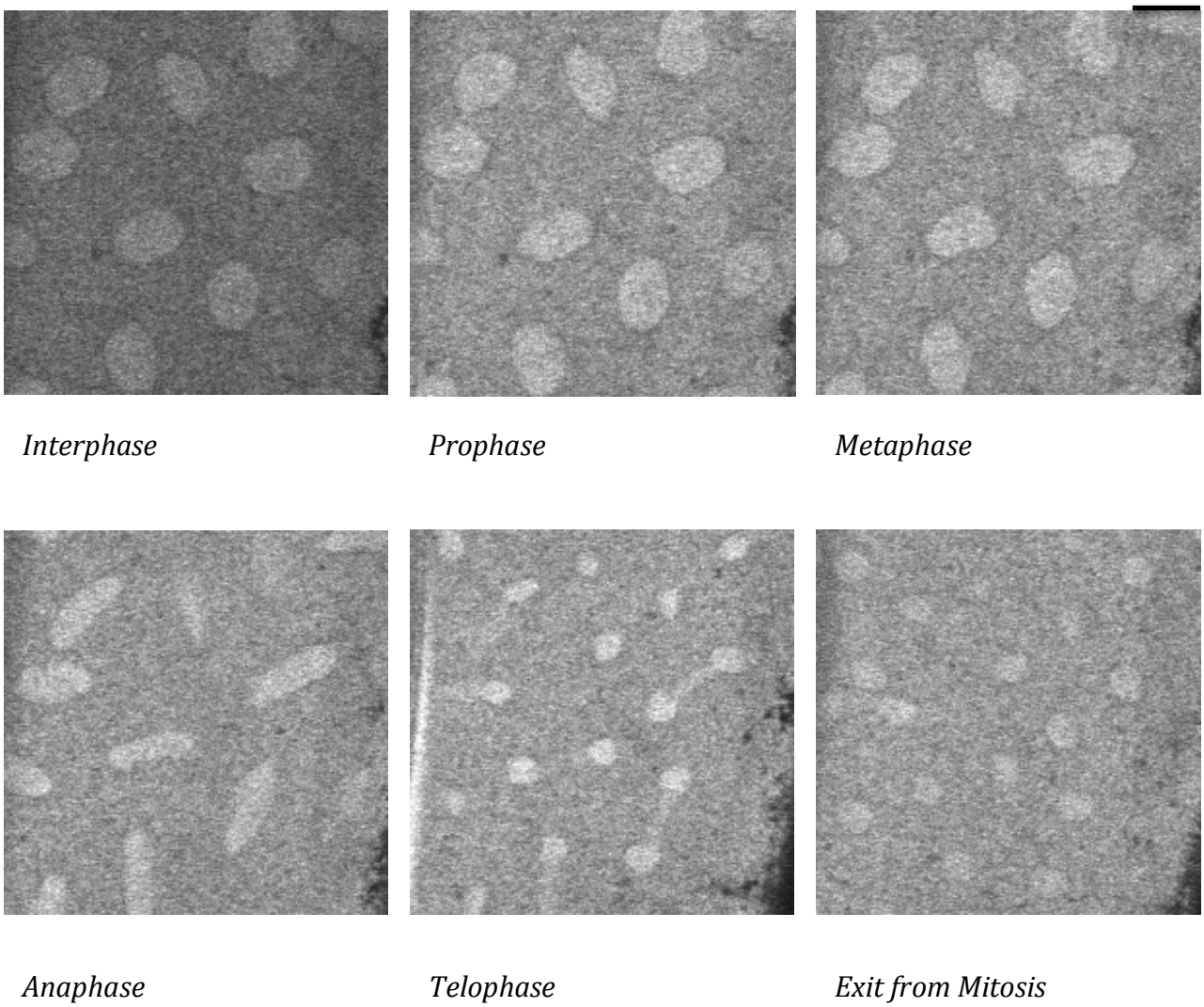


Figure 4.6: Representative images for interphase and different stages of mitosis of the GFP-CenG1A localization movie. The fusion protein seems to localize on the nucleus during interphase and the spindle during mitosis. Scale bar = 10μm.

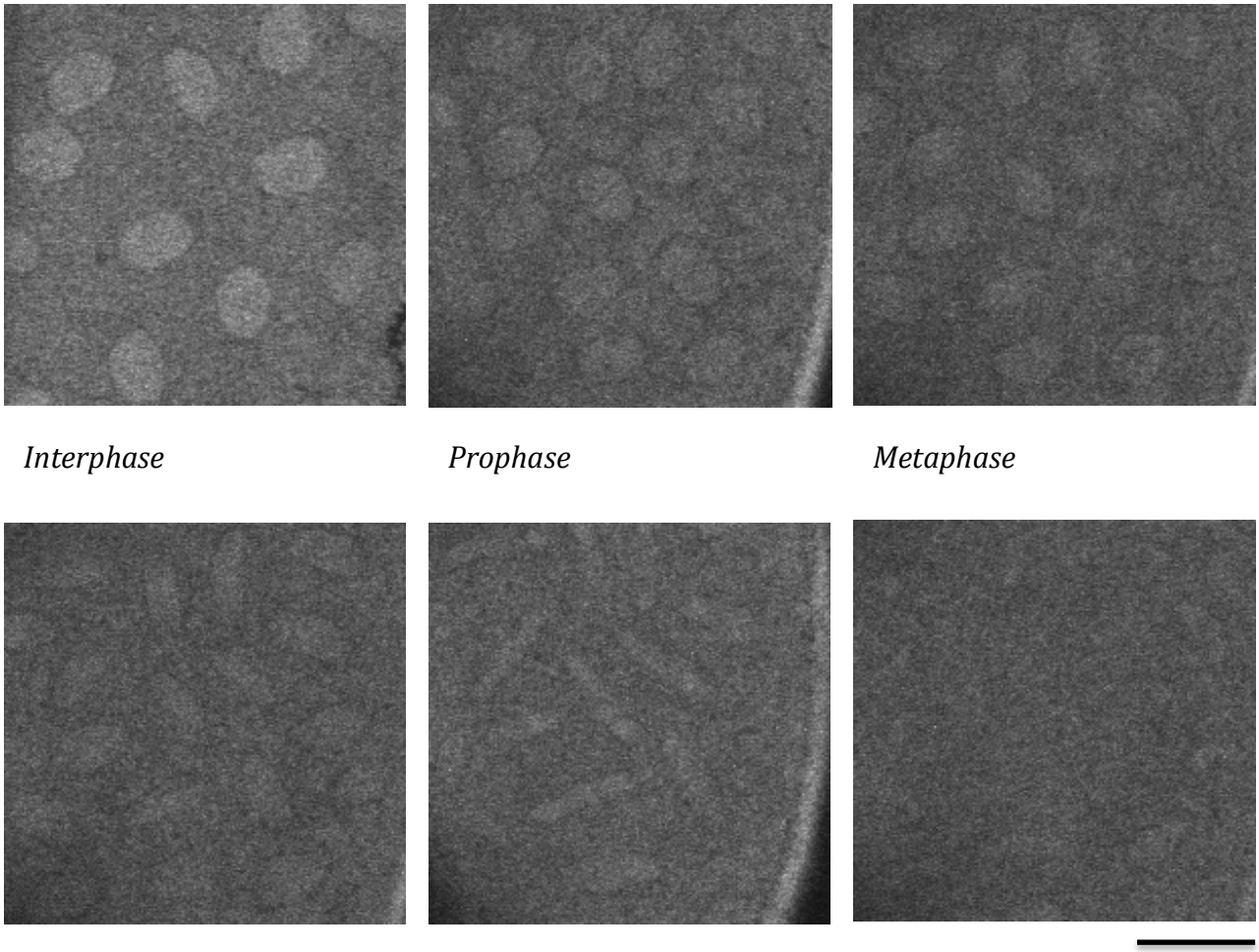


Figure 4.7: Representative images for interphase and different stages of mitosis of the GFP-CG5568 localization movie. The fusion protein seems to localize on the spindle during mitosis. Scale bar = 10μm.

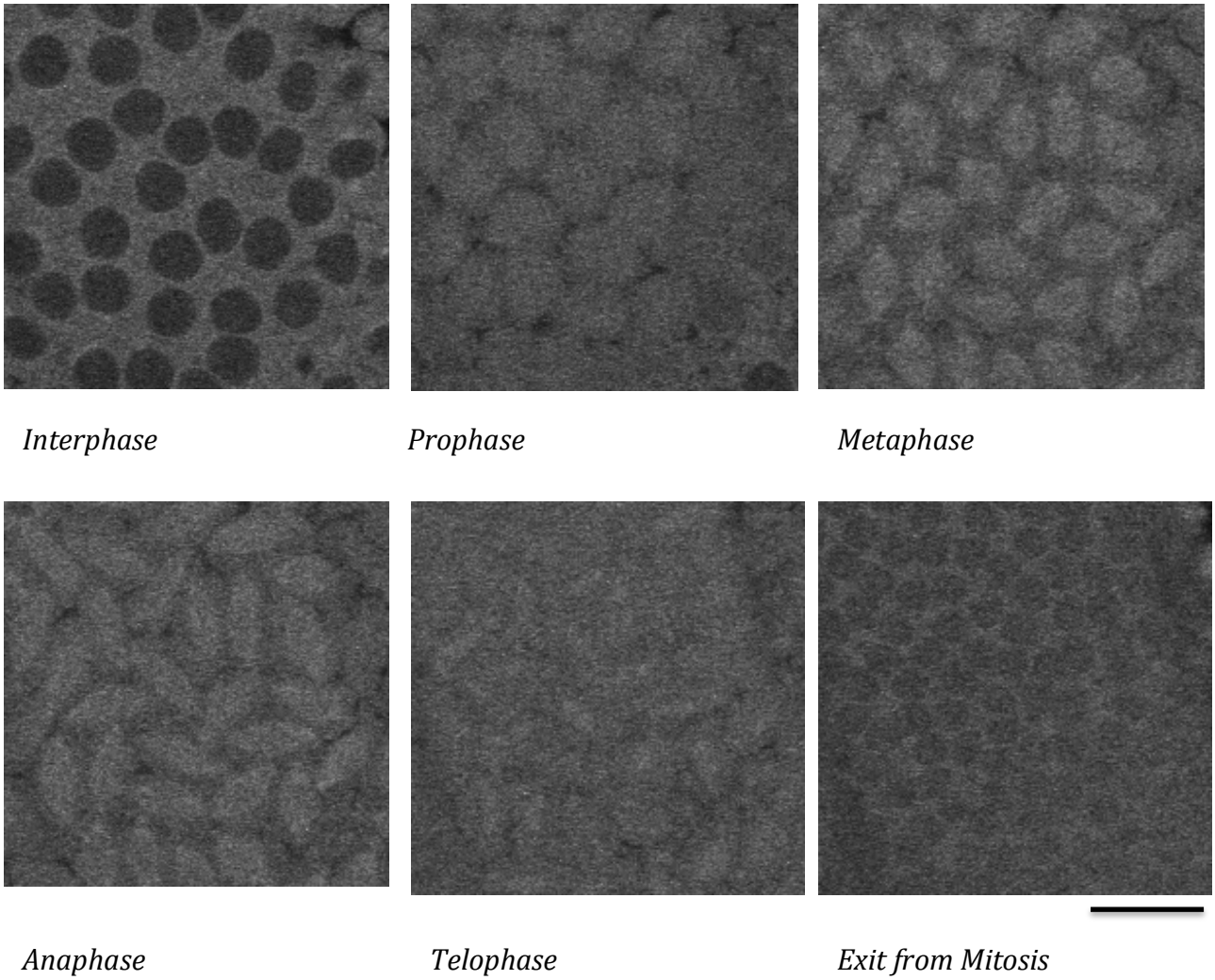


Figure 4.8: Representative images for interphase and different stages of mitosis of the GFP-Gus localization movie. The fusion protein seems to localize on the spindle microtubules during mitosis and remains cytoplasmic and excluded from the nuclear region during interphase. Scale bar = 10µm.

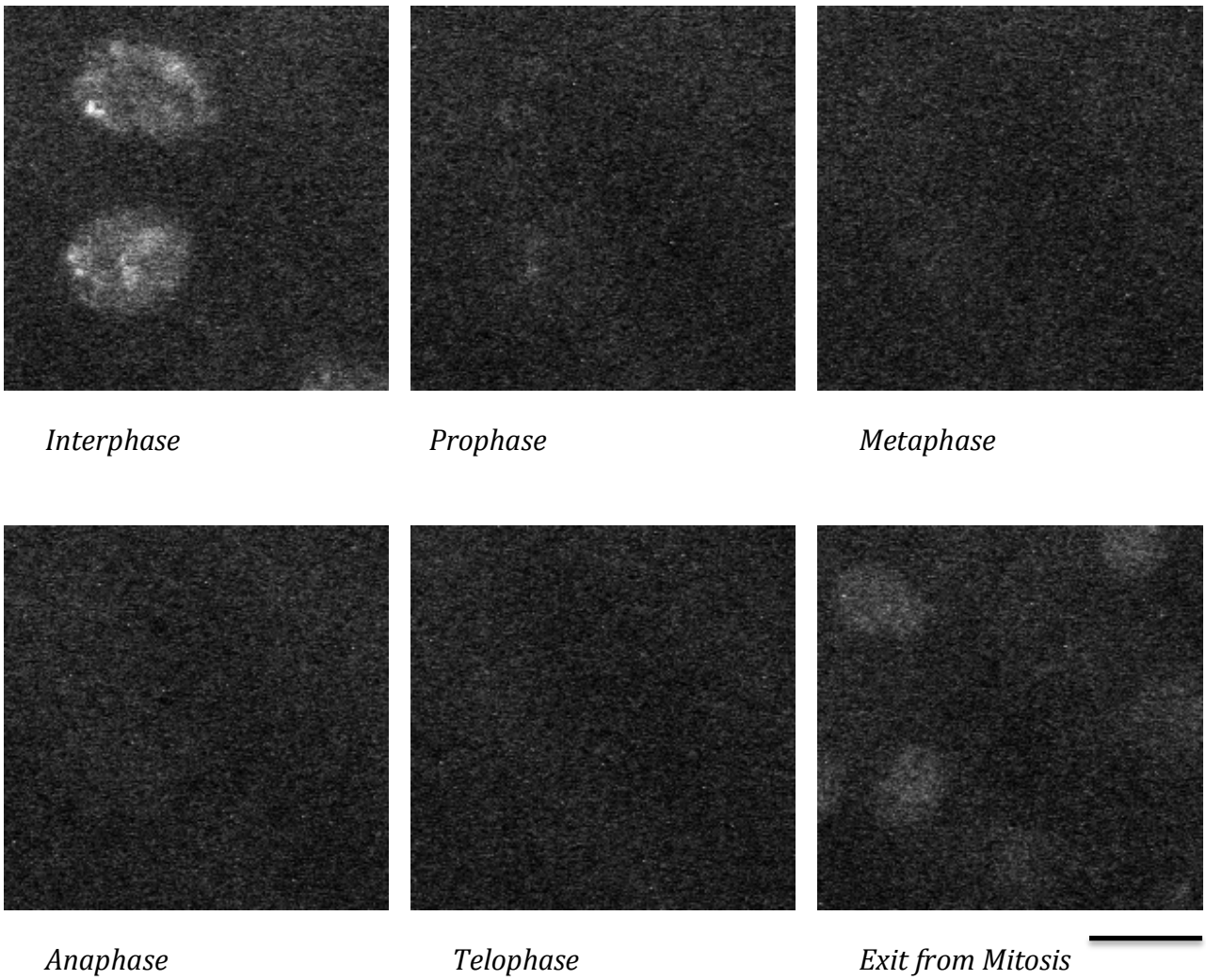


Figure 4.9: Representative images for interphase and different stages of mitosis of the GFP-PHDP localization movie. The fusion protein seems to localize on the chromatin in the nucleus during interphase, fading away upon entry into mitosis and re-appearing upon exit.

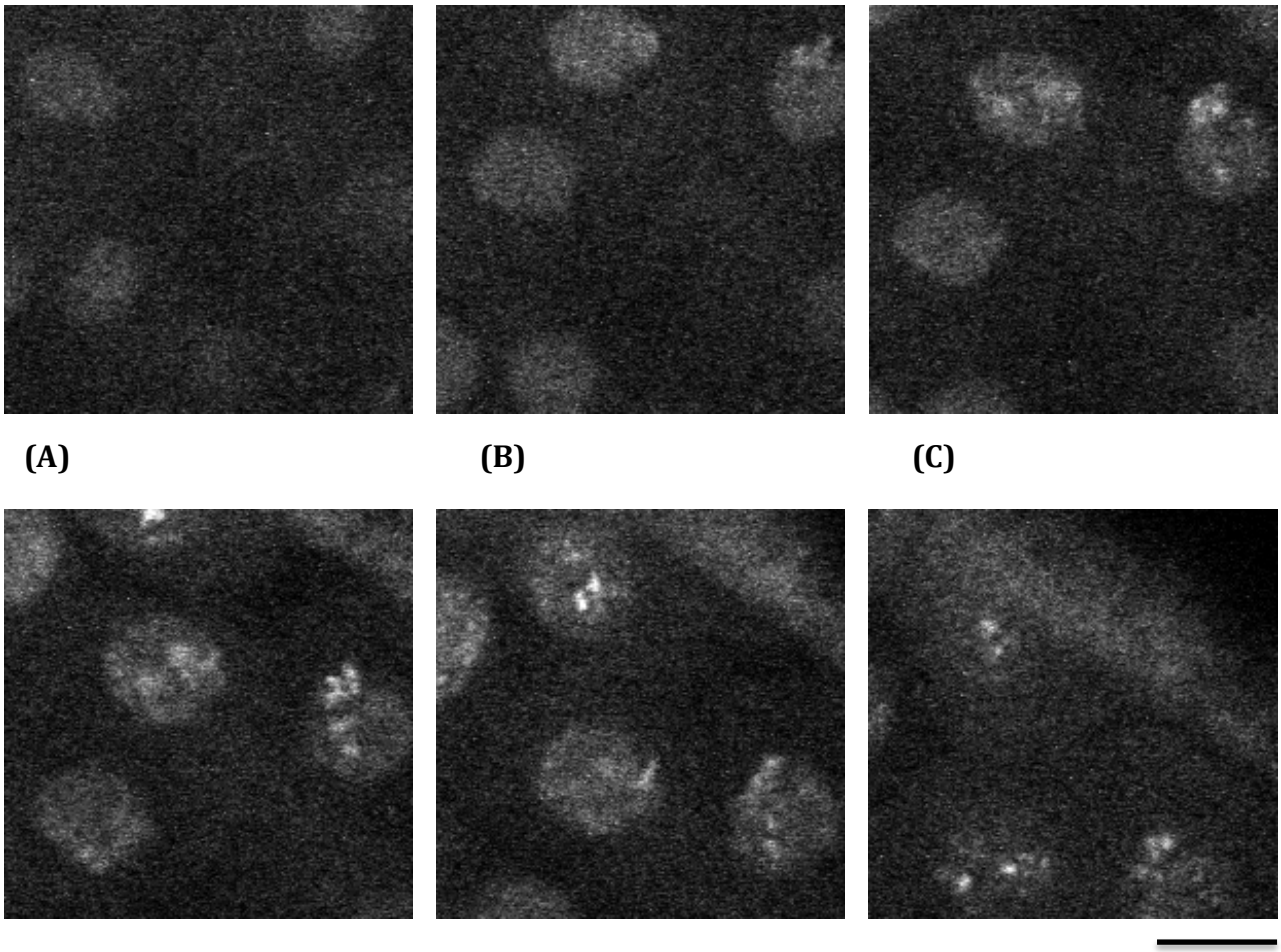


Figure 4.10: Seven representative images for interphase at regular intervals for the GFP-PHDP fly line. The intensity of the fusion protein localization seems to increase during the second half (D-F) and appears more dynamic at various bright spots. The GFP signal fades ultimately fades away upon entry into mitosis (G).

Scale bar = 10μm.

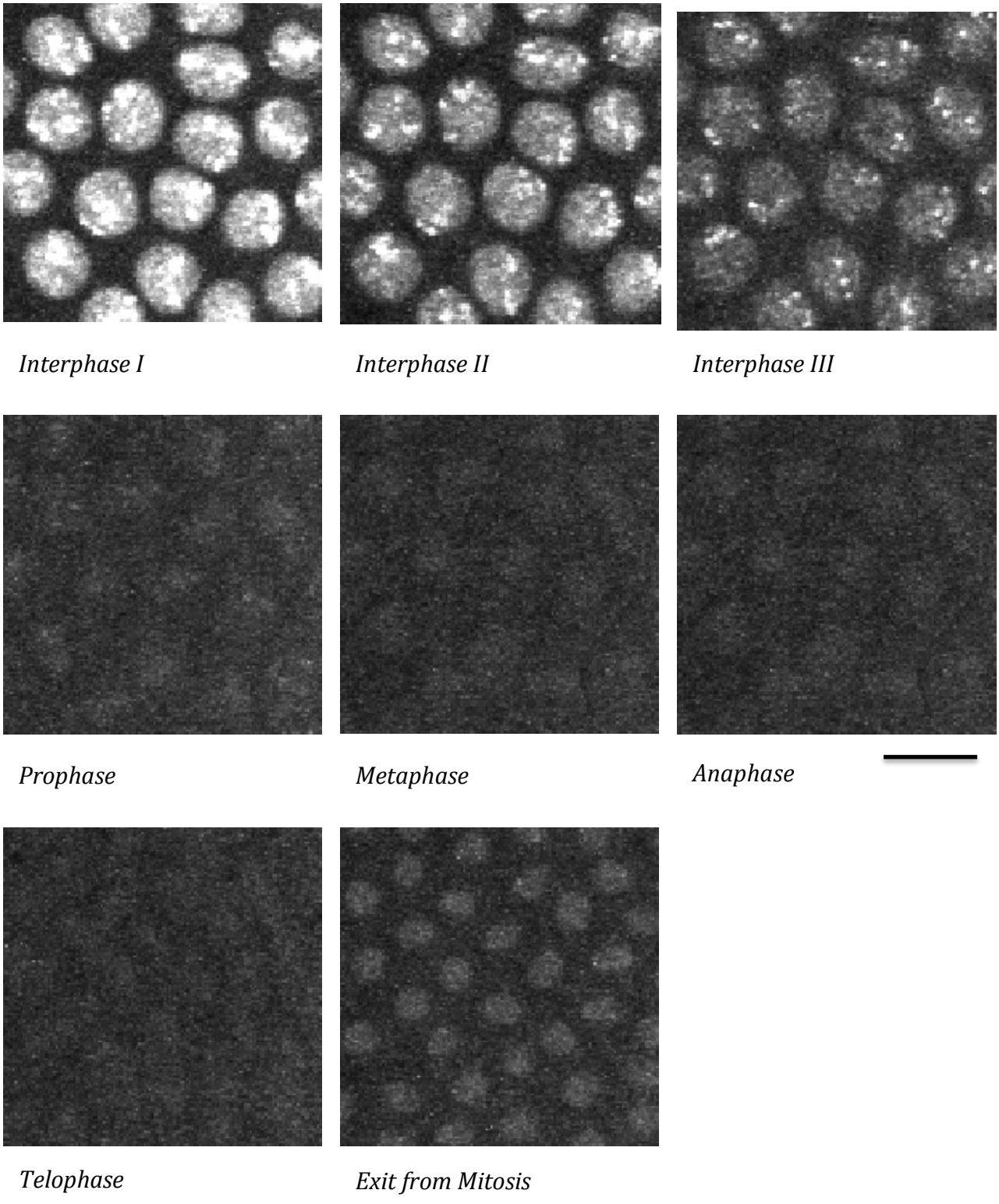


Figure 4.11: Representative images for interphase (3 equal intervals before entry into mitosis) and different stages of mitosis of the GFP-CG5708 localization movie. The fusion protein seems to localize on the chromatin in the nucleus during interphase, fading away upon entry into mitosis and re-appearing upon exit. Scale bar = 10μm.

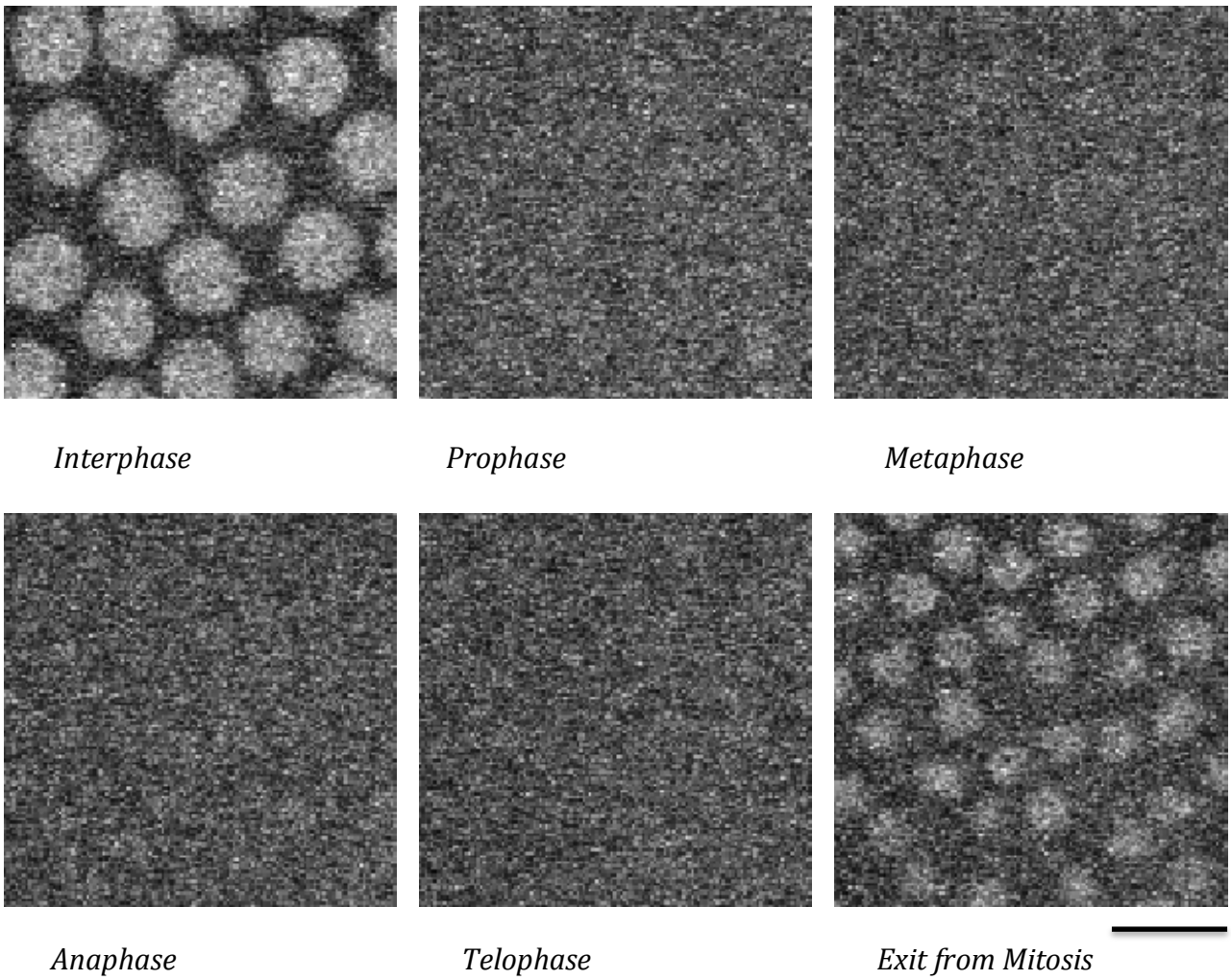


Figure 4.12: Representative images for interphase and different stages of mitosis of the YFP-Nej localization movie. The fusion protein seems to localize on the chromatin in the nucleus during interphase, fading away upon entry into mitosis and re-appearing upon exit. Scale bar = 10μm.

4.3.3 Proteomics analysis

In the proteomics analysis, I studied the *in vivo* interactions of six proteins from the subcluster in early embryos that were expressing each of their GFP-fused constructs. This approach included a co-immunoprecipitation (co-IP) protocol, which was used to pull down the transgene products from embryo extracts along with their interactors using anti-GFP antibodies bound to agarose beads. The co-IP experiment was followed by mass-spectrometry analysis to identify the bound proteins to the beads.

Mass spectrometry was outsourced to a university facility (as outlined in Section 4.2). During mass spectrometry unique peptides are identified based on their mass, which are then mapped to specific proteins in databases based on sequence and predicted mass. The raw output data includes the protein identifier of each protein, its number of peptides identified and the mass spectrometry score, which is a sum of the measure of several physico-chemical properties and gives a measure of the strength of the protein interaction. Results are considered valid if the bait protein itself is pulled down by the co-IP experiment and identified by mass spectrometry with a significant score.

The raw data was mapped to *Drosophila* gene symbols and classified based on a set criterion. Datasets were only considered if the experiment was able to pull down the fusion protein as a hit. Positive results were then filtered for error by removing a list of false positive (proteins bound to negative control GFP beads) developed and curated in the lab (Appendix X).

The remaining hits were finally classified into strong or weak hits based on their mass spectrometry scores, i.e. proteins with scores above 100 were considered strong and proteins with scores between 30 and 100 were considered weak hits. All hits that scored below 30 were disregarded.

Three out of the six proteins successfully met our criteria and identified weak and strong hits. Three transgene experiments could not pull down the protein under study and hence could not be considered as successful experiments.

Each experiment is discussed in more detail in the following section.

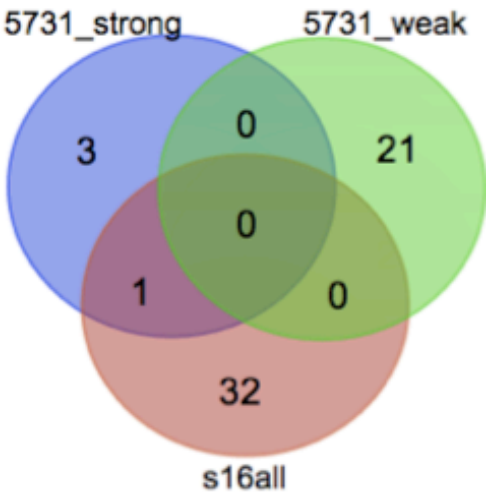
Transgenes with specific hits

The GFP-CG5731 expressing fly lines were tested twice using two different set of GFP-Trap® beads, i.e. the agarose or A-type and the magnetic or M-type. The magnetic beads were not able to pull down the bait protein. The agarose beads experiment was successful and produced the bait protein, CG5731, as a top hit along with other specific hits (Figure 4.13A). Two proteins from our sub-cluster are also pulled down (Mad2 and Gnu), but they scored just below our cut-off (~27). The Png protein from the PNG kinase complex in our subcluster also appears in the results but at a very low score (6.83). Altogether, CG5731 had 4 strong and 21 weak interactors. Three of the strong hits, i.e. betaTub97EF, Ubi-p5E and Msps, appear consistently in the other results, which suggest they are more likely to be non-specific hits. This leaves behind CG5731, the bait protein itself, as the strongest hit protein. The top ten hits along with their mass spectrometry scores and the overlap of the specific hits with the members of the subcluster are given in Figure 4.13.

(A)

Protein Hits	MS Scores
CG5731-PA	228.65
Ubi-p5E-PA	188.96
msps-PE	168.77
betaTub97EF-PB	150.67
Cp1-PC	98.61
CG30116-PC	93.99
CG8036-PD	90.21
CG6540-PA	65.38
spict-PA	65.32
eIF-2gamma-PC	61.82

(B)



(C)

Names	Total	Elements
5731_strong s16all	1	CG5731
5731_strong	3	betaTub97EF msps Ubi-p5E
5731_weak	21	CG30116 Osbp eIF-2gamma cup Srp54k CG9853 CG34113 CG32066 Dcr-1 CG11779 Stat92E CG2118 l(3)01239 FBpp0288803 spict Cp1 Mad1 Nup107 CG1598 CG6540 CG8036
s16all	32	CG13900 mad2 rev7 Orc4 CG13510 nej CG9331 CG9986 CenG1A CG15676 CG13625 plu png CG5708 SAK cdc14 MED15 PHDP gus Bap60 ald Akt1 pie CG8128 CG2865 CG8455 p53 CG11486 ptc E2f2 gnu CG5568

Figure 4.13: Mass spectrometry results for the pull down experiment of the CG5731 transgene. (A) The top 10 hits of the experiment (score above 100 are strong hits and scores 30-100 are weak hits). (B) A Venn diagram of the overlap between the strong and weak hits in the experiments and the members of the subcluster. There is a poor overlap between the two different datasets. (C) The protein composition of each category of the Venn diagram.

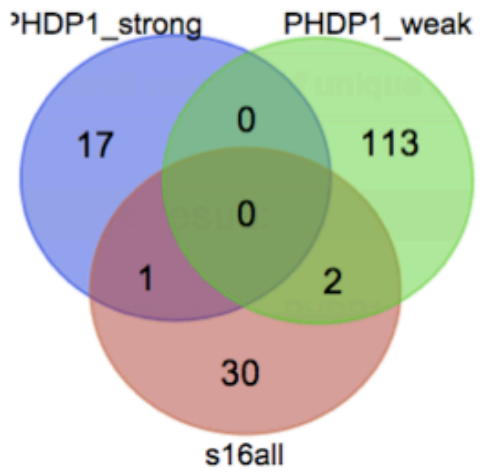
The GFP-PHDP expressing fly lines were analysed twice using GFP-Trap® agarose and magnetic beads. Both experiments successfully pulled down the bait protein (PHDP). These results are presented in Figures 4.14 and 4.15. Dhc64D and Ef1alpha100E were not found in the magnetic beads results and are therefore potential false positives. This made PHDP the first true positive based on the agarose beads experiment. This experiment pulled down 18 strong and 116 weak interactors for PHDP. These included 2 weak hits (Mad2 and Gnu) that are part of our subcluster (Figure 4.14B, C). Four members of subcluster-16 are also found in the potential non-specific hits lower on the ranked list of mass spectrometry hits, i.e. Gus (9.83), Bap60 (7.42) and CG5731 (7.06).

In the magnetic beads experiment, a very low overlap was found with the specific hits (strong and weak) of the agarose beads experiment, highlighting the heterogeneity and rate of error in the experimental setup. PHDP itself was again the only strong hit along with CG13900, a member of the subcluster, as the only weak hit (Figure 4.15B,C). Four proteins from the subcluster were found amongst the potentially non-specific interactors ranked below out cut-off, i.e. Nej (14.65), Gus (5.32), CenG1A (3.94) and CG5708 (2.34).

(A)

Protein Hits	MS Scores
Dhc64C-PA	446.42
Ef1alpha100E-PA	441.46
PHDP-PA	415.83
msps-PD	311.72
Ubi-p5E-PA	189.25
cup-PB	184.74
cdm-PA	168.06
CG5794-PD	159.95
coil-PD	136.50
rod-PA	132.12

(B)



(C)

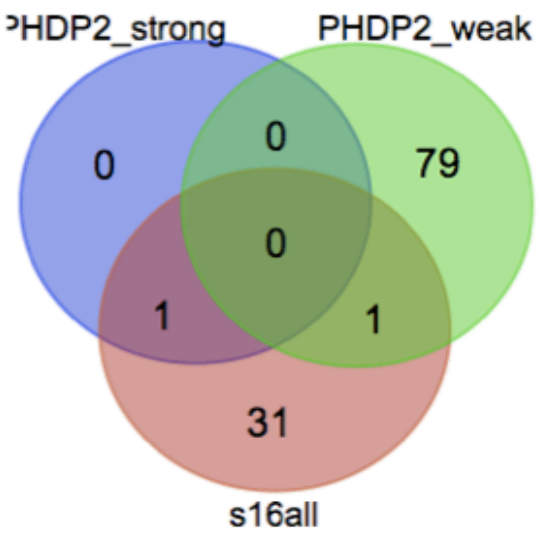
Names	Total	Elements
PHDP1_strong s16all	1	PHDP
PHDP1_weak s16all	2	mad2 gnu
PHDP1_strong	17	cdm rod Dhc64C Ubi-p5E fs(1)Ya CG5794 Jabba msps CG7546 hyd coil rhea Ef1alpha100E cup CG8184 CG3523 Mad1
PHDP1_weak	113	CG8503 muskelin Nup160 l(1)dd4 Cul-4 Upf1 sced CG17514 nxf2 par-1 CG11984 Nipsnap Msp-300 mask beta'Cop Nup93-1 CG9684 cnn tud Rme-8 Cdc27 mit(1)15 Lst8 glu Ge-1 mof DNApol-epsilon shtd Rab5 Rab1 lok Nup98-96 Cp190 Grip84 shot l(1)G0222 egl asp Srp54k gro CG9853 Mtor CG10254 Wdr82 APC7 CG8771 CG2246 Rpl135 mbo CG2469 cmet CG16940 Haspin CG8036 tacc Dys Mi-2 atms aub CG8963 ArfGAP3 E(bx) CG5604 Nup93-2 BicC Bap170 Bruce Nup214 Nipped Tor CG42388 dre4 CLIP-190 Cdc16 Stat92E CG31739 ctrip Grip75 Nup107 mahj woc CG6540 CG12512 Nipped-B pzg Nup133 CG41099 CG1657 Sec13 CG32649 CG9590 Cdc23 Ctf4 nop5 CG4119 CG32066 CG4538 lds CG11779 Cpsf100 ash2 fzy CG2118 smg spict Torsin ArfGAP1 spd-2 SNF4Agamma Nup75 lid Ulp1 jar lola
s16all	30	CG9986 CenG1A CG5731 gus p53 CG11486 ptc CG9331 CG13625 plu cdc14 Bap60 Akt1 pie CG2865 CG13900 Orc4 CG13510 CG15676 CG5708 ald CG8455 E2f2 CG5568 rev7 nej png SAK MED15 CG8128

Figure 4.14: Mass spectrometry results for the agarose beads pull down experiment of the PHDP transgene. (A) The top 10 hits of the experiment (score above 100 are strong hits). (B) A Venn diagram of the overlap between the strong and weak hits in the experiments and the members of the subcluster. There is a poor overlap between the two different datasets. And only mad2 and gnu from the subcluster are present. (C) The protein compositions of each category of the Venn diagram.

(A)

Protein Hits	MS Scores
PHDP-PA	104.30
CG10341-PA	97.44
CG9018-PA	96.81
Patr-1-PA	96.81
Bx42-PA	95.45
noi-PA	92.91
Snr1-PA	92.82
tgo-PA	90.06
CG1677-PA	88.63
nito-PB	87.76

(B)



(C)

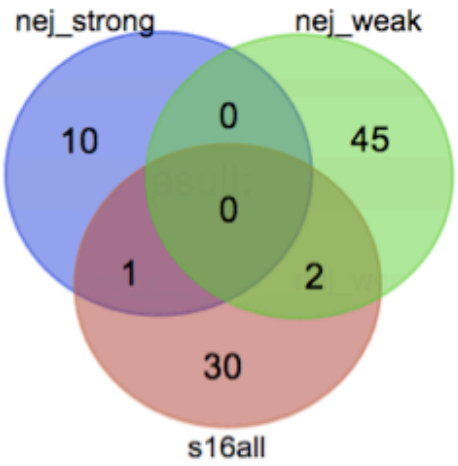
Names	Total	Elements
PHDP2_strong s16all	1	PHDP
PHDP2_weak s16all	1	CG13900
PHDP2_weak	79	Patr-1 atms CG3689 SC35 pea CG11089 PDZ-GEF Hakai vir Syp CG30122 noi Unr psq nito fd68A Iswi enc CG5913 Nipsnap Hrb87F NHP2 CG5726 pygo d4 Snr1 CG8507 CG4598 wds CG4896 U3-55K CTPsyn SmE CG7028 Pitslre CG6540 lark Prp38 sktl RAF2 tgo CG31548 pyd Prp3 SmF btz gw CG17168 Rbp2 CG1635 Nopp140 Fip1 cup Liprin-beta Bx42 CG10555 smt3 SmG x16 CG4119 mle sec24 scrib U4-60K Nup153 CBP QC CG10341 CG1677 CG10492 U2af38 CG10672 CG42303 Lasp CG2807 CG9018 CG32479 CG3605 Psi
s16all	31	mad2 Orc4 CG13510 CG9986 CenG1A CG15676 CG5731 CG5708 gus ald CG8455 p53 CG11486 ptc E2f2 gnu CG5568 rev7 CG9331 nej CG13625 plu png SAK cdc14 MED15 Bap60 Akt1 pie CG8128 CG2865

Figure 4.15: Mass spectrometry results for the magnetic beads pull down experiment of the PHDP transgene. (A) The top 10 hits of the experiment (score above 100 are strong hits). (B) A Venn diagram of the overlap between the strong and weak hits in the experiments and the members of the subcluster. There is a poor overlap between the two different datasets. And only Bap60 and CG13900 from the subcluster are present. (C) The protein compositions of each category of the Venn diagram.

(A)

Protein Hits	MS Scores
HDAC6-PE	179.63
Ubi-p5E-PA	160.84
Mad1-PA	159.36
nej-PE	157.94
CG11779-PB	150.13
DNApol-epsilon-PA	146.98
msi-PH	142.41
CG3532-PB	128.42
Gmap-PA	104.59
CG9590-PA	104.38

(B)



(C)

Names	Total	Elements
nej_strong s16all	1	nej
nej_weak s16all	2	mad2 CG11486
nej_strong	10	Gmap Ubi-p5E tral DNApol-epsilon CG3532 CG9590 msi-PH CG11779 HDAC6 Mad1
nej_weak	45	Crag Ahcy13 CG32230 GstO2 Rox8 eIF-2gamma Sap47 Nup50 stai CG5261 pum SdhB CG1578 CG7546 CG32500 Nup93-1 Nup107 Prat porin Rme-8 CaBP1 CG6540 coil gw me31B Pole2 Nup98-96 cbs cup Srp54k nop5 CG4119 CG32066 Nup153 smg spd-2 ArfGAP1 CG5174 CG1648 kuk Sec16 kis Vha68-3 CG17333 CG32521
s16all	30	CG13900 Orc4 CG13510 CG9986 CenG1A CG15676 CG5731 CG5708 gus ald CG8455 p53 ptc E2f2 gnu CG5568 rev7 CG9331 CG13625 plu png SAK cdc14 MED15 PHDP Bap60 Akt1 pie CG8128 CG2865

Figure 4.16: Mass spectrometry results for the pull down experiment of the nej transgene. (A) The top 10 hits of the experiment (score above 100 are strong hits). (B) A Venn diagram of the overlap between the strong and weak hits in the experiments and the members of the subcluster. There is a poor overlap between the two different datasets. And only mad2 and CG11486 from the subcluster are present. (C) The protein compositions of each category of the Venn diagram.

Embryos from the Nej-YFP lines were analysed using the agarose beads (Figure 4.16). The top hit was HDAC6, which is a histone deacetylase that is known to act on tubulin. Mad1, an interactor of Mad2 emerges as a strong hit followed by the bait protein, Nej, itself. Two other proteins from the subcluster came up as weak hits, i.e. Mad2 (80.61) and CG11486 (40.03) (Figure 4.16B, C). The Nej pull down experiment produced 11 strong and 47 weak interactors.

Transgenes with non-specific hits

The remaining three genes, Gus, CenG1A and CG5708, could not be successfully pulled down from embryos extracts. Embryos from the GFP-Gus and GFP-CenG1A fly lines were analysed using agarose beads and did not pull down the bait protein. Embryos from the GFP-CG5708 fly line were analysed using both the agarose and magnetic beads. The bait protein could not be pulled down during both attempts. The results therefore could not be considered valid for further analysis. Figure 4.17 gives the top ten hits from each experiment.

Although the GFP constructs in both lines were successfully expressed after the initial cross and their localization analysed through live imaging, one possible reason of the bait protein absent from the pull down experiment results in these cases can be the possibility of no expression of the transgene in the embryos during collection.

(A)

Gus hits	Score
CG6523-PA	29.82
Arf79F-PA	29.53
Irp-1B-PA	29.39
Atg7-PA	29.32
AP-1gamma-PA	29.28
RpL28-PB	28.97
Hsp23-PA	28.87
kuk-PA	28.80
p47-PA	28.70
CSN3-PA	28.29

(B)

5708 (M) hits	Score
Fmr1-PG	1916.28
Fmr1-PA	1817.64
nop5-PA	739.32
Chi-PA	738.21
Ef1alpha100E-PA	729.54
Nop60B-PA	564.51
nocte-PC	510.80
Spt5-PA	497.61
Spt6-PA	482.21
aft-PA	465.10

(C)

5708 (A) hits	Score
msps-PE	203.21
CG4119-PA	146.19
cup-PB	100.72
Nup98-96-PA	93.89
Mad1-PA	87.13
nop5-PA	86.06
CG11779-PB	77.07
Srp54k-PA	72.13
CG6540-PA	71.01
kuk-PA	68.77

(D)

CenG1A hits	Score
Rpt4R-PA	29.92
PP2A-B'-PC	29.77
smg-PB	29.45
Arf79F-PA	29.37
Akap200-PA	29.22
Nup50-PA	29.09
Crc-PA	28.61
RpL13A-PB	28.58
RpL9-PA	28.25
Rab14-PA	28.07

Figure 4.17: The top ten hits of the experiments, which did not meet our criteria and failed to come up with specific hits. These experiments included Gus, CenG1A and the two CG5708 experiments with agarose and magnetic beads separately.

The experimental validation approach, undertaken in this Chapter allowed for the analysis of not only the localization of proteins from subcluster-16 with respect to the mitotic apparatus in embryos, but also for conducting biochemical pull down experiments that identified real *in vivo* interactors. Both experiments used the same fly lines expressing their respective GFP-tagged versions of subcluster proteins as the source of embryos. The localization results for each of the eight proteins highlighted the diversity of functions these proteins possessed. These results strengthened the possibility that proteins that produced mitotic phenotypes in the *in vitro* analysis in Chapter 3, might be mitotic microtubule associated regulators, as opposed to being direct microtubule associated proteins. The pull down experiments highlighted the striking rate of false positives in subcluster-16, and identified real physical interactions that occur in the fly embryo. This narrowed down the focus on a module within subcluster-16 (discussed in more detail in the following text), which supported both by *in vitro* and *in vivo* validation, provides a stronger subject for further characterization.

Protein localization studies that employ strategies involving GFP-fusion constructs have several inherent limitations. Although most proteins are able to behave more or less naturally when fused to the 30 kDa GFP tag, many proteins can still have challenges in folding, functioning and localizing according to its native state in the cell. These possible effects are more pronounced when the GFP tag is fused to the protein end where specific folding, binding or catalytic domains are located. All these different variables come together and contribute to false positives and negatives, and therefore lack of overlap, in such

experiments. Further experiments with the GFP protein tagged to the alternate end of the protein of interest to validate results. Rescue experiments with GFP tagged proteins can also be undertaken to validate the function of these proteins.

The MAP prediction model constructed and implemented in Chapter 2 was built on the MAP interactome, which acted as a scaffold for integrating other datasets. Roughly one-third of the proteins in the MAP interactome itself came from biochemical datasets of microtubule-associated proteins. On the other hand, about two-thirds of the proteins were indirect interactors (proteins that bind at least two of these MAPs). This can partly explain the diversity of localizations in our *in vivo* analysis of the selected GFP-fused predicted mitotic MAPs. The fact that several of them do not localize to microtubules specifically, suggests that their mitotic role might be more regulatory rather than structural in nature with direct interactions to microtubules.

The proteomics analysis of the selected potential MAPs highlights the challenges these experiments pose for downstream analysis. With the three successful experiments for the Nej, CG5731 and PHDP transgenes, the rate of false positives can be clearly seen by the poor overlap of the interacting protein hits found in the co-immunoprecipitation (co-IP) experiments in this Chapter and the interactions found in subcluster-16 which are mostly obtained using the yeast-2-hybrid (Y2H) system. There are several inherent differences between the two systems (Y2H and co-IP), which are well documented, like the non-native nuclear environment of the Y2H system and the problem of noise in co-IP experiments. Yet these differences are still insufficient to draw a clear and complete picture of

the substantial false discovery rates. The problem of noise strikes us clearly in the experiments with the three transgenes with the key challenge being the identification of specific hits from the hundreds and in some cases couple of thousand hits reported in mass spectrometry data. The application of a filter of false positives and a cut-off for the mass spectrometry score that has been developed in our lab for *Drosophila* embryos based on the beads and protocol used produces robust results. This indicates that the dominant majority of interactions in subcluster-16 are indeed false positives or do not occur during the fast S/M cycles of the syncytial blastoderm embryo.

Furthermore, there are inherent possibilities of false negative interactions in the co-IP experiments. First, is the possibility of the GFP tag interfering in the correct binding and interaction of the protein as mentioned earlier. Although, interactions between subunits of strong complexes like the Augmin complex have been recapitulated in the lab using the co-IP protocol used in this study, the possibility of missing transient and/or low affinity interactions that are dependent on specific cellular environment and cell cycle stage is always present. Weak and transient interactions can be captured by using additional steps like cross-linking to stabilize and freeze such interactions in the intact cells prior to lysis (Guerrero et al. 2006; Drakas et al. 2005).

Taking the results of these proteomics analysis back to subcluster-16, we can see the recapitulation of some interactions and the addition of others (Figure 4.18). The green lines in the figure represent interactions that are found both in the subcluster-16 and in our proteomics results for Nej, CG5731 and PHDP and the

black line represent new interactions added on the basis of the proteomics analysis in this Chapter.

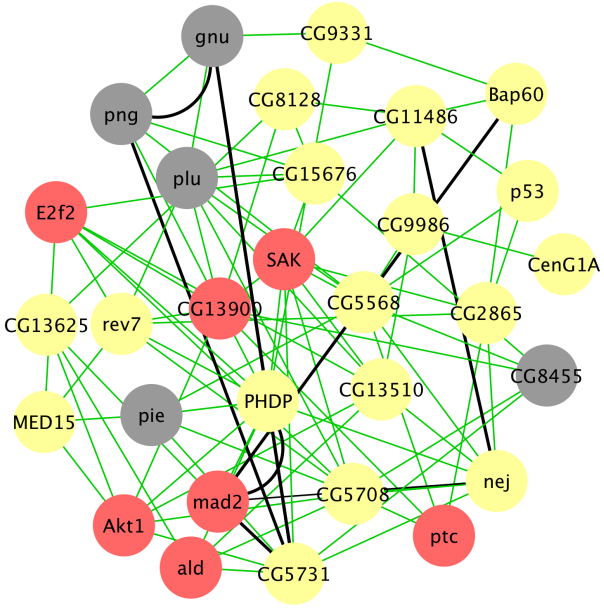


Figure 4.18: The module recapitulated from within the subcluster. Green lines depict interactions that occur both in the subcluster and the proteomics analysis in this Chapter. The black lines indicate new interactions added after the proteomics analysis. PHDP appears to be having the highest degree here (6), followed by CG5731 (5), nej (4) and mad2 (3).

If all our specific hits (weak and strong) from the four pull down experiments (1 each for nej and CG5731 and 2 for PHDP) are added, an increased connectivity of Mad2 with other members of the subcluster can be observed. PHDP and CG5731 can be seen as hubs in this module (with degrees of 5 and 6 respectively).

From the six well-characterized proteins and protein complexes with known functions in cell division and the cell cycle, only two remain, i.e. the PNG complex, which has a specific role in the G_2/M transition and Mad2, a well-studied spindle

assembly checkpoint protein. Both Mad2 and Png have regulatory roles and their higher connectivity to other proteins in the cluster, supports the possibility of our predicted MAPs having regulatory roles through indirect interactions with the microtubules and structural MAPs. The PNG complex was connected to two of the subcluster proteins, CG5731 and PHDP, via its subunit Gnu as shown by *in vivo* interactions. This result highlights the potential regulatory role of the PNG complex in the function of member proteins in the module.

Along with specific hits, our results also recapitulated interactions of PHDP with Gus, CenG1A and CG5708 but with very low mass spectrometry scores hidden in the noise. Adding these three proteins to the module shows no recapitulation of previous interactions, increasing the likelihood of these proteins being false positives in the subcluster and hence giving credence to our cut-off for the mass spectrometry scores.

Based on the results in hand, this module (Figure 4.18) makes an interesting target within subcluster-16 for further functional and biochemical characterization, starting with experiments like *in vivo* RNAi. Nej has been shown to localize to chromatin, similar to CG5708 and PHDP and further work in the lab (Alex Tyler and James Wakefield) has already shown that it does indeed have a role in mitosis. *In vivo* RNAi experiments in fly embryos have shown mitotic defects. Due to its co-precipitation of HDAC6, it was hypothesized that Nej might have a role as a tubulin acetylase. Recent results have shown this to be the case giving further credence to the prediction model and validating how this approach can be used to identify putative mitotic proteins.

4.4 Conclusions

In this Chapter I constructed GFP-fused transgenes for selected members of subcluster-16 (from Chapter 2) based on their expression in embryos and their connectivity with known mitotic and cell cycle members of the subcluster.

GFP-based localization studies of the products of these transgenes were conducted in 1-2 hour old fly embryos though live imaging and observed a diversity of localization for the different proteins. Out of the eight transgene products, three (GFP-Gus, GFP-CenG1A, GFP-CG5568) could bind directly to microtubules and spindles during mitosis with the rest producing a diverse set of results, including chromatin and cytoplasmic localizations. This indicated that although at least some of these proteins might potentially have a mitotic function, they clearly did not localize to microtubules in mitosis and in interphase.

Co-immunoprecipitation experiments with three proteins (Nej, CG5731 and PHDP) from fly embryo extracts were also conducted followed by mass spectrometry. This proteomics analysis demonstrated poor overlap with the interactions found in subcluster-16, highlighting the problem of noise in such experiments. Despite difficulty in pulling down baits from embryos expressing three other transgenes, I managed to recapitulate and highlight a module within the subcluster through the results of the four successful experiments.

The diversity of localizations obtained from the experiments in this Chapter and the fact that the proteins in the module revealed through the proteomics study

bind to two known mitotic and cell cycle regulators only, indicate that our predicted proteins might not be directly associated to microtubules as MAPs, but can possibly be microtubule associated regulators with a mitotic function as displayed by the GFP localizations and also the RNAi results from the previous Chapter.

In the final Chapter we summarize and reflect on our findings from the three Chapters as a whole and highlight future directions.

Chapter 5
Conclusions and future directions

The purpose of this thesis was to study mitosis in the model organism *Drosophila melanogaster*, from a network biology perspective. The primary aim was to develop and test a network-based prediction model that could integrate available data in public databases (like Flybase) and, based on that, predict potential mitotic proteins.

The work in this thesis does not stop at the statistical design, implementation and testing of the prediction model, which performed with 96% accuracy after 10-fold cross validation. But also takes it to the next level in the lab for an experimental validation of the function of these predicted proteins, both *in vitro* in S2 cells and *in vivo* in fly embryos.

Existing knowledge of mitotic proteins was used to create and extend the network of mitotic proteins, which was called the MAP interactome, and to select features that could increase the accuracy of our predictions. These *a priori* biological assumptions acted as *filters* and helped in the enrichment of putative mitotic proteins before screening for them, ultimately enhancing the signal-to-noise ratio, which is inherently low in biological systems (Ideker et al. 2011).

The main cellular machinery that carries out animal cell division (both mitosis and meiosis) is the mitotic spindle, which is made up of microtubules emanating from the centrosomes. In Chapter 2, I took a set of experimentally validated

microtubule associated proteins (MAPs) as the *seed* for building the MAP interactome. I then used two approaches to extend the network, i.e. by transferring homologs and interologues from MAP datasets in four other species and by adding indirect interactors (proteins that bind two or more of these MAPs) from the *Drosophila* proteome.

Once an enriched set of known and potential mitotic MAPs was obtained (the MAP interactome), I gathered several layers of complementary data for the proteins in the network from public databases. These were genetic interactions, gene expression, protein domain, protein sequence and genome-wide RNAi datasets that were integrated as features and used to fit a prediction model. The integration of complementary datasets also helped in the increase of signal-to-noise ratio. This model was finally used to rank all the proteins based on the likelihood of their role in mitosis.

The approach taken to design and implement the prediction model will introduce several limitations into our study. Firstly, the network created in this study and the proteins tested are exclusively based on MAP data, i.e. all nodes in the network are either known MAPs (in *Drosophila* and other species) or interactors of these MAPs. This will inevitably exclude mitotic regulators that do not bind directly to microtubules and function from a distance in the network through indirect interactors.

Secondly, the interactions in our network are based on data derived from the yeast-2-hybrid (Y2H) system, which has a high false discovery rate and introduces the possibility of a large number of false positive interactions. This

not only adds more error to the members of the network (indirect interactors) but also to certain features of the model, like mitotic neighbours. As already established, Y2H-based protein interaction networks are very sparse and far from complete, which highlights the possibility of false negatives in our network as well.

Thirdly, as evident from Chapter 2, the coverage of all features varies between the features themselves and also between the training and prediction sets. The prediction set compared to the training set had a lower coverage, contributing to a lot of proteins with no data. This means the model was fitted using relatively more complete coverage, but has the disadvantage of being another source of false positive predictions in our model. Future implementations of this model will have the advantage of better coverage as these genome-wide (omics) datasets become more complete.

With the intention of experimental validation, a select group of predicted mitotic MAPs, the top 100 proteins of the prediction set were chosen. Chapter 2 ends with a more detailed analysis of the top 100 hits not only in the MAP interactome but from a broader neighbourhood perspective in the *Drosophila* proteome. An interconnected subcluster (named subcluster-16) was identified within this set, which was selected for functional and biochemical analysis in the lab.

In Chapter 3, I used an RNAi-based gene knockdown screen to test each of the thirty-three members of subcluster-16. RNAi provides a fast, cheap efficient and a direct approach to gain insights into the function of a protein.

Along with these advantages, the RNAi approach also carries certain limitations. Several studies have highlighted the poor overlap between datasets from different genome-wide studies (Mathey-Prevot & Perrimon 2006). Although these false discovery rates are inherent to large-scale studies, smaller low-throughput experiments also face the possibilities of false positives and false negatives in their results.

The main source of false negatives in RNAi experiments is inefficient knock down or subtle phenotypes, which are missed by the screen. Although an enriched set of potential mitotic proteins was tested (unlike a genome-wide screen), with positive controls and a measure of transcript levels after dsRNA-treatment, the chances of false negatives can still not be ruled out. The ideal approach will be to conduct independent experiments and test if the phenotype is reproduced.

False positive results are a source of greater concern especially in validation screens as in Chapter 3. The biggest source of false positives is off-target effects or OTEs. These are caused by the unintended knockdown of multiple genes based on sequence similarity to the dsRNA fragments. The ideal way to address this problem is to test the same gene in a screen with two different dsRNA molecules. Although appropriate precautions were taken during the primer design of dsRNA for each experiment, both our replicates used the same dsRNA. This is potentially one of the main weaknesses in our study.

The very high hit-rate in the RNAi screen (compared to undirected screens) can be also be accounted for by other inherent biases in the approach, i.e. the enrichment of MAPs which are more likely to be mitotic in function and the

presence and large number of interactions of known mitotic proteins like SAK, Mad2 and numerous Cyclins and Cdk proteins. Furthermore, like with any other screen, observer bias cannot be ruled out as well regardless of appropriate controls in each batch of experiment and the double-blind analysis of the replicate.

In Chapter 4, I constructed GFP-tagged transgenes for a subset of the proteins in sucblsuter-16. These constructs were expressed in fly embryos for *in vivo* localization studies. As discussed in detail, these GFP-tagged clones allow for a variety of experiments but as with any experimental setup do come with certain caveats.

Although most fused proteins are able to localize in their natural way, there are chances of a fused protein not being able to fold, localize and/or function, as they would do in their native state. Follow up experiments like tagging alternate ends of the proteins and rescue experiments with GFP-fused version of proteins can be undertaken to ascertain the exact localization of proteins. To confirm the mitosis-related localizations reported in this thesis, recapitulating them with GFP-fusion proteins, which have the GFP tags at the C-terminus would be an appropriate experiment.

As part of our experimental *in vivo* validation approach, Chapter 4 also included a proteomics analysis using a co-immunoprecipitation assay that pulls down the GFP-fused construct and its interactors from embryo extracts which can then be resolved and identified using mass spectrometry. The rationale here was to obtain an estimate of false positives in the Y2H-based interactions in subcluster-

16 and identify real *in vivo* interactors. As evident from the results, there was a very low overlap between the two types of interactions, indicating that most interactions in the subcluster were not occurring in the fly embryo.

The proteomics analysis of the selected proteins and the subsequent validation of real interactions highlight a module within subcluster-16. This highly connected module contains eight proteins, which contains six putative mitotic proteins along with two members of the known PNG complex from outside the MAP network. This result is substantiated by the results of the RNAi screen in Chapter 3, which reports them as hits. Four of these genes belong to Class I proteins, which have spindle phenotype, and four of them to Class III, which have spindle and centrosome phenotypes.

In summary, this thesis demonstrated through the design and implementation of the MAP interactome and prediction model, that by applying prior biological knowledge and by integrating complementary 'omics' datasets the signal to noise ratio in such complex biological problems can be increased and high accuracy predictions about gene function can be obtained. This suggests that with the current body of data, we might be able to predict higher-level biological processes, if not the lower level molecular function of uncharacterized proteins.

The experimental validation through the RNAi screen by the validation of ~80% predicted proteins having mitotic function in S2 cells gives credence to the accuracy of the MAP prediction model. The hit-rate in this bioinformatics-guided screen is substantially higher than large-scale genome-wide screens (less than 5%), which consume a lot of time and resource.

Future Work

As mentioned in the previous section, future implementations of this model will no doubt benefit from better coverage as more complete omics datasets are published. One recent example is the modENCODE dataset (Roy et al. 2010), which produced a better coverage of gene expression data. Apart from large-scale screens, low throughput experiments will also continue to fill in the gaps in our knowledge of features like, protein and genetic interactions and gene knockdown data.

One interesting direction for further work can be the testing of new features for integration into the prediction model. This can include novel datasets like protein abundance (Beck et al. 2011), which could identify false 'hubs' or sticky proteins, and epigenetics data, which would bring in another layer of interactions between these proteins. Given the known false discovery rates of Y2H data that has also been observed in this thesis, it would be interesting to employ AP-MS or co-complex data as separate features, e.g. mitotic neighbours based on co-complex data only. This can readily be done with the recently published proteome-wide co-complex screen in *Drosophila* S2 cells (Guruharsha et al. 2011).

With more complete data and better coverage, it would be interesting to test how the model performs in predicting the protein function at a lower level, for example, identifying mitotic complexes in which the protein might have a role, or with a temporal dimension, identifying the phase of mitosis during which the protein might function.

Implementation of the model in a different biological process altogether, like the actin cytoskeleton organization, the centrosome cycle or a developmental process, can be another interesting direction for future work.

From an experimental perspective, the nej/mad2/PHDP/CG5731 complex within subcluster-16 needs further characterization. With experiments already producing encouraging results with Nej, which suggest a potential acetylation function in mitosis and Mad2 being well characterized spindle assembly checkpoint protein, it would be interesting to understand the molecular function of CG5731 and PHDP. *In vivo* RNAi in fly embryos against these proteins and GFP localization experiments in mutant backgrounds would provide better insights at this stage.

Finally, it will be worth validating the remaining RNAi hits and analysing the remaining high-scoring proteins from the predicted list of mitotic proteins for similar modules and complexes, especially in the in the top 200-300 range based on their scores and completeness of data.

"The end of the screen is the beginning of the experiment."

- Boutrus and Ahringer, 2010

Bibliography

Adams, M.D., 2000. The Genome Sequence of Drosophila melanogaster. *Science*, 287(5461), pp.2185–2195.

Agapakis, C.M. & Silver, P. a, 2009. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular bioSystems*, 5(7), pp.704–13.

Agarwal, S. et al., 2010. Revisiting Date and Party Hubs: Novel Approaches to Role Assignment in Protein Interaction Networks P. E. Bourne, ed. *PLoS Computational Biology*, 6(6), p.e1000817.

Akaike, H., 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6), pp.716–723.

Akhmanova, A. & Steinmetz, M.O., 2008. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nature reviews. Molecular cell biology*, 9(4), pp.309–22.

Albert, R. & Barabassi, A.-L., 2002. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.*, 74, pp.47–97.

Albert, R., Jeong, H. & Barabási, A.-L., 2000. Error and attack tolerance of complex networks. *Nature*, 406(6794), pp.378–82.

Alon, U., 2007. Network motifs: theory and experimental approaches. *Nature reviews. Genetics*, 8(6), pp.450–61.

Althoff, F., Karess, R.E. & Lehner, C.F., 2012. Spindle checkpoint-independent inhibition of mitotic chromosome segregation by Drosophila Mps1. *Molecular biology of the cell*, 23(12), pp.2275–91.

Altschul SF, Gish W, Miller W, Myers EW, L.D., 1990. Basic local alignment search tool. *J Mol Biol*, 215(3), pp.403–10.

Amos, L. a & Löwe, J., 1999. How Taxol stabilises microtubule structure. *Chemistry & biology*, 6(3), pp.R65–9.

Amos, L.A., 2005. Tubulin and Microtubules. *Encyclopedia of Life Sciences*, pp.1–7.

Aranda, B. et al., 2010. The IntAct molecular interaction database in 2010. *Nucleic acids research*, 38(Database issue), pp.D525–31.

Arbeitman, M.N. et al., 2002. Gene expression during the life cycle of Drosophila melanogaster. *Science (New York, N.Y.)*, 297(5590), pp.2270–5.

Atsumi, S., Hanai, T. & Liao, J.C., 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 451(7174), pp.86–9.

Bader, G.D., 2003. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Research*, 31(1), pp.248–250.

Bakal, C. et al., 2007. Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science (New York, N.Y.)*, 316(5832), pp.1753–6.

Barabási, A.-L., Gulbahce, N. & Loscalzo, J., 2011. Network Medicine: A Network-based Approach to Human Disease. *Nature Reviews Genetics*, 12(1), pp.56–68.

Barabási, A.-L. & Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nature reviews. Genetics*, 5(2), pp.101–13.

Bebek, G. et al., 2012a. Network biology methods integrating biological data for translational science. *Briefings in bioinformatics*, 13(4), pp.446–59.

Bebek, G. et al., 2012b. Network biology methods integrating biological data for translational science. *Briefings in bioinformatics*, 13(4), pp.446–59.

Beck, M. et al., 2011. The quantitative proteome of a human cell line. *Molecular systems biology*, 7(549), p.549.

Belmont, L.D. & Mitchison, T.J., 1996. Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules. *Cell*, 84(4), pp.623–31.

Bettencourt-Dias, M. et al., 2005. SAK/PLK4 is required for centriole duplication and flagella development. *Current biology : CB*, 15(24), pp.2199–207.

Bettencourt-Dias, M. & Glover, D.M., 2007. Centrosome biogenesis and function: centrosomics brings new understanding. *Nature reviews. Molecular cell biology*, 8(6), pp.451–63.

Blow, N., 2009. Transcriptomics - Technology Feature. *Nature*, 458(March), pp.239–242.

Boone, C., Bussey, H. & Andrews, B.J., 2007. Exploring genetic interactions and networks with yeast. *Nature reviews. Genetics*, 8(6), pp.437–49.

Boxem, M. et al., 2008. A protein domain-based interactome network for C. elegans early embryogenesis. *Cell*, 134(3), pp.534–45.

Braun, P., 2012. Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays. *Proteomics*, 12(10), pp.1499–518.

Breitkreutz, B.-J., Stark, C. & Tyers, M., 2003. Osprey: a network visualization system. *Genome biology*, 4(3), p.R22.

Bu¨rckstu¨mmer, T. et al., 2006. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. , 3(12).

Buffin, E., Emre, D. & Karess, R.E., 2007. Flies without a spindle checkpoint. *Nature cell biology*, 9(5), pp.565–72.

Calderwood, M. a et al., 2007. Epstein-Barr virus and virus human protein interaction maps. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18), pp.7606–11.

Carvalho-Santos, Z. et al., 2010. Stepwise evolution of the centriole-assembly pathway. *Journal of cell science*, 123(Pt 9), pp.1414–26.

Caviston, J.P. & Holzbaur, E.L.F., 2006. Microtubule motors at the intersection of trafficking and transport. *Trends in cell biology*, 16(10), pp.530–7.

Ceol, A. et al., 2010. MINT, the molecular interaction database: 2009 update. *Nucleic acids research*, 38(Database issue), pp.D532–9.

Chait, B.T., 2011. Mass Spectrometry in the Postgenomic Era.

Chen, E.Y. et al., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14, p.128.

Chen, T.-C. et al., 2009. Cliques in mitotic spindle network bring kinetochore-associated complexes to form dependence pathway. *Proteomics*, 9(16), pp.4048–4062.

Chen, Y.-J. et al., 2009. A conserved phosphorylation site within the forkhead domain of FoxM1B is required for its activation by cyclin-CDK1. *The Journal of biological chemistry*, 284(44), pp.30695–707.

Chudakov, D.M., Lukyanov, S. & Lukyanov, K. a, 2005. Fluorescent proteins as a toolkit for in vivo imaging. *Trends in biotechnology*, 23(12), pp.605–13.

Chuong, S.D.X. et al., 2004. Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells. *Molecular & cellular proteomics : MCP*, 3(10), pp.970–83.

Clemens, J.C. et al., 2000. Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp.6499–503.

Colombié, N. et al., 2008. Dual roles of Incenp crucial to the assembly of the acentrosomal metaphase spindle in female meiosis. *Development (Cambridge, England)*, 135(19), pp.3239–46.

Conde, C. & Caceres, A., 2009. Microtubule assembly, organization and dynamics in axons and dendrites. *Nat Rev Neurosci*, 10(5), pp.319–332.

Consortium, T.F., 2003. The FlyBase database of the Drosophila genome projects and community literature. *Nucleic Acids Research*, 31(1), pp.172–175.

Consortium, T.G.O., 2000. Gene Ontology : tool for the unification of biology. *Nature genetics*, 25(may), pp.25–29.

Costa, L. et al., 2008. Characterization of Complex Networks : A Survey of measurements. *Advances in Physics*, 56(1), pp.167–242.

Coudreuse, D. & Nurse, P., 2010. Driving the cell cycle with a minimal CDK control network. *Nature*, 468(7327), pp.1074–1079.

Coulombe, B., Blanchette, M. & Jeronimo, C., 2008. Steps towards a repertoire of comprehensive maps of human protein interaction networks: the Human Proteotheque Initiative (HuPI). *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 86(2), pp.149–56.

Cox, J. & Mann, M., 2011. Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry*, 80, pp.273–99.

Dehmelt, L. & Halpain, S., 2004. Protein family review The MAP2 / Tau family of microtubule-associated proteins. , pp.1–10.

DGRC, 2013. DGRC - Drosophila Genomics Resource Centre, Indiana University, Bloomington. Available at: https://dgrc.cgb.indiana.edu/cells/.

Dietzl, G. et al., 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature*, 448(7150), pp.151–6.

Dixon, S.J. et al., 2009. Systematic mapping of genetic interaction networks. *Annual review of genetics*, 43, pp.601–25.

Dobbelaere, J. et al., 2008. A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in Drosophila. *PLoS biology*, 6(9), p.e224.

Dobrin, R. et al., 2009. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome biology*, 10(5), p.R55.

Drakas, R., Prisco, M. & Baserga, R., 2005. A modified tandem affinity purification tag technique for the purification of protein complexes in mammalian cells. *Proteomics*, 5(1), pp.132–7.

Dreze, M. et al., 2010. *High-quality binary interactome mapping.* 2nd ed., Elsevier Inc.

Dzhindzhev, N.S. et al., 2005. Distinct mechanisms govern the localisation of Drosophila CLIP-190 to unattached kinetochores and microtubule plus-ends. *Journal of cell science*, 118(Pt 16), pp.3781–90.

Ebrahimi, S. & Gregory, S.L., 2011. Dissecting protein interactions during cytokinesis. *Communicative & integrative biology*, 4(2), pp.243–4.

Eckner, R. et al., 1994. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes & Development*, 8(8), pp.869–884.

Elfring, L.K. et al., 1997. Drosophila PLUTONIUM protein is a specialized cell cycle regulator required at the onset of embryogenesis. *Molecular biology of the cell*, 8(4), pp.583–93.

Enright, a J., Van Dongen, S. & Ouzounis, C. a, 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, 30(7), pp.1575–84.

Erdos, P. & Renyi, A., 1960. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5, pp.17–61.

Feldman, I., Rzhetsky, A. & Vitkup, D., 2008. Network properties of genes harboring inherited disease mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11), pp.4323–8.

Fenn, J.B. et al., 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926), pp.64–71.

Fields, S. & Song, O., 1989. A novel genetic system to detect protein protein interactions. *Nature*, 340(6230), pp.245–246.

Fire, a et al., 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669), pp.806–11.

Fisher, K.H., 2009. *DPhil thesis*. University of Oxford.

Gache, V. et al., 2010. Xenopus meiotic microtubule-associated interactome. *PloS one*, 5(2), p.e9248.

Galjart, N., 2005. CLIPs and CLASPs and cellular dynamics. *Nature reviews. Molecular cell biology*, 6(6), pp.487–98.

Gandhi, T.K.B. et al., 2006. Analysis of the human protein interactome and comparison with yeast , worm and fly interaction datasets. , 38(3), pp.285–293.

Gavin, A.-C. et al., 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868), pp.141–7.

Gehlenborg, N. et al., 2010. Visualization of omics data for systems biology. *Nature methods*, 7(3 Suppl), pp.S56–68.

Gentleman, R. & Huber, W., 2007. Making the most of high-throughput protein-interaction data. *Genome biology*, 8(10), p.112.

Gerdes, H.H. & Kaether, C., 1996. Green fluorescent protein: applications in cell biology. *FEBS letters*, 389(1), pp.44–7.

Gerlich, D. et al., 2002. Nuclear Envelope Breakdown Proceeds by Microtubule-Induced Tearing of the Lamina. *Cell*, 108, pp.83–96.

Ghosh, S. & Basu, A., 2012. Network medicine in drug design: implications for neuroinflammation. *Drug discovery today*, 17(11-12), pp.600–7.

Gingras, A.-C. et al., 2007. Analysis of protein complexes using mass spectrometry. *Nature reviews. Molecular cell biology*, 8(8), pp.645–54.

Giot, L. et al., 2003. A protein interaction map of Drosophila melanogaster. *Science (New York, N.Y.)*, 302(5651), pp.1727–36.

Gloeckner, C.J. et al., 2007. A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics*, 7(23), pp.4228–34.

Goh, K. et al., 2007. The human disease network. *PNAS*, 104, pp.8685–8690.

Goh, W.W.B. et al., 2012. How advancement in biological network analysis methods empowers proteomics. *Proteomics*, 12(4-5), pp.550–63.

Goshima, G., 2010. Assessment of mitotic spindle phenotypes in Drosophila S2 cells. *Methods Cell Biol.*, 97, pp.259–75.

Goshima, G. et al., 2008. Augmin: a protein complex required for centrosome-independent microtubule generation within the spindle. *The Journal of cell biology*, 181(3), pp.421–9.

Goshima, G. et al., 2007a. Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science (New York, N.Y.)*, 316(5823), pp.417–21.

Goshima, G. et al., 2007b. Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science (New York, N.Y.)*, 316(5823), pp.417–21.

Goshima, G. et al., 2007c. Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science (New York, N.Y.)*, 316(5823), pp.417–21.

Graveley, B.R. et al., 2010. The developmental transcriptome of Drosophila melanogaster. *Nature*, pp.1–7.

Grimberg, K.B. et al., 2011. Basic leucine zipper protein Cnc-C is a substrate and transcriptional regulator of the Drosophila 26S proteasome. *Molecular and cellular biology*, 31(4), pp.897–909.

Grünenfelder, B. & Winzeler, E. a, 2002. Treasures and traps in genome-wide data sets: case examples from yeast. *Nature reviews. Genetics*, 3(9), pp.653–61.

Gstaiger, M. & Aebersold, R., 2009. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nature reviews. Genetics*, 10(9), pp.617–27.

Gu, Z. et al., 2012. Centrality-based pathway enrichment: a systematic approach for finding significant pathways dominated by key genes. *BMC systems biology*, 6, p.56.

Guerrero, C. et al., 2006. An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Molecular & cellular proteomics : MCP*, 5(2), pp.366–78.

Guest, S.T. et al., 2011a. A protein network-guided screen for cell cycle regulators in Drosophila. *BMC systems biology*, 5, p.65.

Guest, S.T. et al., 2011b. A protein network-guided screen for cell cycle regulators in Drosophila. *BMC systems biology*, 5(1), p.65.

Guo, S. & Kemphues, K.J., 1995. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*, 81(4), pp.611–20.

Guruharsha, K.G. et al., 2011. A protein complex network of Drosophila melanogaster. *Cell*, 147(3), pp.690–703.

Ha, G.-H. et al., 2009. Mitotic catastrophe is the predominant response to histone acetyltransferase depletion. *Cell death and differentiation*, 16(3), pp.483–97.

Halpain, S. & Dehmelt, L., 2006. The MAP1 family of microtubule-associated proteins. *Genome Biology*, 7(6), p.224.

Hammond, S.M. et al., 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404(6775), pp.293–6.

Han, J.J. et al., 2004. Evidence for dynamically organized modularity in the yeast protein – protein interaction network. *Nature*, 430(July).

Hart, G.T., Ramani, A.K. & Marcotte, E.M., 2006. How complete are current yeast and human protein-interaction networks? *Genome biology*, 7(11), p.120.

Hartwell, L.H. et al., 1999. From molecular to modular cell biology. *Nature*, 402(6761 Suppl), pp.C47–52.

He, M., Wang, Y. & Li, W., 2009. PPI finder: a mining tool for human protein-protein interactions. *PloS one*, 4(2), p.e4554.

Heald, R. et al., 1996. Self-organization of microtubules into bipolar spindles around artificial chromosomes in Xenopus egg extracts. *Nature*, 382, p.420.

Hernandez-Toro, J., Prieto, C. & De las Rivas, J., 2007. APID2NET: unified interactome graphic analyzer. *Bioinformatics (Oxford, England)*, 23(18), pp.2495–7.

Hillenkamp, F. et al., 1991. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical Chemistry*, 63(24), p.1193A–1203A.

Ho, Y. et al., 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, 415(6868), pp.180–3.

Houle, D., Govindaraju, D.R. & Omholt, S., 2010. Phenomics: the next challenge. *Nature Reviews Genetics*, 11(12), pp.855–866.

Hsu, S.I. et al., 2001. TRIP-Br : a novel family of PHD zinc-finger and bromodomain-interacting proteins that regulate the transcriptional activity of E2F-1/ DP-1. *The EMBO journal*, 20(9), pp.2273–85.

Huang, T., Lin, C. & Kao, C., 2007. Reconstruction of human protein interolog network using evolutionary conserved network. *BMC bioinformatics*, 14, pp.1–14.

Hughes, J.R. et al., 2008. A microtubule interactome: complexes with roles in cell cycle and mitosis. *PLoS biology*, 6(4), p.e98.

Ideker, T., Dutkowski, J. & Hood, L., 2011. Boosting Signal-to-Noise in Complex Biology: Prior Knowledge Is Power. *Cell*, 144(6), pp.860–3.

Inui, M. et al., 2008. Expression of Clostridium acetobutylicum butanol synthetic genes in Escherichia coli. *Applied microbiology and biotechnology*, 77(6), pp.1305–16.

Isalan, M. et al., 2008. Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, 452(7189), pp.840–5.

Ito, T. et al., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8), pp.4569–74.

Jackson, J.R. et al., 2007. Targeted anti-mitotic therapies: can we improve on tubulin agents? *Nature reviews. Cancer*, 7(2), pp.107–17.

Jeong, H. et al., 2001. Lethality and centrality in protein networks. *Nature*, 411(6833), pp.41–2.

Jeong, H. et al., 2000. The large-scale organization of metabolic networks. *Nature*, 407(6804), pp.651–4.

Johnston, H., 1999. Identification of a Novel SNF2/SWI2 Protein Family Member, SRCAP, Which Interacts with CREB-binding Protein. *Journal of Biological Chemistry*, 274(23), pp.16370–16376.

Jonsson, P.F. & Bates, P. a, 2006. Global topological features of cancer proteins in the human interactome. *Bioinformatics (Oxford, England)*, 22(18), pp.2291–7.

Joyce, A.R. & Palsson, B.Ø., 2006. The model organism as a system: integrating “omics” data sets. *Nature Reviews Molecular Cell Biology*, 7(3), pp.198–210.

Kanehisa, M. et al., 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(Database issue), pp.D109–14.

KAP, 2013. KAP Genetics and Development. , 2013. Available at: http://biology.kenyon.edu/courses/biol114/Chap12/Chapter_12b.html [Accessed August 6, 2013].

Kelso, R.J. et al., 2004. Flytrap, a database documenting a GFP protein-trap insertion screen in Drosophila melanogaster. *Nucleic acids research*, 32(Database issue), pp.D418–20.

Kennerdell, J.R. & Carthew, R.W., 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell*, 95(7), pp.1017–26.

Kim, H.U. et al., 2011. Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. *Molecular Systems Biology*, 7(460), pp.1–15.

Kitano, H., 2002. Computational systems biology. *Nature*, 420(November).

Kodumal, S.J. et al., 2004. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, 101(44), pp.15573–8.

Koegl, M. & Uetz, P., 2007. Improving yeast two-hybrid screening systems. *Briefings in functional genomics & proteomics*, 6(4), pp.302–12.

Kollman, J.M. et al., 2011. Microtubule nucleation by γ-tubulin complexes. *Nature reviews. Molecular cell biology*, 12(11), pp.709–21.

Kulkarni, M.M. et al., 2006. Evidence of off-target effects associated with long dsRNAs in Drosophila melanogaster cell-based assays. *Nature methods*, 3(10), pp.833–838.

Kung, A.L. et al., 2004. Small molecule blockade of transcriptional coactivation of the hypoxia-inducible factor pathway. *Cancer cell*, 6(1), pp.33–43.

Kwon, M. & Scholey, J.M., 2004. Spindle mechanics and dynamics during mitosis in Drosophila. *Trends in cell biology*, 14(4), pp.194–205.

LaCount, D.J. et al., 2005. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature*, 438(7064), pp.103–7.

Lage, K. et al., 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature biotechnology*, 25(3), pp.309–16.

Lanza, A.M., Crook, N.C. & Alper, H.S., 2012. Innovation at the intersection of synthetic and systems biology. *Current opinion in biotechnology*, 23(5), pp.712–7.

De Las Rivas, J. & Fontanillo, C., 2010. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks F. Lewitter, ed. *PLoS Computational Biology*, 6(6), p.e1000807.

Van Leene, J. et al., 2010. Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana. *Molecular systems biology*, 6(397), p.397.

Lewis, A.C.F. et al., 2012. What evidence is there for the homology of protein-protein interactions? *PLoS computational biology*, 8(9), p.e1002645.

Li, J. et al., 2004. The nuclear protein p34SEI-1 regulates the kinase activity of cyclin-dependent kinase 4 in a concentration-dependent manner. *Biochemistry*, 43(14), pp.4394–9.

Li, S. et al., 2004. A map of the interactome network of the metazoan C. elegans. *Science (New York, N.Y.)*, 303(5657), pp.540–3.

Liu, M. et al., 2007. Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS genetics*, 3(6), p.e96.

Liu, Z.-P. et al., 2012. Network-based analysis of complex diseases. *IET systems biology*, 6(1), pp.22–33.

Liu, Z.-P. & Chen, L., 2012. Proteome-wide prediction of protein-protein interactions from high-throughput data. *Protein & cell*, 3(7), pp.508–20.

Luo, F. et al., 2007. Modular organization of protein interaction networks. *Bioinformatics (Oxford, England)*, 23(2), pp.207–14.

Ma, Y. et al., 2006. Prevalence of off-target effects in Drosophila RNA interference screens. *Nature*, 443(7109), pp.359–63.

Ma'ayan, A., 2009. Network integration and graph analysis in mammalian molecular systems biology. *IET systems biology*, 2(5), pp.206–221.

Maere, S., Heymans, K. & Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)*, 21(16), pp.3448–9.

Magrane, M. & Consortium, U., 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database*, 2011, pp.bar009–bar009.

Mann, M., 2006. Functional and quantitative proteomics using SILAC. *Nature reviews. Molecular cell biology*, 7(12), pp.952–8.

Marchler-Bauer, A. et al., 2010. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research*, 39(November 2010), pp.225–229.

Mathey-Prevot, B. & Perrimon, N., 2006. Drosophila genome-wide RNAi screens: are they delivering the promise? *Cold Spring Harbor symposia on quantitative biology*, 71, pp.141–8.

Matthews, K. a, Kaufman, T.C. & Gelbart, W.M., 2005. Research resources for Drosophila: the expanding universe. *Nature reviews. Genetics*, 6(3), pp.179–93.

Matthews, L.R. et al., 2001. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome research*, 11(12), pp.2120–6.

Mesquita, D., Dekanty, A. & Milán, M., 2010. A dp53-Dependent Mechanism Involved in Coordinating Tissue Growth in Drosophila K. Basler, ed. *PLoS Biology*, 8(12), p.e1000566.

Mewes, H.W. et al., 2010. MIPS: curated databases and comprehensive secondary data resources in 2010. *Nucleic acids research*, 39(November 2010), pp.220–224.

Mika, S. & Rost, B., 2006. Protein-protein interactions more conserved within species than across species. *PLoS computational biology*, 2(7), p.e79.

Milo, R. et al., 2002. Network motifs: simple building blocks of complex networks. *Science (New York, N.Y.)*, 298(5594), pp.824–7.

Mohr, S., Bakal, C. & Perrimon, N., 2010a. Genomic screening with RNAi: results and challenges. *Annual review of biochemistry*, 79, pp.37–64.

Mohr, S., Bakal, C. & Perrimon, N., 2010b. Genomic Screening with RNAi: Results and Challenges. *Annual review of biochemistry*.

Mohr, S., Bakal, C. & Perrimon, N., 2010c. Genomic screening with RNAi: results and challenges. *Annual review of biochemistry*, 79, pp.37–64.

Montañez-Wiscovich, M.E. et al., 2010. Aberrant expression of LMO4 induces centrosome amplification and mitotic spindle abnormalities in breast cancer cells. *The Journal of pathology*, 222(3), pp.271–81.

Musacchio, A. & Salmon, E.D., 2007. The spindle-assembly checkpoint in space and time. *Nature reviews. Molecular cell biology*, 8(5), pp.379–93.

Napoli, C., Lemieux, C. & Jorgensen, R., 1990. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell*, 2(4), pp.279–289.

Nasmyth, K., 2005. How might cohesin hold sister chromatids together? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1455), pp.483–96.

Nath, S. et al., 2011. Spindle assembly checkpoint protein Cdc20 transcriptionally activates expression of ubiquitin carrier protein UbcH10. *The Journal of biological chemistry*, 286(18), pp.15666–77.

Navlakha, S. & Kingsford, C., 2010. The power of protein interaction networks for associating genes with diseases. *Bioinformatics (Oxford, England)*, 26(8), pp.1057–63.

Neumann, B. et al., 2010. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464(7289), pp.721–7.

Ohta, S. et al., 2010. The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. *Cell*, 142(5), pp.810–821.

Oliynyk, M. et al., 2007. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nature biotechnology*, 25(4), pp.447–53.

Oltersdorf, T. et al., 2005. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, 435(7042), pp.677–81.

Ong, S.-E. et al., 2002. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, 1(5), pp.376–386.

Orchard, S. et al., 2007. The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nature biotechnology*, 25(8), pp.894–8.

Oti, M. et al., 2006. Predicting disease genes using protein-protein interactions. *Journal of medical genetics*, 43(8), pp.691–8.

Pandey, U.B. & Nichols, C.D., 2011. Human Disease Models in Drosophila melanogaster and the Role of the Fly in Therapeutic Drug Discovery. , 63(2), pp.411–436.

Papageorgiou, A.C. & Wikman, L.E.K., 2004. Is JAK3 a new drug target for immunomodulation-based therapies? *Trends in pharmacological sciences*, 25(11), pp.558–62.

Patel, P.C. et al., 2009. Proteomic analysis of microtubule-associated proteins during macrophage activation. *Molecular & cellular proteomics : MCP*, 8(11), pp.2500–14.

Paulovich, A. et al., 2009. Interlaboratory Study Characterizing a Yeast Performance Standard for Benchmarking LC-MS Platform Performance. *Molecular & Cellular Proteomics*, 458(2), pp.242–254.

Peset, I. & Vernos, I., 2008. The TACC proteins: TACC-ling microtubule dynamics and centrosome function. *Trends in cell biology*, 18(8), pp.379–88.

Peters, J.-M., 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature reviews. Molecular cell biology*, 7(9), pp.644–56.

Prieto, C. & De Las Rivas, J., 2006. APID: Agile Protein Interaction DataAnalyzer. *Nucleic acids research*, 34(Web Server issue), pp.W298–302.

Ravasz, E. et al., 2002. Hierarchical organization of modularity in metabolic networks. *Science (New York, N.Y.)*, 297(5586), pp.1551–5.

Ravasz, E. & Barabási, A.-L., 2003. Hierarchical organization in complex networks. *Physical Review E*, 67(2), pp.1–7.

Reiter, L.T. et al., 2001. A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome research*, 11(6), pp.1114–25.

Rieder, C.L., 2008. Kinetochore fiber formation in animal somatic cells: dueling mechanisms come to a draw. *Chromosma*, 114(5), pp.310–318.

Ro, D.-K. et al., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086), pp.940–3.

Rogers, S.L. & Rogers, G.C., 2008. Culture of Drosophila S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nature protocols*, 3(4), pp.606–11.

Rojas, A.M. et al., 2012. Uncovering the molecular machinery of the human spindle--an integration of wet and dry systems biology. *PloS one*, 7(3), p.e31813.

Romano, N. & Macino, G., 1992. Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular microbiology*, 6(22), pp.3343–53.

Roukos, D.H., 2012. Disrupting cancer cells' biocircuits with interactome-based drugs: is "clinical" innovation realistic? *Expert review of proteomics*, 9(4), pp.349–53.

Roy, S. et al., 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science (New York, N.Y.)*, 330(6012), pp.1787–97.

Rual, J.-F. et al., 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437(7062), pp.1173–8.

Rüegg, C. et al., 2008. Omics meets hypothesis-driven research. Partnership for innovative discoveries in vascular biology and angiogenesis. *Thrombosis and Haemostasis*, pp.738–746.

Rychlik, J.L. et al., 2005. Phox2 and dHAND Transcription Factors Select Shared and Unique Target Genes in the Noradrenergic Cell Type. *Journal of molecular neuroscience*, 27, pp.281–292.

Sakamoto, T. et al., 2008. Mass spectrometric analysis of microtubule co-sedimented proteins from rat brain. *Genes to cells : devoted to molecular & cellular mechanisms*, 13(4), pp.295–312.

Salwinski, L. et al., 2004. The Database of Interacting Proteins: 2004 update. *Nucleic acids research*, 32(Database issue), pp.D449–51.

Samuel, T., Weber, H.O. & Funk, J.O., 2002. Linking DNA Damage to Cell Cycle Checkpoints. , (June), pp.162–168.

Sanz-Pamplona, R. et al., 2012. Tools for protein-protein interaction network analysis in cancer research. *Clinical and Translational Oncology*, 14(1), pp.3–14.

Sauer, G. et al., 2005. Proteome analysis of the human mitotic spindle. *Molecular & cellular proteomics : MCP*, 4(1), pp.35–43.

Scardoni, G., Petterlini, M. & Laudanna, C., 2009. Analyzing biological network parameters with CentiScaPe. *Bioinformatics (Oxford, England)*, 25(21), pp.2857–9.

Schadt, E.E. & Björkegren, J.L.M., 2012. NEW: network-enabled wisdom in biology, medicine, and health care. *Science translational medicine*, 4(115), p.115rv1.

Schimanski, B., Nguyen, T.N. & Gu, A., 2005. Highly Efficient Tandem Affinity Purification of Trypanosome Protein Complexes Based on a Novel Epitope Combination. , 4(11), pp.1942–1950.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth*, 9(7), pp.671–675.

Schneider, I., 1972. Cell lines derived from late embryonic stages of Drosophila melanogaster. *Journal of embryology and experimental morphology*, 27(2), pp.353–65.

Shamanski, F.L. & Orr-Weaver, T.L., 1991. The Drosophila plutonium and pan gu genes regulate entry into S phase at fertilization. *Cell*, 66(6), pp.1289–300.

Shapira, S.D. et al., 2009. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell*, 139(7), pp.1255–67.

Shelanski, M.L., Gaskin, F. & Cantor, C.R., 1973. Microtubule assembly in the absence of added nucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 70(3), pp.765–8.

Shoval, O. & Alon, U., 2010. SnapShot: Network Motifs. *Cell*, 143(2), pp.326.e1–326.e2.

Smolik, S. & Jones, K., 2007. Drosophila dCBP is involved in establishing the DNA replication checkpoint. *Molecular and cellular biology*, 27(1), pp.135–46.

Smoot, M.E. et al., 2010. Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics (Oxford, England)*, 27(3), pp.431–432.

Soundararajan, M. et al., 2007. The centaurin gamma-1 GTPase-like domain functions as an NTPase. *The Biochemical journal*, 401(3), pp.679–88.

De Souza, C.P.C. & Osmani, S. a, 2007. Mitosis, not just open or closed. *Eukaryotic cell*, 6(9), pp.1521–7.

Spradling, a C. et al., 1999. The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics*, 153(1), pp.135–77.

Stark, C. et al., 2010. The BioGRID Interaction Database: 2011 update. *Nucleic acids research*, 39(November 2010), pp.698–704.

Steen, E.J. et al., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, 463(7280), pp.559–62.

Strogatz, S.H., 2001. Exploring complex networks. *Nature*, 410(March).

Stuart, L.M. et al., 2007. A systems biology analysis of the Drosophila phagosome. *Nature*, 445(7123), pp.95–101.

Suozzi, K.C., Wu, X. & Fuchs, E., 2012. Spectraplakins: master orchestrators of cytoskeletal dynamics. *The Journal of cell biology*, 197(4), pp.465–75.

Swallow, C.J. et al., 2005. Sak/Plk4 and mitotic fidelity. *Oncogene*, 24(2), pp.306–12.

Tabb, D.L. et al., 2010. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography - Tandem Mass Spectrometry research articles. , pp.761–776.

Taylor, I.W. et al., 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature biotechnology*, 27(2), pp.199–204.

Theocharidis, A. et al., 2009. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nature protocols*, 4(10), pp.1535–50.

Tijsterman, M. & Plasterk, R.H.A., 2004. Dicers at RISC : The Mechanism of RNAi The pathway of RNA interference starts when Dicer. *Cell*, 117(1-4).

Trautmann, L. & Sekaly, R.-P., 2011. Solving vaccine mysteries: a systems biology perspective. *Nature immunology*, 12(1529-2916 (Electronic)), pp.729–731.

Tsuchiya, M. et al., 2011. Critical role of the nucleolus in activation of the p53-dependent postmitotic checkpoint. *Biochemical and biophysical research communications*, 407(2), pp.378–82.

Turnell, A.S. et al., 2005. The APC/C and CBP/p300 cooperate to regulate transcription and cell-cycle progression. *Nature*, 438(7068), pp.690–5.

Tuschl, T. et al., 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & development*, 13, pp.3191–3197.

Tyson, J.J. & Novak, B., 2011. Cell cycle: who turns the crank? *Current biology : CB*, 21(5), pp.R185–7.

Uehara, R. et al., 2009. The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17), pp.6998–7003.

Uetz, P. et al., 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403(6770), pp.623–7.

Vanunu, O. et al., 2010. Associating genes and protein complexes with disease via network propagation. *PLoS computational biology*, 6(1), p.e1000641.

Vardy, L. & Orr-Weaver, T.L., 2007. The Drosophila PNG kinase complex regulates the translation of cyclin B. *Developmental cell*, 12(1), pp.157–66.

Vardy, L., Pesin, J. a & Orr-Weaver, T.L., 2009. Regulation of Cyclin A protein in meiosis and early embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp.1838–43.

Vassilev, L.T. et al., 2004. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science (New York, N.Y.)*, 303(5659), pp.844–8.

VGEC, 2013. University of Licester - Virtual Genetics Educaiton Centre. Available at: http://www2.le.ac.uk/departments/genetics/vgec/schoolscolleges/topics/cellcycle-mitosis-meiosis [Accessed August 6, 2013].

Vidal, M., Cusick, M.E. & Barabási, A.-L., 2011. Interactome networks and human disease. *Cell*, 144(6), pp.986–98.

Villavicencio, E.H., Walterhouse, D.O. & Iannaccone, P.M., 2000. The sonic hedgehog-patched-gli pathway in human development and disease. *American journal of human genetics*, 67(5), pp.1047–54.

Voevodski, K., Teng, S.-H. & Xia, Y., 2009. Finding local communities in protein networks. *BMC bioinformatics*, 10, p.297.

Wachi, S., Yoneda, K. & Wu, R., 2005. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics (Oxford, England)*, 21(23), pp.4205–8.

Wade, R.H., 2009. On and around microtubules: an overview. *Molecular biotechnology*, 43(2), pp.177–91.

Walczak, C.E. & Shaw, S.L., 2010. A MAP for bundling microtubules. *Cell*, 142(3), pp.364–7.

Walhout, a. J., 2000. Protein Interaction Mapping in C. elegans Using Proteins Involved in Vulval Development. *Science*, 287(5450), pp.116–122.

Wang, L., Tu, Z. & Sun, F., 2009. A network-based integrative approach to prioritize reliable hits from multiple genome-wide RNAi screens in Drosophila. *BMC genomics*, 10, p.220.

Wang, Y. et al., 2010. Global Protein-Protein Interaction Network in the Human Pathogen Mycobacterium tuberculosis H37Rv research articles. *Journal of Proteome Research*, pp.6665–6677.

Wasi, S., 2003. RNA interference: the next genetics revolution. *Horizon Symposium: RNAissance*, (May), pp.1–4.

Wasner, M. et al., 2003. Cyclin B1 transcription is enhanced by the p300 coactivator and regulated during the cell cycle by a CHR-dependent repression mechanism. *FEBS Letters*, 536(1-3), pp.66–70.

Weiss, M. et al., 2010. Shotgun proteomics data from multiple organisms reveals remarkable quantitative conservation of the eukaryotic core proteome. *Proteomics*, 10(6), pp.1297–306.

Wepf, A. et al., 2009. Quantitative interaction proteomics using mass spectrometry. *Nature methods*, 6(3), pp.203–205.

White, E. a & Glotzer, M., 2012. Centralspindlin: at the heart of cytokinesis. *Cytoskeleton (Hoboken, N.J.)*, 69(11), pp.882–92.

Wiese, C. & Zheng, Y., 2006. Microtubule nucleation: gamma-tubulin and beyond. *Journal of cell science*, 119(Pt 20), pp.4143–53.

Wiles, A.M. et al., 2010. Building and analyzing protein interactome networks by cross-species comparisons. *BMC systems biology*, 4, p.36.

Wilson, L. & Meza, I., 1973. The mechanism of action of colchicine. Colchicine binding properties of sea urchin sperm tail outer doublet tubulin. *The Journal of cell biology*, 58(3), pp.709–19.

Winzeler, E. a, 2006. Applied systems biology and malaria. *Nature reviews. Microbiology*, 4(2), pp.145–51.

Wu, J., Vallenius, T. & Ovaska, K., 2009. Integrated network analysis platform for interactions. *Nature Methods*, 6(1), pp.2008–2010.

Wu, P.-S., Egger, B. & Brand, A.H., 2008. Asymmetric stem cell division: lessons from Drosophila. *Seminars in cell & developmental biology*, 19(3), pp.283–93.

Xiong, B. & Gerton, J.L., 2010. Regulators of the cohesin network. *Annual review of biochemistry*, 79, pp.131–53.

Xu, J. & Li, Y., 2006. Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics (Oxford, England)*, 22(22), pp.2800–5.

Yan, H. et al., 2010. A genome-wide gene function prediction resource for Drosophila melanogaster. *PloS one*, 5(8), p.e12139.

Yang, X., Ogryzko, V. & Nishikawa, J., 1996. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A.

Yeong, F.M., Lim, H.H. & Surana, U., 2002. MEN, destruction and separation: mechanistic links between mitotic exit and cytokinesis in budding yeast. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 24(7), pp.659–66.

Yu, H. et al., 2004. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome research*, 14(6), pp.1107–18.

Yu, H. et al., 2008. High-quality binary protein interaction map of the yeast interactome network. *Science (New York, N.Y.)*, 322(5898), pp.104–10.

Yu, W. et al., 2008. The Microtubule-severing Proteins Spastin and Katanin Participate Differently in the Formation of Axonal Branches. , 19(April), pp.1485–1498.

Zeghouf, M. et al., 2004. Sequential Peptide Affinity (SPA) System for the Identification of Mammalian and Bacterial Protein Complexes research articles. , pp.463–468.

Zhai, B. et al., 2008. Phosphoproteome Analysis of Drosophila melanogaster Embryos. *Journal of Proteome Research*, 7, pp.1675–1682.

Zhang, W., Li, F. & Nie, L., 2010. Integrating multiple " omics " analysis for microbial biology : application and methodologies. *Microbiology*, pp.287–301.

Zhou, D. & He, Y., 2008. Extracting interactions between proteins from the literature. *Journal of biomedical informatics*, 41(2), pp.393–407.

Appendix I

Custom Perl scripts used to gather, map and count different features.

```
use strict;

#Program to take out PPIs for a given list of proteins from a BioGrid file.

print "enter input file name containing list of protein accessions:\n";
my $file1 = <STDIN>;
chomp $file1;

print "enter file name containing list of protein interactions:\n";
my $file2 = <STDIN>;
chomp $file2;

my $outfile = "ppi_out.txt";
open OUT, ">$outfile" or die "Cannot open $outfile\n";

open LIST, $file1;
my @list = <LIST>;
close LIST;

open INT, $file2;
my @int = <INT>;
close INT;

foreach my $list (@list) {
     chomp $list;
     $list =~ s/^\s+//;
     $list =~ s/\s+$//;       print "protein name $list\n";
                  foreach my $int (@int) {       chomp $int;       my ($a, $b);
           ($a, $b) = split (/\t/, $int);
           $a =~ s/^\s+//;
           $a =~ s/\s+$//;
           $b =~ s/^\s+//;
           $b =~ s/\s+$//;
           #@int1 = split /\t/, "$a, $b, $c"

 if ($a eq $list) {
           print "$int\n";
           print OUT "$int\n";
              }
 if ($b eq $list) {
           print "$int\n";
           print OUT "$int\n";
              }
    } }

close OUT;
```

```
use strict;

#Program to map Uniprot AC/ID to Flybase symbols (and CG and FBgn) for MAP proteins.

print "enter input file name containing flybase query results:\n";
my $file1 = <STDIN>;
chomp $file1;  print "enter file name containing uniprot mapping:\n";
my $file2 = <STDIN>;
chomp $file2;

my $outfile = "unimapped1324.txt";
open OUT, ">$outfile" or die "Cannot open $outfile\n";

open LIST, $file1;
my @list = <LIST>;
close LIST;
 open UNI, $file2;
my @uni = <UNI>;
close UNI;
my $counttot=0;
foreach my $list (@list) {
chomp $list;
my ($a, $b, $c, $d);
($a, $b, $c, $d) = split (/\t/, $list);
$a =~ s/^\s+//;
$a =~ s/\s+$//;
$b =~ s/^\s+//;
$b =~ s/\s+$//;
$c =~ s/^\s+//;
$c =~ s/\s+$//;
$d =~ s/^\s+//;
$d =~ s/\s+$//;
#           print "$a:$c\n";
foreach my $uni (@uni) {
chomp $uni;
my ($x, $y);
($x, $y) = split (/\t/, $uni);
$x =~ s/^\s+//;
$x =~ s/\s+$//;
$y =~ s/^\s+//;
$y =~ s/\s+$//;
#           print "$x:$y\n";
if ($c eq $y) {
print "yesssssss\n";
print OUT "$x\t$d\n";
my $counttot++;
print "$counttot\n"; }  }  }

close OUT;
```

```
use strict;

#Program to count PPIs for a given list of proteins reading a BioGrid file and print an array of the interactors as well.

print "enter input file name containing list of protein accessions:\n";
my $file1 = <STDIN>;
chomp $file1;

print "enter file name containing list of protein interactions:\n";
my $file2 = <STDIN>;
chomp $file2;

my $outfile = "ppinos_out.txt";
open OUT, ">$outfile" or die "Cannot open $outfile\n";

open LIST, $file1;
my @list = <LIST>;
close LIST;

open INT, $file2;
my @int = <INT>;
close INT;

foreach my $list (@list) {
chomp $list;
print "protein name :$list:\n";
print OUT "$list\t";
my $ppicount=0;
     $list =~ s/^\s+//;
     $list =~ s/\s+$//;
 my @ppilist=();
                  foreach my $int (@int) {
                                    chomp $int;
                                    my ($a, $b);
                                    ($a, $b) = split (/\t/, $int);
                                    print "$a:$b\n";
 if ($a eq $list) {
            print "$int\n";
            push (@ppilist, $b);          $ppicount++;
               }
 if ($b eq $list) {
            print "$a\t";
            push (@ppilist, $a);          $ppicount++;
               }
    } print OUT "\t$ppicount\t@ppilist\n";  }

close OUT;
```

Appendix II

List of the top 100 scoring proteins from the prediction (test) set of the MAP interactome. The Uniprot accession numbers, gene symbol, the source of each protein in the MAP interactome dataset (D = Drosophila MAP dataset, I = Indirect interactors) and the predicted MAP score are shown in the table.

UniProt AC	Source	Gene Symbol	MAP Score
Q9W321	I	Nej	1.000
P21187	D	PABP	0.994
Q9VL21	I	CG5708	0.818
Q9VAU6	I	CG9986	0.808
O46070	I	CG2865	0.763
P41374	D	eIF2A	0.751
Q961D1	I	CycK	0.728
Q9VYG2	I	Bap60	0.722
Q8T965	I	CG15676	0.652
Q960S0	I	BtbVII	0.644
P08570	I	RpLP1	0.641
Q7JVK6	I	trsn	0.639
Q02308	I	H	0.639
Q95TJ9	I	CycG	0.637
Q9V3U6	D	26-29-p	0.627
Q9VE69	I	CG31122	0.623
Q9VRQ4	I	CG5568	0.613
Q9XYW6	I	CHIP	0.606
A1ZAU6	I	CG4853	0.586
Q8MZ38	I	CG15109	0.585
Q9VEA2	I	CG7146	0.578
Q9W5Y4	I	CG17018	0.570
Q9VBU7	D	Nup358	0.569
Q9VZV3	I	CG11486	0.569
O76513	I	CycH	0.564
Q9VYA4	I	CG11164	0.564
Q9W4F9	I	CG32772	0.559
Q9W362	I	CG42569	0.553
Q9WVF5	I	EGFR	0.553
Q9Y149	I	MED15	0.552
Q7KSE8	I	OSA	0.544
Q9VK58	I	CG5792	0.540
Q9VW51	I	Su(Tpl)	0.538
O76927	I	oho23B	0.536
Q9VN49	I	CG31534	0.535
Q9VME5	I	slam	0.520
Q9VBN9	I	CG4730	0.504
A8Y4W5	I	Erk1	0.495
C7LAE7	I	Mod(Mdg4)	0.495
Q0E7J8	I	rev7	0.495
Q9W227	D	CG2852	0.495
Q9N6D8	I	p53	0.491
O18373	I	SelD	0.477
Q7JUY1	I	gus	0.468
Q9VC60	I	CG13625	0.462
Q9W1H7	I	PHDP	0.452
P24156	D	l(2)37Cc	0.451

Q9VTU0	I	CG5645	0.445
Q95RA9	I	CG9796	0.444
Q9VNV2	I	CG11523	0.438
Q545C3	I	CDK4	0.437
Q9W5A9	I	CG11380	0.437
Q7Z2C5	I	CG12723	0.437
C7LA75	I	HSC70-4	0.437
Q9XTM1	I	sec10	0.436
Q9VM49	I	CG11266	0.434
Q8T3W6	I	hale	0.427
Q9VFE4	D	RpS5b	0.425
A1Z9R6	I	Asx	0.424
Q9VRV8	I	Trn	0.424
Q9VL27	I	CG5731	0.423
Q9VAC4	I	CG7911	0.422
P13098	I	E(spl)	0.421
Q9VJD4	I	Sgt	0.415
Q9V564	I	CG1968	0.409
Q9VRV6	I	CG7386	0.409
Q9VK59	D	CG5787	0.408
Q9V345	I	CSN4	0.406
Q9VH64	D	CG8507	0.405
Q8SYG2	I	CSN3	0.405
Q7K0L7	I	CG6608	0.402
Q9VW47	I	kto	0.399
Q9VED4	I	CG14317	0.398
Q9VCK2	I	CG13822	0.398
P35122	D	Uch	0.394
Q94524	I	Dlc90F	0.393
Q9VUQ1	I	CG7945	0.393
Q8MRW8	I	CG30122	0.387
Q9VIJ0	I	CG9331	0.385
Q0E8J0	I	MEP-1	0.385
Q9VHA8	I	P58IPK	0.384
Q9W4I4	I	HLH4C	0.381
Q9VM97	I	Tsp	0.381
Q7K180	I	Map60	0.380
Q9VIN1	I	rtGEF	0.380
Q9VGV8	I	CG4511	0.377
Q9VVG2	I	CG13731	0.377
Q9VPH6	I	CG5618	0.370
P54611	I	Vha26	0.369
Q9VSV2	I	CG4476	0.362
Q9VFQ3	I	E5	0.362
Q8IRH5	I	CG2199	0.357
Q9I7T7	I	CG11505	0.357
Q9VIY9	I	CG17568	0.351
P14199	I	ref(2)P	0.351
Q7K126	I	l(2)NC136	0.350
Q9VEQ1	D	CG17556	0.350
Q9VT19	I	CG3335	0.350
P09957	I	y	0.350
Q9VMI5	D	CG9135	0.349

Appendix III

List of primers used to amplify cDNA for dsRNA production using the in vitro transcription. Each primer is added with the T7 promoter sequence (TAATACGACTCACTATAGGG) at the 5' end for the in vitro transcription reaction using the Ambion® MEGAscript® T7 kit (Life Technologies). Primers were designed using SnapDragon (http://www.flyrnai.org/cgi-bin/RNAi_find_primers.pl). The Gene symbol, FlyBase accession number, amplicon size and number of potential off-targets for each cDNA fragment are given. The threshold of predicted off-target size was set to the recommended 19bp, and the product size limited between 250-500bp. The number of primer pairs to call was set to 1. For each pair, the T7 promoter sequence. Expression levels in S2 cell lines were gathered from the DRSC resource website (http://www.flyrnai.org/).

Gene Symbol	Potential Off-targets	Exon or Exon/Intron Span	Expression in S2 cells	Product Size	Forward Primer (with T7 sequence)	Reverse Primer (with T7 sequence)
Sak	3	FBtr0078387-7	6.47588 - Expressed	360	TAATACGACTCACTATAGGGAGCGGCAGGAAAACTGACTA	TAATACGACTCACTATAGGGTGAAATTTGAACGTAGGCCC
mad2	1	FBtr0077113-3	35.1309 - Expressed	311	TAATACGACTCACTATAGGGGGACTTTAATATGCAGGCCG	TAATACGACTCACTATAGGGTGTAGTTGACCACGGTGTCC
E2f2	0	FBtr0081501-1	36.3147 - Expressed	339	TAATACGACTCACTATAGGGCGCACGAAGAATAGAGGGAG	TAATACGACTCACTATAGGGTGTCCATGCTAACGGGTGTA
CG13900	2	FBtr0072598-1	296.068 - Expressed	348	TAATACGACTCACTATAGGGAGATATTTGGGTTCCCGACC	TAATACGACTCACTATAGGGGTAGGCGTTGTGATCGGTTT
ptc	2	FBtr0089427-3	11.2339 - Expressed	453	TAATACGACTCACTATAGGGGGAGCAGACCAAGCTGATTC	TAATACGACTCACTATAGGGTAGTGCTGAAAGGCGAAGGT
akt1	332	FBtr0083226-7	27.9611 - Expressed	371	TAATACGACTCACTATAGGGATAACACTTGGGGCATAGCG	TAATACGACTCACTATAGGGGCACAGCACATTAGAGGCAA

nej	11	DRSC18006 (FBtr0302722 - Exon 14)	13.9738 - Expressed	478	TAATACGACTCACTATAGGGAAAACAAATTGCAGCTAAAAAAG	TAATACGACTCACTATAGGGGCTGCTCCGCTTTTTATTG
CG9986	1	FBtr0085304-2	7.51489 - Expressed	326	TAATACGACTCACTATAGGGCGCCGAGGAGAGTCTTAATG	TAATACGACTCACTATAGGGTCGAGTGCTTGGTGATGAAG
CG2865	3	FBtr0070408-2	8.15042 - Expressed	372	TAATACGACTCACTATAGGGGTGAGAATCTCTTCCAGCCG	TAATACGACTCACTATAGGGGATCGATTAGCAGCATGGGT
Bap60	0	FBtr0073741-2	109.57 - Expressed	355	TAATACGACTCACTATAGGGTGGATCTGCTGACGTTTGAG	TAATACGACTCACTATAGGGGTCTCCTGGGTGGTGTGAGT
CG15676	1	FBtr0071668-1	0 Not expressed	338	TAATACGACTCACTATAGGGTCGCCAAGCCAGAGTTCTAT	TAATACGACTCACTATAGGGCAGGCTGGTGATGTCCTTTT
CG5568	12	FBtr0077025-3	11.7198 - Expressed	424	TAATACGACTCACTATAGGGAAGCACAAGGACCCAATGTC	TAATACGACTCACTATAGGGTCCACAATCAAAGCTCCTCC
rev7	1	FBtr0078699-1	14.7952 - Expressed	435	TAATACGACTCACTATAGGGGGAGATTAAGGCCGACATCA	TAATACGACTCACTATAGGGAATGCCTCCTGGGTAGTGTG
med15	456	FBtr0078062-2	34.1771 - Expressed	399	TAATACGACTCACTATAGGGCGGAACTGCCTCTAACTTGC	TAATACGACTCACTATAGGGTGTTGCATGGCATTTACGTT
p53	1	FBtr0301765-2	13.3466 - Expressed	235	TAATACGACTCACTATAGGGCCAACAATAATTCCGATGGC	TAATACGACTCACTATAGGGTGATCAAGCGATCTTGTTGG
CG11486	1	FBtr0072991-7	41.6429 - Expressed	203	TAATACGACTCACTATAGGGCCGCAACGAGATCTCTAACC	TAATACGACTCACTATAGGGCATGTATTCTTCGCAGGCAG
CG13625	7	FBtr0084671-1	9.8944 - Expressed	493	TAATACGACTCACTATAGGGCCGAAATCCACCAAGACACT	TAATACGACTCACTATAGGGGAAAGCTGCCCTCGTACTTG
gus	13	FBtr0309851-2	27.153 - Expressed	358	TAATACGACTCACTATAGGGTCCCCGAGAGTTGTTTTGTC	TAATACGACTCACTATAGGGCCACTGGGTGCCTATGAAAT
PHDP	0	FBtr0072224-3	0.083365 Not expressed	268	TAATACGACTCACTATAGGGGAAGAGCCAAGTTTCGCAAG	TAATACGACTCACTATAGGGCGCAATTCGCTCCAAATAGT
CG5731	1	FBtr0079959-4	1.97806 - Weakly expressed	209	TAATACGACTCACTATAGGGTCCACTTTCGGTCGACTTCT	TAATACGACTCACTATAGGGATTACGTCCTGGTTGTTGCC
CG12723	1	FBtr0073738-2	0.312598 Not expressed	431	TAATACGACTCACTATAGGGGACCTATGCTCCGACTGCTC	TAATACGACTCACTATAGGGCGGGAAATTGTGGTATTTGG
CG8128	2	FBtr0074053-4	13.1939 - Expressed	482	TAATACGACTCACTATAGGGGGAGCACGAGTCTAGCAACC	TAATACGACTCACTATAGGGTTAGCACCTGATGCACTTCG
CG13510	1	FBtr0309799-2	2.79373 - Weakly expressed	208	TAATACGACTCACTATAGGGACATCGGCACCTACGAGAAC	TAATACGACTCACTATAGGGTAGCGGAAAGAGTGCCTTGT

CG9331	0	FBtr0081417-2	185.624 - Expressed	301	TAATACGACTCACTATAGGGAATCAAAGCCAGAATCGCAC	TAATACGACTCACTATAGGGGTCTCTTTACCTCCGGCACA
cenG1a	18	FBtr0080546-8	13.0219 - Expressed	498	TAATACGACTCACTATAGGGTAAGCACACGGATCATGAGG	TAATACGACTCACTATAGGGTTTCCATTCAACACTTGCCA
png	3	FBtr0070234-1	12.9134 - Expressed	319	TAATACGACTCACTATAGGGGCCCTACTGCACACCGTAGT	TAATACGACTCACTATAGGGCAAAGCACAGCACACACCTT
plu	0	FBtr0086306-2	0.0489817 Not expressed	306	TAATACGACTCACTATAGGGGACATGTACGGGAATACCGC	TAATACGACTCACTATAGGGTGTTTGAGTTTCTTGCACGC
gnu	1	FBtr0075697-2	0.207959 Not expressed	382	TAATACGACTCACTATAGGGGCGAGATACGGACTCTTTGG	TAATACGACTCACTATAGGGAGCAGGGCTTATAGCGTGAA
cdc14	436	FBtr0301331-7	7.84252 - Expressed	314	TAATACGACTCACTATAGGGCCATCACCACATGACGCTAC	TAATACGACTCACTATAGGGAGGAGGGAGGAGGAGTGGTA
CG8455	0	FBtr0079587-2	18.1935 - Expressed	370	TAATACGACTCACTATAGGGGATGGTTGCATGTTTTGCAC	TAATACGACTCACTATAGGGCTTTGTTTCGCCAGCTGAAT
CG6800	0	FBtr0084132-2	0.0720497 - Not expressed	301	TAATACGACTCACTATAGGGTCTCTCCTTGGTGCTGGAAT	TAATACGACTCACTATAGGGTAGAATCTCTGGCGCTCGAT
pie	0	FBtr0080046-2	11.2813 - Expressed	468	TAATACGACTCACTATAGGGTATGCTGCAGTTCTTGTGCC	TAATACGACTCACTATAGGGGTGAGTCGGATGTGTCCCTT
orc4	5	FBtr0306317-2	23.96 - Expressed	392	TAATACGACTCACTATAGGGAACTTCCTATTCCGCCTGGT	TAATACGACTCACTATAGGGTATGTCAGCGCCAGCTTATG

Appendix IV

The table shows the average percentage of mitotic cells with abnormal spindle morphology (Ab S) and metaphase chromosome alignment (Ab C) and the centrosome (0-4+) numbers for the two independent RNAi experiments against each gene. The standard error and corresponding p-values are also given. P-values were calculated using a two-tailed unpaired t-test. P-values<0.05 are highlighted red († indicates significant (p>0.05) results for a one-way ANOVA test).

Gene	Averages							Std. Errors							p-values (p<0.5)						
	Ab S	Ab C	0	1	2	3	4+	Ab S	Ab C	0	1	2	3	4+	Ab S	Ab C	0	1	2	3	4+
Blac - 12	36.474	36.904	0.000	5.127	64.258	10.886	19.260	4.644	5.455	0.000	4.402	1.030	1.915	5.001							
Sak	80.610	51.564	24.732	45.579	23.316	3.731	1.962	4.820	5.467	6.136	2.940	5.660	2.918	1.361	0.002	0.109	0.039†	0.004†	0.010†	0.082	0.068
mad2	79.429	67.278	1.833	8.420	64.163	8.550	17.034	1.168	0.114	0.341	2.450	11.924	6.376	8.339	0.041	0.175	0.117	0.173	0.985	0.794	0.596
E2f2	74.166	59.531	6.194	11.826	49.487	14.746	17.747	13.639	0.995	3.562	5.247	3.145	4.990	0.674	0.178	0.235	0.332	0.276	0.080	0.584	0.464
CG13900	79.649	68.304	1.754	2.544	37.924	26.813	30.965	6.316	19.415	1.754	0.789	9.854	1.257	7.632	0.035	0.364	0.500	0.240	0.218†	0.097	0.503
ptc	80.242	67.424	1.333	17.485	47.394	11.212	22.576	1.576	0.758	1.333	3.848	1.939	2.121	9.242	0.032	0.171	0.500	0.133	0.023	0.923	0.951
akt1	50.884	48.753	0.000	2.599	63.739	7.831	25.832	8.576	2.599	0.000	1.247	2.928	2.426	6.601	0.240	0.451	#DIV/0!	0.354	0.860	0.554	0.794
ald	73.447	82.450	2.389	7.719	49.185	15.866	24.841	3.024	2.057	0.428	2.085	9.970	1.781	6.531	0.020	0.096	0.113	0.154	0.363	0.341	0.873
nej	85.000	82.500	0.000	0.000	17.500	10.000	72.500	5.000	2.500	0.000	0.000	2.500	5.000	7.500	0.017	0.089	#DIV/0!	0.500	0.007†	0.911	0.043†
CG9986	82.601	69.744	1.493	10.526	37.356	13.859	36.766	5.458	18.316	1.493	2.907	5.928	0.426	9.900	0.022	0.328	0.500	0.166	0.115†	0.521	0.385
CG2865	82.104	58.857	3.846	15.097	43.982	13.291	23.783	2.512	12.938	3.846	2.852	2.957	3.035	0.706	0.017	0.362	0.500	0.107	0.043	0.632	0.943
Bap60	92.121	49.899	0.000	9.899	59.293	8.586	22.222	1.212	21.010	0.000	1.010	2.929	4.141	2.222	0.031†	0.724	#DIV/0!	0.023	0.296	0.722	0.865
CG15676	91.396	81.656	5.195	21.185	42.857	12.744	18.019	5.032	11.201	1.623	1.542	7.143	10.471	0.162	0.013†	0.107	0.193	0.020	0.187	0.883	0.482
CG5568	46.378	46.904	4.762	12.866	62.285	10.217	9.870	39.336	35.636	4.762	3.007	6.729	3.868	2.828	0.784	0.880	0.500	0.138	0.804	0.923	0.178

rev7	53.540	32.524	1.794	10.794	45.635	16.762	26.016	11.540	13.476	0.206	9.206	4.365	7.238	20.016	0.296	0.685	0.073	0.471	0.109	0.558	0.916
med15	74.578	72.574	1.559	9.693	43.155	19.140	26.453	8.755	5.907	0.293	0.434	2.414	6.177	8.732	0.087	0.086	0.118	0.018	0.024	0.385	0.792
p53	62.637	57.967	4.808	7.555	43.132	16.484	28.022	3.709	2.610	4.808	3.984	2.060	4.945	5.907	0.033	0.233	0.500	0.330	0.017	0.451	0.608
CG11486	79.887	51.159	0.000	2.569	52.914	10.683	33.835	2.256	10.683	0.000	1.378	4.229	1.159	1.692	0.021	0.496	#DIV/0!	0.394	0.187	0.983	0.260
CG13625	59.781	54.032	0.000	12.407	43.238	12.386	31.969	8.168	4.032	0.000	8.561	2.916	2.130	7.775	0.137	0.287	#DIV/0!	0.402	0.039	0.728	0.463
gus	48.237	51.976	4.060	12.821	47.863	17.147	18.109	22.596	16.079	1.496	3.846	3.419	4.647	2.724	0.611	0.595	0.225	0.186	0.081	0.391	0.481
PHDP	71.407	58.733	2.510	13.111	51.853	14.008	18.517	3.874	4.188	1.386	6.369	7.697	6.216	6.158	0.021	0.203	0.321	0.299	0.339	0.703	0.608
CG5731	59.241	57.035	0.746	3.709	42.241	19.996	31.837	2.524	0.318	0.746	0.768	2.535	0.593	1.986	0.047	0.270	0.500	0.110	0.025	0.210	0.316
CG8128	38.126	44.225	4.703	8.072	46.928	20.171	20.126	19.207	10.441	0.703	2.739	7.739	5.505	1.207	0.819	0.782	0.094	0.209	0.252	0.306	0.637
CG13510	96.341	82.683	3.659	6.098	33.415	4.878	51.951	3.659	2.683	3.659	6.098	6.585	4.878	8.049	0.008†	0.085	0.500	0.540	0.114†	0.435	0.117
CG9331	74.727	60.929	0.602	7.229	37.808	15.003	39.357	20.511	24.785	0.602	7.229	13.999	5.479	27.309	0.278	0.551	0.500	0.534	0.304†	0.589	0.662
cenG1a	60.950	47.821	0.649	3.896	51.139	13.272	31.044	14.375	6.725	0.649	3.896	7.303	2.313	14.161	0.280	0.529	0.500	0.565	0.307	0.606	0.684
CG5708	67.847	69.514	0.000	1.111	31.181	22.569	45.139	13.403	11.736	0.000	1.111	0.069	0.347	0.694	0.207	0.191	#DIV/0!	0.814	0.033†	0.171	0.138
png	73.528	49.738	4.294	7.339	57.238	16.190	14.133	2.278	3.488	0.544	2.339	4.012	0.060	0.383	0.026	0.412	0.080	0.192	0.295	0.349	0.317
plu	54.695	53.647	0.862	12.958	47.984	10.650	27.546	3.926	0.199	0.862	0.889	0.292	1.419	1.684	0.060	0.328	0.500	0.010	0.061	0.977	0.552
gnu	76.303	60.217	0.649	7.677	37.819	15.371	38.484	1.619	3.074	0.649	1.184	0.156	1.085	1.775	0.035	0.198	0.500	0.054	0.040†	0.383	0.174
cdc14	65.314	65.437	2.572	6.940	60.494	10.410	19.585	21.478	14.752	1.202	1.279	3.890	1.919	5.887	0.362	0.304	0.278	0.076	0.487	0.935	0.677
CG8455	41.782	34.940	3.134	5.359	56.316	13.254	22.596	27.309	5.849	0.502	0.096	3.684	7.799	2.859	0.791	0.651	0.101	0.098	0.233	0.808	0.913
pie	55.700	48.744	3.493	9.069	45.508	18.110	23.820	7.813	10.716	2.141	5.015	0.438	8.651	15.369	0.157	0.585	0.350	0.341	0.047	0.548	0.980
orc4	58.661	37.752	2.198	12.668	57.540	10.470	17.125	7.273	8.082	2.198	7.112	14.683	4.915	0.458	0.118	0.854	0.500	0.340	0.720	0.964	0.432

Appendix V

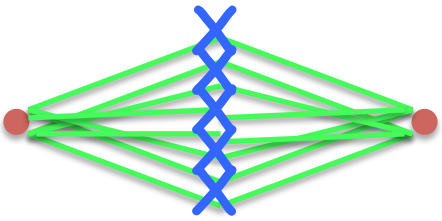
The raw data from the RNAi screen for each member of subcluster-16 is tabulated below. The table includes the sample size, percentage of mitotic cells normal/abnormal (N/Ab) spindle morphology and metaphase chromosome alignment and centrosome numbers (0-4+) for each replicate of the RNAi screen. Colour shades in the first column represent the type of protein each gene encodes – control (green), training set (red), predicted mitotic proteins (yellow), proteins from outside the MAP interactome (grey).

Gene	n	Spindle		DNA		Centrosome				
		N	Ab	N	Ab	0	1	2	3	4(+)
Blac - 12	36	55.6	44.4	69.4	30.6	0	13.9	63.9	11.1	11.1
Blac - 13	67	71.6	28.4	52.2	47.8	0	1.5	62.7	7.5	28.4
Blac - 14	71	63.4	36.6	67.6	32.4	0	0	66.2	14.1	18.3
Sak	33	12.1	87.9	54.5	45.5	21.2	48.5	27.3	3	0
	49	20.4	79.6	51	49	18.4	40.8	28.6	8.2	2
	78	25.6	74.4	39.7	60.3	34.6	47.4	14.1	0	3.8
mad2	67	19.4	80.6	32.8	67.2	1.5	6	52.2	14.9	25.4
	46	21.7	78.3	32.6	67.4	2.2	10.9	76.1	2.2	8.7
E2f2	76	39.5	60.5	39.5	60.5	2.6	6.6	52.6	19.7	18.4
	41	36.6	87.8	41.5	58.5	9.8	17.1	46.3	9.8	17.1
CG13900	90	26.7	73.3	51.1	48.9	0	3.3	47.8	25.6	23.3
	57	14	86	12.3	87.7	3.5	1.8	28.1	28.1	38.6
ptc	75	21.3	78.7	33.3	66.7	2.7	21.3	49.3	13.3	13.3
	22	18.2	81.8	31.8	68.2	0	13.6	45.5	9.1	31.8
akt1	78	57.7	42.3	53.8	46.2	0	3.8	66.7	10.3	19.2
	74	40.5	59.5	48.6	51.4	0	1.4	60.8	5.4	32.4
ald	71	29.6	70.4	15.5	84.5	2.8	5.6	59.2	14.1	18.3
	51	23.5	76.5	19.6	80.4	2	9.8	39.2	17.6	31.4
nej	20	10	90	15	85	0	0	15	5	80
	20	20	80	20	80	0	0	20	15	65
CG9986	67	11.9	88.1	11.9	88.1	3	13.4	43.3	13.4	26.9
	105	22.9	77.1	48.6	51.4	0	7.6	31.4	14.3	46.7
CG2865	98	20.4	79.6	54.1	45.9	0	12.2	46.9	16.3	24.5
	78	15.4	84.6	28.2	71.8	7.7	17.9	41	10.3	23.1
Bap60	55	9.1	90.9	29.1	70.9	0	10.9	56.4	12.7	20
	45	6.7	93.3	71.1	28.9	0	8.9	62.2	4.4	24.4
CG15676	44	13.6	86.4	25	70.5	6.8	22.7	50	2.3	18.2
	56	3.6	96.4	7.1	92.9	3.6	19.6	35.7	23.2	17.9
CG5568	71	97.2	7	88.7	11.3	0	9.9	69	14.1	7

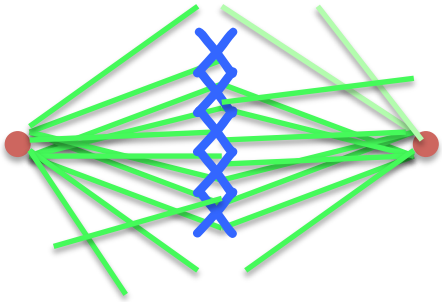
	63	14.3	85.7	17.5	82.5	9.5	15.9	55.6	6.3	12.7
rev7	50	58	42	54	46	2	20	50	24	6
	63	34.9	65.1	81	19	1.6	1.6	41.3	9.5	46
med15	79	34.2	65.8	21.5	78.5	1.3	10.1	45.6	25.3	17.7
	54	16.7	83.3	33.3	66.7	1.9	9.3	40.7	13	35.2
p53	104	33.7	66.3	39.4	60.6	9.6	11.5	45.2	11.5	22.1
	56	41.1	58.9	44.6	55.4	0	3.6	41.1	21.4	33.9
CG11486	76	22.4	77.6	35.5	61.8	0	3.9	48.7	11.8	35.5
	84	17.9	82.1	59.5	40.5	0	1.2	57.1	9.5	32.1
CG13625	62	48.4	51.6	41.9	58.1	0	21	40.3	14.5	24.2
	78	32.1	67.9	50	50	0	3.8	46.2	10.3	39.7
gus	78	74.4	25.6	64.1	35.9	2.6	9	51.3	21.8	15.4
	72	29.2	70.8	31.9	68.1	5.6	16.7	44.4	12.5	20.8
PHDP	89	24.7	75.3	42.7	62.9	1.1	6.7	59.6	20.2	12.4
	77	32.5	67.5	45.5	54.5	3.9	19.5	44.2	7.8	24.7
CG5731	67	43.3	56.7	43.3	56.7	1.5	4.5	44.8	19.4	29.9
	68	38.2	61.8	42.6	57.4	0	2.9	39.7	20.6	33.8
CG8128	74	81.1	18.9	66.2	33.8	5.4	10.8	39.2	25.7	18.9
	75	42.7	57.3	45.3	54.7	4	5.3	54.7	14.7	21.3
CG13510	41	7.3	92.7	14.6	85.4	7.3	12.2	26.8	9.8	43.9
	5	0	100	20	80	0	0	40	0	60
CG9331	83	45.8	54.2	63.9	36.1	1.2	14.5	51.8	20.5	12
	21	4.8	95.2	14.3	85.7	0	0	23.8	9.5	66.7
cenG1a	77	24.7	75.3	45.5	54.5	1.3	7.8	58.4	15.6	16.9
	73	53.4	46.6	58.9	41.1	0	0	43.8	11	45.2
CG5708	90	45.6	54.4	42.2	57.8	0	2.2	31.1	22.2	44.4
	48	18.8	81.3	18.8	81.3	0	0	31.3	22.9	45.8
png	80	28.8	71.3	53.8	46.3	3.8	5	61.3	16.3	13.8
	62	24.2	75.8	46.8	53.2	4.8	9.7	53.2	16.1	14.5
plu	58	41.4	58.6	46.6	53.4	1.7	12.1	48.3	12.1	25.9
	65	49.2	50.8	46.2	53.8	0	13.8	47.7	9.2	29.2
gnu	77	22.1	77.9	42.9	57.1	1.3	6.5	37.7	14.3	40.3
	79	25.3	74.7	26.6	63.3	0	8.9	38	16.5	36.7
cdc14	106	13.2	86.8	11.3	80.2	3.8	5.7	56.6	8.5	25.5
	73	56.2	43.8	49.3	50.7	1.4	8.2	64.4	12.3	13.7
CG8455	76	85.5	14.5	59.2	40.8	2.6	5.3	52.6	21.1	19.7
	55	30.9	69.1	70.9	29.1	3.6	5.5	60	5.5	25.5
pie	71	52.1	47.9	54.9	38	5.6	14.1	45.1	26.8	8.5
	74	36.5	63.5	40.5	59.5	1.4	4.1	45.9	9.5	39.2
orc4	91	34.1	65.9	70.3	29.7	4.4	19.8	42.9	15.4	17.6
	72	48.6	51.4	54.2	45.8	0	5.6	72.2	5.6	16.7

Appendix VI

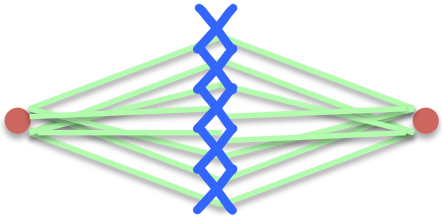
Schematic diagrams of the four spindle mutant phenotypes observed in the RNAi screen, compared to the wild-type diamond-shaped bipolar spindle.



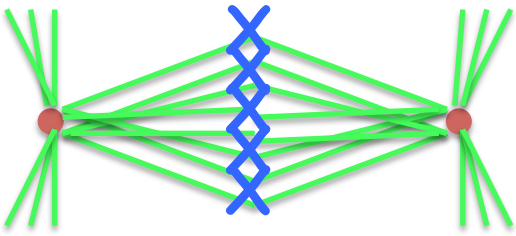
Wild-type spindles



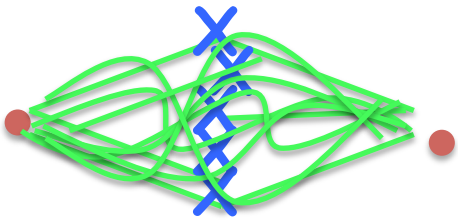
'spiky' spindles



'dim' spindles



'moustache' spindles



'mashed' spindles

Appendix VII

Table summarising the RNAi phenotypes along with results reported in the Vale screen, the Human orthologue and its GO annotation and the reported phenotype for human orthologue in the MitoCheck study. The type of the protein refers to its category in the MAP prediction model (T – training dataset, P – prediction dataset and O – outside the MAP interactome).

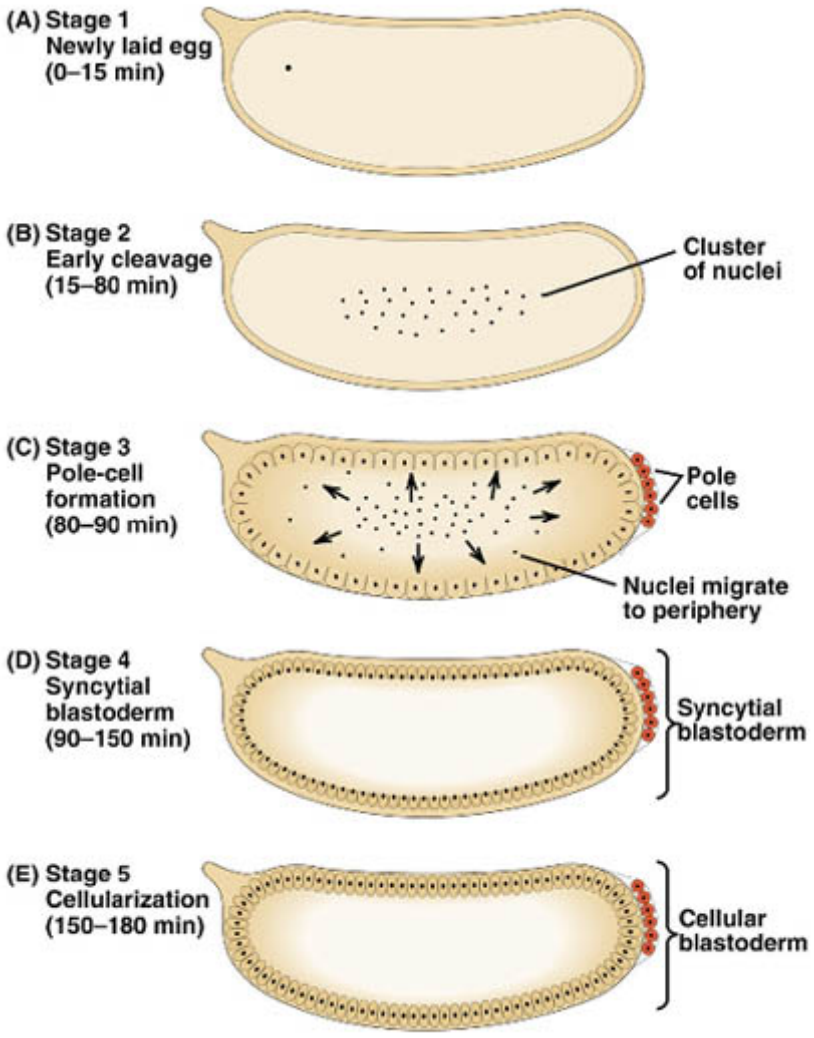
Type	Gene symbol	RNAi phenotype	Vale Screen	Human Orthologue and Gene Ontology	MitoCheck Phenotype
T	ald	Spindle	-	TTK - mitotic spindle assembly checkpoint	-
P	CG13510	Spindle	-	-	-
T	E2f2	Spindle	-	-	-
T	mad2	Spindle	-	MAD2L1 - mitotic spindle assembly checkpoint	Nuclear shape
P	PHDP	Spindle	Dim gamma tubulin (computer hit only)	DRGX - TF, organ development	-
P	CG9986	Spindle	-	Uncharacterized protein	-
T	CG13900	Spindle	Positive hit - weak long spindles	SF3B3 - splicing factor B3 subunit 3	Enhanced secretion
P	CG11486	Spindle	-	PAN3 - ribonuclease, polyA tail, decapping	-
O	png	Spindle	Dim tubulin in poles (computer hit only)	-	-
O	plu	Centrosome	-	-	-
P	med15	Centrosome	-	MED15 - RNApol TF	-
P	CG5708	Centrosome	-	LMO3 - LIM-domain only protein, regulation of transcription	Secretion inhibition
P	CG13625	Centrosome	-	BUD13, Ccoil, mRNA transport, splicing	-
O	pie	Centrosome	Chromosome misalign weak (manual hit only)	-	-
P	CG9331	Centrosome	-	GRHPR - metabolic enzyme	-

O	cdc14	Centrosome	High nuclear number (computer hit only)	cdc14A - Cell cycle phosphatase	Dynamic changes
P	cenG1a	Centrosome	-	AGAP7 - positive regulation of GTPase activity	-
T	ptc	Spindle & Centrosome	-	PTCH2 - hedgehog family protein binding, skin development	-
P	CG15676	Spindle & Centrosome	-	-	-
P	nej	Spindle & Centrosome	Few mitotics cells	CREBBP - histone acetylation signaling, morphogen, Cell cycle	Dynamic changes (EP300)
P	Bap60	Spindle & Centrosome	Long spindles (manual hit only)	SMARCD2- SW1/SNF related matrix assoc actin-dependent regulation of chromatin	-
T	Sak	Spindle & Centrosome	Positive hit - monoastral bipolar spindle	PLK4- known centriole duplication, pericentriole formation	Mitotic delay, pro-metaphase delay, segregation problems, metaphase alignment problems.
P	p53	Spindle & Centrosome	-	p53- growth arrest, apoptosis	-
O	gnu	Spindle & Centrosome	-	-	-
P	CG2865	Spindle & Centrosome	-	-	-
P	CG5731	Spindle & Centrosome	-	GLA - galactosidase, negative regulation of biosynthesis	-
P	gus	None	-	SPSB1 - protein ubiquitination	-
O	orc4	None	-	ORC4 - DNA replication origin binding	-
P	CG8128	None	-	NUDT6 - hydrolase activity/Growth factor activity	-
P	CG5568	None	-	-	-
O	CG8455	None	-	MPPE1 - metallophosphoesterase, ER-GC transport	-
P	rev7	None	-	MAD2L2 - cell division	-
T	akt1	None	-	AKT2 - kinase activity/phospholipid binding	Mild inhibition of secretion

Appendix VIII

Schematic diagram of *Drosophila* early embryogenesis - The newly laid egg is full of yolk and has one nucleus, which divides in the centre of the egg. Seven nuclear divisions happen in the centre and form a cluster of nuclei after which they start migrating towards the periphery of the egg. All nuclear divisions in the first 13 rounds take place without cell division, which makes the *Drosophila* egg a syncytium. Each of the 13 nuclear divisions is synchronous.

(Image taken from zoology.ubc.ca/~bio463)



Appendix IX

List of primers used for the amplification of cDNA for the transgenes are given below. Each forward primer was added with the 4 base pair sequence (CACC) on the 5'-end as required for the directional cloning into the pENTR™ vector as outlined in detail in section 4.2.1.

Primer	Sequence	GC%	Tm
CG2865-F	CACCATGACCTTGCCCACAAACAC	54	57.35
CG2865-R	TGTCTCCAACGAGGTGACCACC	59	56.49
CG5568-F	CACCATGGGTGAATTGCAAGCG	55	57.47
CG5568-R	ATTCTTCTCCATAATTCGTAAAAACTCAGC	33	57.93
CG5708-F	CACCATGAATGCAACGTTCCATTCATA	41	57.91
CG5708-R	GTCCTTTGATCTTCTGGGCCTACC	54	56.95
CG5731-F	CACCATGTTAGCTACTCTTTGGATCATATT	37	58.43
CG5731-R	CAAGTCAACTGGGTACTGAACTTCTCC	48	58.22
cenG1A-F	CACCATGAGTAACTACCATCCGCATTC	48	58.98
cenG1A-R	GATCACACCCGACACCGACTTCTC	58	58.61
gus-F	CAC CAT GCA CAT ATG CAT ATA CAT GTC TAT	37	58.66
gus-R	TCTACGGTTTTTATACAATAAATATGTTTTCATT	21	55.31
PHDP-F	CACCATGGAATTTTCACTGCTCAACAA	41	57.94
PHDP-R	TATTTGATCGTATAAGCATTTCATTTGACG	30	57.36

Appendix X

List of mass spectrometry results showing strong (score > 100) and weak (100 > score < 30) interactors of the four co-immunoprecipitation experiments reported in Chapter 4, i.e. GFP-CG5731, GFP-PHDP1, GFP-PHDP2 and Nej-YFP. The final list is the list is the raw mass spectrometry results for the negative control (GFP only) pull down, which constitutes the list of 'false positives' used to filter all results.

(a) Mass spectrometry results for the GFP-CG5731 co-IP experiment

Protein	FlyBase Accession	Coverage	# PSMs	# Peptides	# AAs	MW [kDa]	calc. pI	Score
CG5731-PA	FBpp0079549	34.14	65	11	413	47.0	5.59	228.65
Ubi-p5E-PA	FBpp0070894	72.10	52	7	534	60.0	7.62	188.96
msps-PE	FBpp0293342	26.23	53	40	2074	229.4	8.16	168.77
betaTub97EF-PB	FBpp0289838	18.82	51	10	457	51.2	4.86	150.67
Cp1-PC	FBpp0086719	38.27	30	12	371	41.6	7.21	98.61
CG30116-PC	FBpp0085898	1.24	22	1	1698	190.5	6.68	93.99
CG8036-PD	FBpp0081372	40.86	27	16	580	63.0	6.80	90.21
CG6540-PA	FBpp0074372	31.72	16	5	331	35.1	9.00	65.38
spict-PA	FBpp0079978	4.16	13	1	385	42.3	6.87	65.32
eIF-2gamma-PC	FBpp0302538	34.48	19	11	467	50.4	8.73	61.82
Mad1-PA	FBpp0087680	20.00	18	11	730	85.0	6.49	54.76
cup-PB	FBpp0288761	16.92	17	12	1117	125.6	8.85	54.25
CG11779-PB	FBpp0083164	27.80	16	11	428	48.9	6.48	49.51
Nup107-PA	FBpp0079710	17.75	16	11	845	97.3	5.94	48.58
Stat92E-PE	FBpp0088978	8.75	10	5	754	85.6	6.29	47.89

CG9853-PA	FBpp0078589	29.79	13	6	339	38.5	6.54	47.37
CG2118-PA	FBpp0085215	15.62	10	7	698	76.5	6.20	44.17
CG1598-PA	FBpp0088029	24.11	12	6	336	37.6	5.02	43.35
CG32066-PC	FBpp0076006	28.09	13	7	324	36.5	6.65	43.26
l(3)01239-PA	FBpp0075967	46.15	13	6	143	16.2	5.64	38.37
FBpp0288803	FBpp0288803	1.19	12	2	2264	255.4	9.58	37.39
Srp54k-PA	FBpp0076872	10.63	9	4	508	55.8	8.70	37.21
Dcr-1-PA	FBpp0083717	0.93	11	1	2249	255.2	5.72	36.90
CG34113-PO	FBpp0110103	1.18	4	1	1689	181.4	5.95	34.50
Osbp-PA	FBpp0084176	2.04	4	1	784	89.3	5.81	34.41

(b) Mass spectrometry results for the GFP-PHDP1 co-IP experiment.

Protein	**FlyBase Accession**	**Coverage**	**# PSMs**	**# Peptides**	**# AAs**	**MW [kDa]**	**calc. pI**	**Score**
Dhc64C-PA	FBpp0073215	25.95	136	86	4639	529.9	6.33	446.42
Ef1alpha100E-PA	FBpp0085193	41.13	149	16	462	50.6	8.95	441.46
PHDP-PA	FBpp0072133	79.09	130	17	220	25.1	9.14	415.83
msps-PD	FBpp0293341	34.44	99	57	2082	230.3	8.21	311.72
Ubi-p5E-PA	FBpp0070894	72.10	58	7	534	60.0	7.62	189.25
cup-PB	FBpp0288761	39.84	54	27	1117	125.6	8.85	184.74
cdm-PA	FBpp0083039	29.66	53	24	971	111.2	6.25	168.06
CG5794-PD	FBpp0084048	12.04	45	31	3912	440.1	6.21	159.95
coil-PD	FBpp0100132	39.59	38	20	634	70.5	5.86	136.50
rod-PA	FBpp0085156	18.76	42	32	2089	239.5	6.15	132.12
hyd-PA	FBpp0081568	14.28	42	31	2885	318.7	5.94	129.54
CG8184-PB	FBpp0073898	7.56	36	23	5146	556.5	5.44	125.10
Mad1-PA	FBpp0087680	31.51	41	21	730	85.0	6.49	123.00
CG3523-PA	FBpp0077343	15.91	39	24	2438	266.3	6.37	122.46
CG7546-PE	FBpp0293610	22.57	34	16	1298	138.0	5.91	119.35

rhea-PH	FBpp0305320	12.94	35	23	2689	291.4	6.06	111.33
Jabba-PI	FBpp0305366	22.17	32	8	406	45.9	6.19	104.06
fs(1)Ya-PA	FBpp0070463	27.16	25	12	696	77.7	9.42	100.24
Nipped-A-PE	FBpp0292385	8.15	29	23	3790	435.1	7.75	92.45
tud-PA	FBpp0071508	10.62	32	20	2515	285.1	6.46	92.11
Mtor-PA	FBpp0087140	10.23	29	17	2346	262.2	5.10	91.32
lds-PA	FBpp0081255	23.37	29	19	1061	118.3	6.65	90.99
Upf1-PA	FBpp0073433	21.27	30	20	1180	129.8	7.33	86.50
Nup107-PA	FBpp0079710	22.25	24	14	845	97.3	5.94	82.61
Ge-1-PA	FBpp0079818	16.10	27	17	1354	149.2	5.90	81.13
lok-PB	FBpp0080861	29.19	23	10	459	52.3	9.00	80.23
CG8036-PD	FBpp0081372	36.55	24	14	580	63.0	6.80	78.12
Nup93-1-PA	FBpp0073659	24.91	20	13	823	93.8	6.27	77.68
Rme-8-PA	FBpp0087699	12.79	27	21	2408	272.4	7.05	76.27
cnn-PD	FBpp0100116	18.17	25	16	1090	123.6	5.59	76.09
Cdc27-PA	FBpp0076600	16.00	25	12	900	101.2	7.43	76.03
Bruce-PC	FBpp0304071	4.33	20	16	4852	536.2	5.59	71.90
CG41099-PB	FBpp0112714	22.05	24	17	1111	123.0	6.02	71.77
CLIP-190-PM	FBpp0290655	13.31	22	16	1668	186.6	4.96	71.14
jar-PB	FBpp0084020	15.00	20	15	1253	143.2	8.63	70.56
CG12512-PA	FBpp0078711	26.31	18	9	593	65.2	7.69	67.05
nop5-PA	FBpp0078997	30.33	18	11	511	57.1	8.59	65.11
Grip84-PC	FBpp0074546	18.80	19	12	819	94.6	6.44	64.63
beta'Cop-PA	FBpp0080048	18.93	19	12	914	102.6	5.26	62.74
CG17514-PA	FBpp0112471	8.67	20	16	2630	293.7	7.03	62.23
Ulp1-PA	FBpp0074462	11.57	16	12	1513	172.2	5.10	61.16
ArfGAP1-PA	FBpp0075713	30.77	16	9	468	50.5	7.37	60.90
Nup160-PA	FBpp0079788	12.62	19	13	1411	160.2	5.36	60.84
CG2469-PB	FBpp0072561	12.78	17	11	1150	130.4	9.55	59.03
aub-PA	FBpp0079754	21.02	19	12	866	98.5	9.28	58.00
CG2118-PB	FBpp0085214	23.19	16	11	634	69.4	5.52	57.59

pzg-PA	FBpp0077992	15.66	17	10	996	105.0	4.89	56.40
Tor-PA	FBpp0080003	8.18	18	13	2470	280.9	6.92	56.34
Nipsnap-PD	FBpp0302001	44.59	17	8	222	26.2	8.31	55.89
CG31739-PA	FBpp0080475	15.80	19	12	1082	121.4	5.82	55.31
Sec13-PA	FBpp0083801	25.84	16	9	356	39.5	6.25	54.66
Nipped-B-PJ	FBpp0307004	8.02	18	11	1957	223.2	6.51	53.96
Cdc16-PA	FBpp0083769	15.60	15	8	718	81.7	5.52	53.16
Ctf4-PA	FBpp0086732	13.41	14	7	895	96.6	5.34	51.61
CG4119-PA	FBpp0070765	15.33	15	10	998	112.6	6.49	51.60
Nup133-PA	FBpp0083695	9.58	16	10	1200	135.1	5.63	51.29
CG2246-PD	FBpp0085038	40.60	16	11	367	40.5	7.52	50.82
Rab1-PA	FBpp0083503	48.78	15	7	205	22.7	5.47	50.65
CG10254-PC	FBpp0306810	12.04	17	11	1379	154.3	5.07	50.59
CG11779-PB	FBpp0083164	19.86	15	8	428	48.9	6.48	50.12
Wdr82-PA	FBpp0079266	40.06	16	8	317	35.3	7.24	49.95
shtd-PA	FBpp0073893	8.13	17	12	2030	227.1	6.51	49.04
tacc-PJ	FBpp0291771	13.85	13	9	1213	135.0	5.06	48.70
smg-PB	FBpp0076276	10.31	15	9	999	109.0	7.43	48.35
CG9684-PA	FBpp0081342	21.34	16	10	642	71.9	7.99	48.27
spd-2-PA	FBpp0075122	15.01	16	10	1146	124.5	6.27	47.38
CG8771-PB	FBpp0291752	7.45	14	9	1839	210.2	6.35	46.81
ctrip-PJ	FBpp0306924	4.95	14	10	2929	313.9	6.99	46.52
CG32649-PA	FBpp0073530	18.61	14	9	661	74.4	8.18	45.95
CG11984-PD	FBpp0289967	16.86	12	7	599	62.4	5.72	45.69
CG5604-PA	FBpp0079663	5.79	13	10	2727	302.0	5.43	45.29
muskelin-PC	FBpp0305316	14.54	13	7	832	95.1	6.32	45.19
RpI135-PA	FBpp0077714	9.48	13	9	1129	128.4	8.41	43.93
CG16940-PA	FBpp0072389	10.27	12	8	1032	116.8	8.18	43.85
lola-PG	FBpp0088381	13.24	13	8	891	96.2	6.46	43.51
lola-PD	FBpp0088383	19.12	13	9	748	79.4	5.71	43.16
Stat92E-PE	FBpp0088978	10.21	11	5	754	85.6	6.29	42.79

Nup93-2-PA	FBpp0082467	15.58	12	8	796	90.5	7.15	42.72
CG6540-PA	FBpp0074372	31.72	12	5	331	35.1	9.00	42.60
Grip75-PA	FBpp0079623	10.77	10	5	650	74.9	5.48	42.33
mahj-PA	FBpp0071575	10.30	16	14	1544	172.0	5.00	42.30
Nup75-PA	FBpp0085954	14.67	11	6	668	76.8	6.05	42.26
asp-PA	FBpp0084071	4.45	9	7	1954	230.0	10.78	42.18
CG9590-PA	FBpp0082626	24.35	12	8	460	50.6	4.69	41.65
gro-PH	FBpp0293584	18.05	13	8	709	77.8	7.96	41.34
BicC-PB	FBpp0080362	8.40	13	5	905	97.8	8.66	41.00
gnu-PA	FBpp0075447	42.50	11	6	240	27.2	10.35	40.97
mask-PE	FBpp0306961	3.65	10	8	4000	422.8	5.69	40.90
Mi-2-PD	FBpp0304462	6.79	12	10	1973	223.0	5.76	40.73
atms-PA	FBpp0078508	28.25	15	11	538	60.8	9.29	39.94
spict-PA	FBpp0079978	4.16	5	1	385	42.3	6.87	39.15
CG8963-PA	FBpp0086167	17.35	11	7	559	63.2	6.23	39.05
lola-PT	FBpp0088393	21.57	12	8	575	61.9	6.25	38.97
CG8503-PA	FBpp0086665	17.74	12	9	513	58.6	7.50	38.80
fzy-PA	FBpp0080391	16.16	12	5	526	57.0	8.72	38.77
Haspin-PA	FBpp0112507	20.32	13	8	566	65.3	8.27	38.22
ash2-PC	FBpp0084040	17.81	11	7	556	63.2	6.92	38.13
CG42388-PE	FBpp0289277	12.65	13	11	1225	138.5	7.23	38.06
egl-PB	FBpp0113078	10.26	12	8	1004	112.1	7.28	38.03
Lst8-PA	FBpp0071303	31.95	12	7	313	35.3	7.17	37.18
ArfGAP3-PG	FBpp0303208	17.30	10	6	549	58.6	7.71	37.07
par-1-PA	FBpp0085646	10.98	10	7	938	101.4	9.77	36.83
dre4-PB	FBpp0072744	7.66	11	7	1083	123.5	6.29	36.20
Dys-PC	FBpp0083179	4.64	12	10	3127	357.8	5.57	36.15
cmet-PE	FBpp0304513	5.40	11	9	2186	251.0	5.17	35.58
E(bx)-PF	FBpp0290563	3.65	8	5	2139	239.3	9.09	35.12
glu-PA	FBpp0080489	8.30	11	8	1409	159.8	6.15	35.10
Cul-4-PA	FBpp0087897	14.62	12	10	821	94.3	8.66	34.87

sced-PA	FBpp0085445	8.27	9	4	822	91.2	8.43	34.85
Cdc23-PA	FBpp0080894	16.22	12	8	678	77.5	6.51	34.53
SNF4Agamma-PR	FBpp0290803	11.86	10	9	1400	152.3	6.06	34.49
woc-PD	FBpp0306778	6.40	10	8	1687	187.2	4.75	34.36
Nup98-96-PA	FBpp0083851	4.69	11	8	1960	210.0	6.52	34.30
Rab5-PA	FBpp0077475	26.48	9	4	219	23.9	8.41	33.81
Srp54k-PA	FBpp0076872	18.90	10	7	508	55.8	8.70	33.75
l(1)G0222-PF	FBpp0290629	9.73	9	6	1110	123.0	8.25	33.28
DNApol-epsilon-PA	FBpp0083800	4.16	10	7	2236	256.5	6.47	33.11
lid-PA	FBpp0078862	6.80	10	8	1838	203.9	6.62	33.08
nxf2-PC	FBpp0303951	8.61	6	5	825	94.6	7.49	33.03
CG9853-PA	FBpp0078589	32.45	8	7	339	38.5	6.54	32.87
APC7-PA	FBpp0070907	13.98	10	6	615	70.3	6.98	32.19
Msp-300-PK	FBpp0304677	0.65	7	5	11917	1357.5	5.20	32.16
CG4538-PA	FBpp0083264	13.52	8	5	673	71.2	7.64	31.69
l(1)dd4-PA	FBpp0073672	14.72	12	10	917	103.6	7.37	31.62
shot-PH	FBpp0086747	0.56	5	3	8805	988.9	5.71	31.61
Torsin-PA	FBpp0070658	21.18	10	6	340	38.1	8.62	31.56
CG32066-PC	FBpp0076006	24.07	10	6	324	36.5	6.65	31.51
mbo-PA	FBpp0082137	10.11	10	6	702	78.6	5.91	31.50
mad2-PA	FBpp0076819	31.88	10	5	207	23.4	5.92	31.42
Cp190-PA	FBpp0082580	9.95	10	6	1096	121.6	4.67	31.03
Nup214-PA	FBpp0071905	6.49	10	7	1711	175.1	8.75	30.94
mit(1)15-PB	FBpp0070425	13.73	11	9	721	82.2	5.27	30.82
Bap170-PA	FBpp0085442	4.62	8	5	1688	182.9	8.46	30.60
mof-PA	FBpp0070794	14.39	11	7	827	92.6	4.79	30.43
Cpsf100-PA	FBpp0084726	10.05	10	6	756	85.4	5.47	30.39
CG1657-PA	FBpp0073304	6.43	10	9	1712	191.1	5.76	30.23

(c) Mass spectrometry results for the GFP-PHDP2 co-IP experiment.

Protein	FlyBase Accession	Coverage	# PSMs	# Peptides	# AAs	MW [kDa]	calc. pI	Score
PHDP-PA	FBpp0072133	53.18	33	11	220	25.1	9.14	104.30
CG10341-PA	FBpp0080647	19.86	26	8	564	62.2	4.72	97.44
CG9018-PA	FBpp0072788	48.00	30	14	375	41.9	7.46	96.81
Patr-1-PA	FBpp0082833	16.22	26	9	968	108.2	7.83	96.81
Bx42-PA	FBpp0071285	25.59	28	11	547	61.1	9.38	95.45
noi-PA	FBpp0078400	29.42	24	11	503	58.4	5.62	92.91
Snr1-PA	FBpp0078331	38.92	26	10	370	41.9	5.27	92.82
tgo-PA	FBpp0081483	25.70	23	10	642	71.4	7.02	90.06
CG1677-PA	FBpp0088653	15.50	25	9	1000	109.0	5.76	88.63
nito-PB	FBpp0087934	19.17	27	12	793	89.0	9.38	87.76
lark-PB	FBpp0076553	42.61	25	9	352	39.9	9.07	87.49
vir-PA	FBpp0071946	12.24	24	15	1854	208.9	5.94	82.81
CG32479-PA	FBpp0072486	15.56	23	12	1517	164.6	7.28	80.35
CG4598-PA	FBpp0079473	39.86	24	11	281	31.0	9.23	79.98
CBP-PB	FBpp0089293	30.30	18	6	231	26.8	6.54	77.73
pygo-PA	FBpp0085167	9.08	20	4	815	80.4	8.31	72.59
CG42303-PA	FBpp0288961	35.04	21	10	351	41.5	5.50	69.36
CG3689-PB	FBpp0076184	45.32	19	9	203	23.0	7.15	66.85
CG31548-PA	FBpp0078385	42.97	22	9	256	27.1	8.72	66.15
Pitslre-PA	FBpp0077905	23.11	20	16	952	108.8	8.05	65.40
sec24-PA	FBpp0087536	12.67	19	11	1184	129.3	7.62	64.79
btz-PA	FBpp0084590	19.58	18	8	761	83.6	5.49	64.35
scrib-PS	FBpp0307488	0.97	13	1	1859	199.7	5.02	60.55
psq-PJ	FBpp0100066	21.60	18	8	639	69.7	6.79	58.32

enc-PG	FBpp0305411	8.79	17	9	1798	187.7	8.37	56.87
CG2807-PA	FBpp0077728	15.15	18	12	1340	149.5	6.67	56.35
Rbp2-PB	FBpp0074056	37.96	18	9	324	34.9	9.26	56.25
RAF2-PB	FBpp0075144	13.70	15	10	1117	125.4	9.26	56.21
U4-U6-60K-PA	FBpp0074971	26.58	18	11	553	61.5	6.21	55.83
Lasp-PC	FBpp0297281	16.04	14	7	636	71.8	8.41	55.13
x16-PB	FBpp0304606	28.40	16	7	257	27.7	11.53	55.02
PDZ-GEF-PB	FBpp0289315	11.85	17	12	1569	171.0	7.36	54.85
CG30122-PB	FBpp0085841	7.24	12	6	1271	140.4	5.01	54.76
pea-PA	FBpp0086647	9.82	14	9	1242	141.8	8.76	54.16
CG13900-PA	FBpp0072494	11.65	14	10	1227	136.5	5.47	53.82
Nopp140-PB	FBpp0088542	23.62	14	10	686	70.6	8.88	53.29
pyd-PC	FBpp0293839	9.98	14	7	1293	140.5	9.38	53.09
CG8507-PA	FBpp0081618	31.66	15	10	379	44.6	8.15	50.31
sktl-PA	FBpp0071491	2.02	15	1	792	87.7	7.31	50.13
Iswi-PA	FBpp0086954	9.83	13	8	1027	118.8	8.29	48.63
Prp38-PA	FBpp0087663	28.79	15	7	330	40.1	8.98	48.42
Hakai-PA	FBpp0080804	19.54	13	3	302	32.5	8.41	47.22
atms-PA	FBpp0078508	23.05	15	10	538	60.8	9.29	46.90
CTPsyn-PB	FBpp0075389	14.93	15	8	623	69.7	7.17	46.42
U3-55K-PA	FBpp0090943	23.76	15	10	484	53.3	8.73	44.98
gw-PI	FBpp0302679	10.72	13	8	1381	142.6	6.74	44.74
Unr-PB	FBpp0113084	7.47	12	6	1057	118.4	6.55	44.50
CG4119-PA	FBpp0070765	9.52	11	5	998	112.6	6.49	44.29
Fip1-PA	FBpp0078568	7.85	13	4	701	78.6	5.44	43.14
CG5913-PA	FBpp0084312	16.08	11	5	454	49.9	9.70	42.31
SmE-PA	FBpp0079182	54.26	13	3	94	11.1	9.63	42.28
CG5726-PA	FBpp0085951	9.01	12	5	766	86.5	7.74	41.02
NHP2-PA	FBpp0075448	23.75	13	5	160	17.7	8.10	40.69

QC-PA	FBpp0076788	19.41	11	4	340	38.0	7.36	40.63
SC35-PB	FBpp0088838	20.00	12	3	195	21.4	11.75	40.08
CG10492-PA	FBpp0080692	10.12	10	4	810	89.0	7.55	40.04
CG1635-PA	FBpp0085157	24.11	10	7	448	51.1	9.06	39.52
Liprin-beta-PB	FBpp0075555	16.32	13	7	680	75.6	7.78	38.82
Prp3-PB	FBpp0074661	14.96	10	7	528	60.1	9.73	38.08
SmG-PA	FBpp0074151	57.89	13	3	76	8.5	8.79	36.38
CG11089-PA	FBpp0084087	21.02	12	8	590	63.3	7.87	35.92
CG17168-PA	FBpp0112453	19.48	10	6	421	49.7	9.88	35.16
CG10672-PA	FBpp0076882	14.51	8	5	317	33.6	9.33	34.87
fd68A-PO	FBpp0304899	11.54	9	4	745	80.4	7.88	34.81
CG10555-PA	FBpp0071122	2.59	8	1	926	92.9	8.51	34.60
smt3-PA	FBpp0078984	34.44	12	3	90	10.1	5.45	34.01
SmF-PA	FBpp0087118	32.95	12	3	88	9.7	4.54	33.69
U2af38-PA	FBpp0077792	23.11	10	4	264	29.9	8.91	33.66
CG7028-PA	FBpp0072427	11.91	9	7	907	104.0	9.50	33.20
cup-PB	FBpp0288761	9.67	11	8	1117	125.6	8.85	33.06
CG6540-PA	FBpp0074372	28.10	9	4	331	35.1	9.00	32.90
d4-PC	FBpp0085419	9.49	10	5	495	55.1	7.83	32.89
wds-PA	FBpp0070422	25.76	9	6	361	39.0	8.57	32.89
CG4896-PD	FBpp0077633	7.80	8	4	949	107.6	8.22	32.64
fd68A-PK	FBpp0291179	11.22	9	4	740	80.0	7.88	32.01
Syp-PF	FBpp0083368	1.70	4	1	529	59.4	8.07	31.13
CG3605-PA	FBpp0077353	7.21	8	5	749	84.7	5.34	30.77
mle-PA	FBpp0085367	6.81	10	7	1293	143.6	7.23	30.59
Psi-PB	FBpp0086220	14.70	10	8	796	81.5	8.13	30.44
Hrb87F-PA	FBpp0082316	23.38	8	6	385	39.5	9.03	30.24
Nup153-PA	FBpp0074059	2.81	6	3	1883	196.5	8.13	30.16
Nipsnap-PD	FBpp0302001	30.18	8	4	222	26.2	8.31	30.14

(d) Mass spectrometry results for the Nej-YFP co-IP experiment.

Protein	FlyBase Accession	Coverage	# PSMs	# Peptides	# AAs	MW [kDa]	calc. pI	Score
HDAC6-PE	FBpp0298372	28.52	51	26	1108	122.6	6.35	179.63
Ubi-p5E-PA	FBpp0070894	78.65	44	9	534	60.0	7.62	160.84
Mad1-PA	FBpp0087680	42.60	53	33	730	85.0	6.49	159.36
nej-PE	FBpp0305701	6.43	46	11	3266	339.5	8.69	157.94
CG11779-PB	FBpp0083164	49.30	49	20	428	48.9	6.48	150.13
DNApol-epsilon-PA	FBpp0083800	14.45	46	25	2236	256.5	6.47	146.98
msi-PH	FBpp0305821	32.10	41	16	567	58.6	9.26	142.41
CG3532-PB	FBpp0112136	28.63	37	26	1135	130.0	6.04	128.42
Gmap-PA	FBpp0073859	20.60	31	23	1398	158.4	4.78	104.59
CG9590-PA	FBpp0082626	28.70	29	9	460	50.6	4.69	104.38
tral-PD	FBpp0304883	45.67	29	16	646	68.7	9.54	101.54
Rox8-PB	FBpp0083977	29.96	27	12	464	49.5	7.85	98.46
CG5261-PB	FBpp0079072	27.73	28	14	512	54.2	9.57	88.17
Srp54k-PA	FBpp0076872	35.83	26	15	508	55.8	8.70	83.16
mad2-PA	FBpp0076819	42.51	21	6	207	23.4	5.92	80.61
cup-PB	FBpp0288761	19.43	20	13	1117	125.6	8.85	78.83
coil-PD	FBpp0100132	29.65	23	16	634	70.5	5.86	76.29
pum-PA	FBpp0081470	14.81	23	12	1533	157.4	7.65	74.11
kuk-PA	FBpp0082838	23.68	22	10	570	60.1	5.48	74.04
kis-PB	FBpp0077804	11.62	19	14	2151	224.6	6.21	70.48
Pole2-PA	FBpp0076789	31.62	18	11	525	58.7	6.48	69.58
CG4119-PA	FBpp0070765	22.34	18	11	998	112.6	6.49	67.15
CG7546-PE	FBpp0293610	16.26	20	12	1298	138.0	5.91	65.16
nop5-PA	FBpp0078997	29.35	17	10	511	57.1	8.59	61.55

cbs-PC	FBpp0301621	3.99	18	2	601	69.6	5.26	55.69
Crag-PC	FBpp0290510	11.86	16	14	1644	183.9	6.32	54.87
Nup93-1-PA	FBpp0073659	19.44	15	11	823	93.8	6.27	54.15
CG1648-PA	FBpp0087577	47.39	16	9	230	23.8	8.91	54.09
gw-PI	FBpp0302679	12.38	16	9	1381	142.6	6.74	53.67
Ahcy13-PA	FBpp0073847	23.38	15	11	432	47.3	6.20	48.81
Sap47-PG	FBpp0082664	25.05	12	7	535	55.3	4.55	47.77
spd-2-PA	FBpp0075122	12.39	12	10	1146	124.5	6.27	46.05
CaBP1-PA	FBpp0080395	20.09	14	8	433	46.7	5.69	45.38
Sec16-PF	FBpp0301987	6.53	13	10	2250	245.7	5.82	43.53
smg-PB	FBpp0076276	10.61	13	9	999	109.0	7.43	43.14
CG17333-PA	FBpp0073235	40.33	12	7	243	26.7	8.21	42.77
stai-PC	FBpp0078827	32.30	14	7	257	29.6	8.25	42.50
ArfGAP1-PA	FBpp0075713	18.16	13	7	468	50.5	7.37	41.66
CG11486-PC	FBpp0072864	14.49	12	8	780	84.2	7.62	40.03
CG32521-PI	FBpp0304199	19.20	13	4	323	29.9	8.03	38.85
Nup98-96-PA	FBpp0083851	6.94	10	9	1960	210.0	6.52	37.49
CG1578-PA	FBpp0073444	10.43	12	8	901	97.5	6.05	37.18
Nup153-PA	FBpp0074059	7.54	12	8	1883	196.5	8.13	36.87
me31B-PB	FBpp0079566	24.30	11	9	428	48.6	7.97	36.72
Rme-8-PA	FBpp0087699	4.78	11	8	2408	272.4	7.05	36.22
Nup107-PA	FBpp0079710	11.36	10	6	845	97.3	5.94	36.22
Prat-PA	FBpp0081261	15.75	10	7	546	59.5	6.67	35.35
SdhB-PA	FBpp0085489	33.00	14	9	297	33.7	8.66	35.26
eIF-2gamma-PC	FBpp0302538	25.48	11	8	467	50.4	8.73	34.48
Nup50-PA	FBpp0087861	19.68	12	7	564	59.4	8.43	33.67
CG32500-PA	FBpp0077066	31.80	10	6	283	31.3	5.40	33.58
CG32230-PA	FBpp0070039	55.42	11	6	83	9.4	9.26	32.73
porin-PA	FBpp0079771	25.89	7	6	282	30.5	6.96	32.61

CG6540-PA	FBpp0074372	31.72	8	5	331	35.1	9.00	32.44
Vha68-3-PA	FBpp0080002	8.75	10	5	743	82.3	5.21	32.37
CG32066-PC	FBpp0076006	29.32	8	7	324	36.5	6.65	32.05
CG5174-PA	FBpp0085865	43.75	9	8	208	22.5	5.33	32.03
GstO2-PB	FBpp0076377	16.00	9	3	250	28.7	7.05	30.85

(e) Mass spectrometry results for the Negative control (GFP only) co-IP experiment.

Protein	**FlyBase Accession**	**Coverage**	**# PSMs**	**# Peptides**	**# AAs**	**MW [kDa]**	**calc. pI**	**Score**
Prp19	FBpp0085902	30.50	14	9	505	55.2	6.90	52.42
CG12321	FBpp0083036	27.53	9	5	247	27.8	6.23	30.36
RpS27	FBpp0084242	41.67	15	4	84	9.4	9.52	50.41
CG8778	FBpp0086993	28.43	11	6	299	31.9	8.22	36.22
Lam	FBpp0078733	31.67	19	13	622	71.3	6.47	71.45
eIF3-S8	FBpp0086013	20.55	26	14	910	105.6	6.06	87.40
eIF-3p40	FBpp0078667	28.11	12	7	338	38.4	6.10	40.63
RnrS	FBpp0087152	15.27	11	7	393	45.1	5.63	30.84
UGP	FBpp0290912	15.26	10	6	511	57.8	7.40	31.01
CG16935	FBpp0086748	27.45	10	6	357	39.1	9.11	31.05
CG14476	FBpp0070057	16.99	10	10	924	105.7	6.51	31.15
Prosbeta7	FBpp0078448	19.78	8	4	268	30.0	6.58	31.18
Pp4-19C	FBpp0077016	18.89	9	5	307	35.3	5.26	31.20
RanBPM	FBpp0087357	19.06	10	5	598	66.4	6.68	32.65
CG32626	FBpp0290266	12.84	9	7	771	89.2	5.97	31.44
sn	FBpp0071057	21.09	10	6	512	57.2	7.01	31.51
RpL8	FBpp0072801	39.06	15	8	256	27.9	11.15	56.77
GstE7	FBpp0085856	23.32	10	3	223	25.5	6.57	31.71
Arp3	FBpp0076460	23.21	10	6	418	47.0	5.99	32.10

cdc2c	FBpp0083329	28.98	10	7	314	35.9	8.24	30.03
CG5214	FBpp0081834	12.39	8	4	468	49.9	9.47	30.71
RpLP1	FBpp0077716	50.89	11	4	112	11.5	4.41	47.40
dj-1beta	FBpp0085065	26.83	7	4	205	21.4	7.94	32.45
RfC4	FBpp0073120	32.02	15	7	331	37.1	7.72	50.79
Rpi	FBpp0296955	36.51	10	7	241	26.5	5.88	32.50
CG5941	FBpp0070832	58.24	9	8	182	20.3	9.11	32.54
PpD3	FBpp0081607	18.27	10	6	520	59.2	6.23	32.72
ND75	FBpp0071128	17.37	10	8	731	78.6	6.84	32.73
Srp72	FBpp0083319	21.69	16	10	650	72.8	9.06	46.23
Mapmodulin	FBpp0086001	27.59	10	7	261	29.2	4.21	32.84
CG3902	FBpp0074843	25.36	9	8	414	45.3	6.71	32.84
CG17331	FBpp0080490	37.31	10	6	201	22.5	6.39	32.92
bl	FBpp0290974	17.04	10	5	493	49.8	8.35	33.06
vas	FBpp0304748	17.10	15	8	661	72.3	5.66	46.23
CG1532	FBpp0076960	32.29	9	6	288	31.6	5.39	33.19
CG3107	FBpp0085414	9.86	10	8	1034	119.2	6.81	33.20
fon	FBpp0080764	13.10	10	5	565	56.6	6.46	33.20
CG4603	FBpp0076867	24.50	9	5	347	37.8	4.82	33.28
CG33714	FBpp0076925	66.67	10	5	90	10.0	9.64	33.40
Dlic	FBpp0073299	19.27	10	8	493	54.4	6.02	33.55
pix	FBpp0076284	21.11	13	10	611	69.3	8.13	33.57
RhoGDI	FBpp0074665	34.83	10	6	201	23.2	5.64	33.61
Gpdh	FBpp0078777	26.39	10	7	360	39.3	6.61	33.64
CG30382	FBpp0087968	18.85	9	5	244	27.1	7.66	30.65
ApepP	FBpp0080509	17.78	9	8	613	68.5	5.95	33.78
CG5261	FBpp0079073	13.06	9	5	421	44.1	8.82	33.88
CG9281	FBpp0073872	16.53	15	9	611	69.5	7.53	40.71
Rya-r44F	FBpp0087719	0.88	8	3	5113	579.3	5.60	33.99

CG6907	FBpp0078729	25.17	10	7	437	48.7	5.88	34.00
CG4365	FBpp0077910	40.22	15	9	271	30.3	6.07	47.31
La	FBpp0080905	37.18	34	13	390	44.9	7.58	118.75
und	FBpp0079462	14.06	9	5	448	49.8	6.68	31.98
RpL23A	FBpp0072687	18.05	9	6	277	29.4	10.95	34.42
CG2246	FBpp0307424	26.40	10	6	356	39.3	7.25	34.47
vimar	FBpp0089175	17.67	9	8	634	70.2	5.16	34.96
spel1	FBpp0080203	12.32	12	10	917	103.2	5.92	35.02
Snap	FBpp0074609	35.27	9	6	292	33.0	5.45	35.06
CG14434	FBpp0070999	42.25	11	6	187	21.1	5.33	35.29
Tal	FBpp0072050	25.68	8	4	331	36.7	7.71	35.54
RpLP2	FBpp0086252	46.90	10	5	113	11.7	4.70	34.77
CG10289	FBpp0304174	8.48	10	5	967	108.4	4.50	35.63
sec31	FBpp0087723	10.40	12	9	1240	135.9	6.14	39.25
Aos1	FBpp0082065	26.41	11	7	337	37.6	5.44	36.45
CG5642	FBpp0075754	23.93	16	9	539	63.2	5.96	49.77
CG12702	FBpp0074578	10.57	12	6	870	97.1	6.71	36.05
Prosalpha7	FBpp0089041	23.32	14	5	253	27.7	5.69	41.59
CG5174	FBpp0085863	40.86	12	7	186	20.5	6.60	36.41
UbcD4	FBpp0076124	37.19	11	5	199	22.5	5.47	32.01
pic	FBpp0082177	8.25	11	8	1140	126.0	5.36	43.12
P32	FBpp0086051	32.70	12	5	263	29.0	5.08	36.46
sqd	FBpp0082319	28.35	11	7	321	35.0	6.58	36.85
lin19	FBpp0087921	12.14	9	7	774	89.5	7.90	30.66
GstS1	FBpp0086157	34.14	9	6	249	27.6	4.65	36.90
Eip55E	FBpp0085889	22.90	11	7	393	43.0	7.62	36.91
CG6045	FBpp0082605	8.77	10	6	1254	137.8	6.57	37.39
eIF4AIII	FBpp0081324	13.03	11	5	399	45.6	6.02	37.46
Vps4	FBpp0074278	27.38	22	10	442	49.6	6.96	66.64

Map205	FBpp0306414	13.80	17	10	1109	118.1	4.83	50.67
CG2051	FBpp0078318	16.54	10	6	405	47.9	6.73	37.73
Dhc64C	FBpp0271878	11.64	47	37	4638	529.6	6.32	178.29
Ntf-2	FBpp0076934	80.00	14	6	130	14.6	5.73	49.67
RpL10	FBpp0289147	36.70	12	6	218	25.5	9.85	38.01
DNApol-delta	FBpp0075277	13.28	18	12	1092	124.8	6.92	60.51
CG17896	FBpp0070088	26.42	9	7	511	55.0	8.35	32.80
Moe	FBpp0071215	18.95	12	10	512	60.6	5.80	38.41
CG9135	FBpp0078816	26.90	11	9	487	53.4	7.18	38.42
FK506-bp1	FBpp0082574	21.57	11	7	357	39.3	4.79	36.61
DNA-ligI	FBpp0072041	17.80	17	11	747	84.7	6.62	46.73
Caf1	FBpp0082511	24.65	12	7	430	48.6	4.89	48.81
Actn	FBpp0070329	15.53	12	11	895	103.8	5.64	38.62
RpL18	FBpp0076602	34.57	11	5	188	21.7	11.53	30.49
Tango7	FBpp0086705	28.68	13	9	387	44.1	5.71	38.95
RpL18A	FBpp0086103	10.73	7	2	177	21.0	10.62	33.69
r-l	FBpp0083440	29.41	11	9	493	53.4	7.25	39.05
Trn	FBpp0076708	14.89	11	8	893	101.5	5.07	38.84
SelD	FBpp0086690	20.10	11	6	398	43.4	6.93	39.32
zip	FBpp0291731	9.42	18	13	1964	226.6	5.52	59.92
Prx5	FBpp0100079	54.74	12	7	190	19.9	8.68	39.46
CG18815	FBpp0075834	50.00	11	7	216	23.1	6.28	39.63
RpL26	FBpp0074833	32.89	11	4	149	17.3	10.95	35.97
Pdcd4	FBpp0073671	25.15	10	7	509	56.3	5.97	40.08
AnnX	FBpp0070024	37.19	14	9	320	35.6	4.77	40.12
CG13625	FBpp0084051	14.53	141	10	647	76.5	10.29	512.87
ATPsyn-gamma	FBpp0084905	35.69	17	8	297	32.9	9.22	59.23
bur	FBpp0290556	30.60	16	14	683	76.7	6.80	54.70
Prosbeta2	FBpp0075382	46.69	12	8	272	29.8	8.66	40.79

CG12171	FBpp0078357	28.79	12	7	257	26.9	7.50	37.88
alphaCop	FBpp0072693	14.91	20	14	1234	139.2	7.65	70.72
Dmn	FBpp0087722	38.68	14	10	380	42.0	5.26	41.25
cathD	FBpp0087972	26.02	12	5	392	42.4	6.32	44.69
rl	FBpp0112423	35.64	15	9	376	43.1	6.07	48.14
CG1354	FBpp0088409	23.43	10	5	397	44.9	6.71	35.78
CG5590	FBpp0084585	22.33	8	6	412	44.3	8.02	31.30
Scsalpha	FBpp0073010	24.39	14	5	328	34.4	8.98	57.84
Npl4	FBpp0084266	21.96	14	10	624	69.9	6.73	42.82
RpS18	FBpp0085585	46.71	17	10	152	17.6	10.48	52.13
CG9149	FBpp0072560	29.59	13	8	392	41.1	7.05	42.90
CG2852	FBpp0071844	60.00	18	11	205	22.2	8.75	63.48
Rm62	FBpp0078299	45.74	27	17	575	62.4	9.57	100.17
nudC	FBpp0075087	28.92	15	8	332	37.8	5.64	47.19
RpL13	FBpp0079484	27.06	11	5	218	24.9	10.99	35.83
Paf-AHalpha	FBpp0073975	24.00	11	4	225	25.4	5.81	36.75
deltaCOP	FBpp0291138	19.77	10	7	531	57.8	6.20	33.87
CG32473	FBpp0082212	14.40	13	10	903	102.6	5.12	44.11
Rpd3	FBpp0073173	13.24	11	5	521	58.3	5.76	44.28
RpL30	FBpp0291653	53.15	12	4	111	12.2	9.58	49.19
spag	FBpp0072164	17.04	11	7	534	59.6	8.34	44.60
RpS11	FBpp0087115	41.94	17	7	155	18.1	10.93	64.01
eIF-4B	FBpp0112402	25.45	11	6	389	44.1	6.30	40.10
CG16817	FBpp0081560	29.35	14	6	184	20.7	4.60	45.55
Ranbp9	FBpp0081862	21.51	25	16	1018	114.3	4.87	79.83
CG6767	FBpp0076144	32.57	16	8	350	38.3	7.44	48.81
Pglym78	FBpp0084753	42.75	14	9	255	28.6	6.89	45.62
Sod	FBpp0075958	55.56	14	7	153	15.7	6.11	51.17
Prx3	FBpp0082927	46.58	12	7	234	26.4	7.49	45.83

CG15093	FBpp0085821	28.70	10	5	324	33.9	8.13	34.14
Rpn9	FBpp0083860	20.68	9	6	382	43.7	5.26	32.80
rept	FBpp0074756	17.05	12	6	481	53.5	5.85	37.80
Aats-arg	FBpp0073965	22.41	15	12	665	75.5	7.44	46.29
Uch-L5	FBpp0076200	41.36	20	11	324	37.6	5.21	64.14
Gs1	FBpp0077774	35.59	14	9	399	44.4	6.46	46.44
RpL12	FBpp0072084	44.85	14	5	165	17.7	9.07	46.19
CG7911	FBpp0084971	43.59	15	6	156	17.4	4.56	55.77
clu	FBpp0086380	8.22	13	9	1448	160.8	6.60	48.22
ALiX	FBpp0084610	8.73	9	6	836	92.5	5.45	30.20
GstE6	FBpp0085855	25.23	8	3	222	25.0	6.23	31.83
rngo	FBpp0074054	24.24	13	7	458	50.5	5.16	47.82
CG9953	FBpp0076608	36.22	22	11	508	56.8	6.60	75.34
RpL7	FBpp0079536	39.29	17	8	252	29.5	10.89	59.40
pch2	FBpp0081381	17.34	10	6	421	46.5	5.69	30.32
CG8858	FBpp0087116	8.73	13	11	1890	212.0	6.86	45.67
EndoGI	FBpp0080382	37.05	13	8	359	40.6	4.60	48.56
eRF1	FBpp0304201	14.87	9	5	437	49.0	6.09	31.21
AGBE	FBpp0086845	26.28	15	12	685	79.1	6.25	48.70
Elf	FBpp0305303	20.61	14	11	495	55.3	7.20	49.26
Pp1-87B	FBpp0082067	22.52	10	6	302	34.5	5.59	35.00
rin	FBpp0082265	25.36	22	11	690	74.9	7.37	74.41
Pfk	FBpp0087507	16.88	14	8	788	86.7	6.76	49.38
Trip1	FBpp0078689	37.42	22	9	326	36.1	5.34	76.66
His4:CG33909	FBpp0091154	52.43	13	7	103	11.4	11.36	42.50
gho	FBpp0302999	10.39	13	8	1193	129.2	7.25	46.04
CG7461	FBpp0289511	33.97	24	15	627	68.3	7.77	78.31
RpL27A	FBpp0307135	35.57	17	6	149	17.0	10.71	56.60
Nc73EF	FBpp0075028	8.83	9	5	1008	112.5	6.89	36.83

RpL10Ab	FBpp0075764	38.71	21	8	217	24.3	9.86	67.28
CG10306	FBpp0071587	33.78	16	6	222	25.6	6.43	51.17
Rpn11	FBpp0078664	21.43	7	4	308	34.4	6.13	33.24
Thiolase	FBpp0072135	56.29	45	17	469	50.6	9.14	135.26
RanBP3	FBpp0307011	27.68	12	6	448	47.3	4.60	51.57
CG5706	FBpp0084021	30.90	14	11	589	65.7	6.04	53.07
Cdc37	FBpp0072660	16.20	12	6	389	45.1	5.06	39.34
Aats-lys	FBpp0071301	16.38	18	8	574	64.6	6.51	59.60
Pros29	FBpp0071451	39.77	14	7	264	29.4	7.15	53.31
Drp1	FBpp0077424	27.07	24	15	735	82.5	6.98	85.56
Txl	FBpp0076804	39.02	13	7	287	31.7	5.83	53.94
Aats-glupro	FBpp0083898	19.08	37	22	1714	189.3	8.63	141.01
RpS12	FBpp0075612	63.31	20	11	139	15.2	6.38	68.89
CSN4	FBpp0087935	24.57	10	7	407	46.4	6.32	30.72
Dak1	FBpp0084349	36.76	15	10	253	27.8	8.00	54.85
RpL5	FBpp0110421	28.43	20	7	299	34.0	9.77	57.50
RpL3	FBpp0081822	35.82	22	10	416	46.9	10.24	77.31
Hel25E	FBpp0078754	23.11	9	7	424	48.6	5.66	33.22
l(1)G0255	FBpp0070915	23.13	10	5	467	50.4	7.53	55.73
RpS9	FBpp0076152	40.51	20	9	195	22.6	10.61	60.85
ade5	FBpp0305933	29.67	17	10	428	47.1	8.38	56.03
Aats-his	FBpp0074365	17.43	15	7	522	57.7	6.95	56.08
RpS20	FBpp0083371	23.33	12	3	120	13.5	10.33	41.37
cdc2	FBpp0079641	42.09	15	10	297	34.4	6.87	55.60
Chc	FBpp0073966	21.33	38	26	1678	191.1	5.72	119.32
Aats-tyr	FBpp0075168	27.05	18	11	525	58.1	6.87	57.13
RpS6	FBpp0071087	31.05	21	9	248	28.4	10.74	62.73
eEF1delta	FBpp0079541	46.72	16	11	229	25.8	4.74	57.73
Uba2	FBpp0076457	16.29	12	8	700	77.6	5.02	41.72

sta	FBpp0070277	58.89	16	12	270	30.2	4.87	58.03
CG10932	FBpp0071095	40.73	23	11	410	43.4	8.68	95.72
CG31075	FBpp0084452	28.25	15	7	485	52.6	6.67	58.43
CG3226	FBpp0070953	27.39	9	5	230	25.8	6.96	31.63
Cp1	FBpp0292990	57.43	25	12	249	27.0	5.44	86.46
RpS14b	FBpp0071052	49.01	19	8	151	16.3	10.35	57.93
sar1	FBpp0083604	56.48	15	9	193	21.8	6.93	59.06
Dph5	FBpp0083673	50.18	18	10	281	31.6	5.52	59.19
CG9674	FBpp0075100	3.78	6	5	2114	231.9	6.42	30.79
FK506-bp2	FBpp0085703	73.15	17	7	108	11.7	8.13	59.40
VhaSFD	FBpp0080496	25.62	14	8	441	50.7	6.54	59.42
CG6180	FBpp0079979	34.63	14	6	257	28.7	8.82	58.55
14-3-3zeta	FBpp0087500	25.81	9	5	248	28.3	4.93	32.51
Prosbeta1	FBpp0086400	43.75	12	7	224	24.2	5.25	43.26
ras	FBpp0071423	30.17	16	10	537	57.8	7.43	51.23
eIF-4E	FBpp0076216	33.87	16	6	248	27.8	5.39	56.93
Psa	FBpp0072581	23.44	24	16	866	99.3	5.30	67.68
CG5028	FBpp0084275	41.04	20	11	402	44.4	6.80	78.11
Jabba	FBpp0289219	36.88	40	11	320	35.9	5.29	153.76
Aats-asp	FBpp0086897	30.51	17	11	531	59.0	6.81	58.70
Ef1beta	FBpp0305165	44.14	16	8	222	24.2	4.58	51.42
Hrb27C	FBpp0078974	31.83	13	7	421	44.7	6.80	51.18
Plap	FBpp0077645	16.90	19	10	787	85.7	5.83	66.82
sec23	FBpp0078359	15.39	12	7	773	86.7	6.79	51.23
eEF1delta	FBpp0079542	29.30	11	7	256	28.9	4.87	39.62
polo	FBpp0074608	37.50	31	16	576	66.9	8.88	100.13
Usp7	FBpp0073474	18.42	19	14	1129	130.4	6.04	62.49
eIF-2alpha	FBpp0074075	39.59	16	8	341	38.6	4.94	64.90
Fer2LCH	FBpp0084986	40.53	11	6	227	25.2	6.34	38.26

14-3-3zeta	FBpp0087502	25.81	10	5	248	28.2	4.88	39.74
lic	FBpp0073551	40.12	16	10	334	38.2	6.39	53.71
me31B	FBpp0079565	50.98	34	18	459	51.9	7.64	130.06
Arp1	FBpp0082121	41.49	18	10	376	42.7	7.30	58.45
RpS16	FBpp0071766	51.35	29	9	148	16.8	10.17	85.59
RpS19a	FBpp0074087	45.51	26	11	156	17.3	10.11	81.14
Droj2	FBpp0082219	33.50	20	10	403	45.2	6.48	63.92
tws	FBpp0081666	27.54	17	7	443	50.8	6.42	60.42
CG3731	FBpp0082459	32.98	16	11	470	51.8	6.00	49.57
tsr	FBpp0072097	62.16	26	10	148	17.1	7.17	85.88
fabp	FBpp0099726	55.38	18	7	130	14.5	5.66	68.56
lig	FBpp0293224	11.46	15	7	1300	130.6	7.01	66.49
l(1)G0334	FBpp0070676	40.60	20	12	399	43.9	7.68	66.76
Rpn12	FBpp0075068	38.64	20	9	264	30.2	6.06	66.80
Rad23	FBpp0088191	33.57	16	7	414	45.8	4.67	67.03
REG	FBpp0073561	47.76	21	9	245	28.1	6.05	68.59
awd	FBpp0085223	44.19	24	7	172	19.2	8.62	84.55
Rpn7	FBpp0083687	28.79	19	10	389	45.4	6.48	63.92
eIF3-S10	FBpp0078584	23.07	29	23	1140	133.8	9.01	100.30
r	FBpp0088675	7.91	16	12	2224	246.5	6.64	49.60
RpL7A	FBpp0070877	33.21	22	11	271	30.7	10.71	70.46
RpA-70	FBpp0081356	32.50	22	13	603	66.6	6.76	68.19
Rpt6R	FBpp0085042	22.56	13	7	399	45.1	7.97	43.84
Aats-gln	FBpp0084240	21.21	15	12	778	87.5	7.21	47.34
Inos	FBpp0088368	23.01	10	7	565	62.2	6.23	33.81
pont	FBpp0081704	36.62	18	12	456	50.2	7.34	61.44
eIF-5A	FBpp0072081	44.65	20	6	159	17.6	5.15	69.61
CG3011	FBpp0288674	20.77	11	7	467	51.0	7.72	35.12
CG5355	FBpp0079637	19.71	17	10	756	86.3	6.19	61.25

RanGap	FBpp0306292	34.69	24	15	588	65.1	4.65	84.59
CG11876	FBpp0084735	20.55	12	6	365	39.3	7.80	34.02
Hsc70-1	FBpp0075504	11.08	19	5	641	70.6	5.49	64.84
Eb1	FBpp0085461	46.05	16	10	291	32.6	5.34	55.68
eIF3-S9	FBpp0086097	28.84	26	12	690	80.4	6.28	95.55
flr	FBpp0075582	33.94	22	14	604	66.1	6.73	71.78
wal	FBpp0087186	28.79	17	7	330	34.2	8.32	54.04
dUTPase	FBpp0303622	40.80	11	7	174	18.5	5.43	38.81
Pros28.1	FBpp0073989	44.18	15	8	249	28.0	8.12	58.49
RpS2	FBpp0079500	44.94	25	10	267	28.9	10.15	90.64
SpdS	FBpp0081556	31.36	21	9	287	32.3	5.78	72.37
CG33123	FBpp0077251	20.81	22	15	1182	134.8	7.64	76.85
Hsc70-5	FBpp0086694	18.22	16	11	686	74.0	6.35	49.27
exba	FBpp0078393	24.41	12	8	422	49.2	5.74	42.52
Prosalpha5	FBpp0086066	44.26	20	9	244	26.9	4.98	73.28
Df31	FBpp0099688	69.95	19	10	183	18.8	4.27	74.29
Oscp	FBpp0082522	48.33	16	10	209	22.4	9.63	58.34
CG17259	FBpp0077351	16.57	9	5	501	56.4	6.49	30.97
eIF4G	FBpp0088303	26.83	49	29	1666	183.8	7.83	170.15
tral	FBpp0075691	39.11	19	12	652	69.3	9.54	75.28
Uch	FBpp0077516	47.58	12	7	227	25.8	5.49	43.48
skap	FBpp0306436	42.13	38	12	451	49.0	6.83	129.03
CG11980	FBpp0081444	20.29	10	5	345	39.2	6.54	31.31
Dip-B	FBpp0082299	17.91	8	6	508	55.5	6.76	33.03
bel	FBpp0081374	32.71	36	18	798	85.0	7.53	131.72
RpL4	FBpp0084617	32.92	27	9	401	45.0	11.47	108.96
RpS7	FBpp0088970	49.48	26	10	194	22.2	9.80	85.61
Vha26	FBpp0078350	57.08	23	13	226	26.1	6.15	79.93
GstD3	FBpp0082042	48.24	17	8	199	22.9	5.47	64.68

bsf	FBpp0080637	29.82	51	31	1412	157.2	7.14	164.14
Khc	FBpp0086328	17.44	23	13	975	110.3	5.77	76.97
CG6439	FBpp0083588	36.49	40	13	370	40.4	8.54	148.07
CG1416	FBpp0085258	39.55	25	14	354	40.1	6.57	81.02
GstT1	FBpp0087548	50.44	19	12	228	26.6	8.97	55.89
Dp1	FBpp0085859	13.68	23	16	1301	144.2	6.20	83.78
CG8728	FBpp0087958	28.78	19	8	556	61.0	7.12	81.81
Mcm3	FBpp0070729	18.80	22	13	819	90.9	6.40	73.52
shi	FBpp0305867	32.49	22	16	834	93.4	7.94	83.59
CG7920	FBpp0084968	31.66	19	10	477	51.8	8.12	73.43
chic	FBpp0078864	57.94	13	3	126	13.7	5.41	57.69
RpS3A	FBpp0088242	53.36	34	13	268	30.3	9.61	116.93
CG10576	FBpp0076827	28.90	15	9	391	42.7	7.11	45.68
Trap1	FBpp0085482	19.10	20	10	691	77.9	7.59	62.40
Art1	FBpp0081780	16.49	9	5	376	42.8	5.15	31.86
CG7322	FBpp0074425	52.48	14	8	242	25.6	7.12	48.33
Nurf-38	FBpp0271761	75.52	34	15	290	32.6	5.68	135.55
Rpt4	FBpp0293948	53.65	29	15	397	44.9	8.07	89.31
Mdh1	FBpp0079640	38.58	18	8	337	36.0	7.39	68.44
Ahcy13	FBpp0304841	34.67	17	9	323	35.6	6.09	52.17
eIF-2gamma	FBpp0302537	48.42	29	17	475	51.5	8.73	103.72
RpLP0	FBpp0078134	36.91	21	10	317	34.2	6.95	75.31
Tctp	FBpp0081820	45.35	18	8	172	19.6	4.81	60.48
CG13349	FBpp0112985	39.15	23	9	424	45.5	5.96	89.98
Mov34	FBpp0072197	33.73	17	8	338	38.1	6.71	55.15
Rpn5	FBpp0078278	23.11	14	9	502	57.7	5.80	57.34
CG6543	FBpp0086770	44.07	27	12	295	31.6	8.63	90.19
pAbp	FBpp0085915	42.90	46	19	634	69.9	9.31	160.46
Aats-val	FBpp0086847	21.35	25	18	1049	118.2	6.65	91.63

Pros35	FBpp0079538	45.16	16	9	279	31.0	6.55	51.98
DppIII	FBpp0081330	17.15	13	10	723	81.9	5.71	37.89
CG7834	FBpp0084901	47.04	17	10	253	27.2	8.05	55.57
Mcm6	FBpp0070913	10.77	13	7	817	92.3	5.38	40.52
CG13349	FBpp0086736	29.56	17	7	389	42.0	5.64	68.57
CG12262	FBpp0076520	29.59	16	9	419	45.8	7.94	62.52
EfTuM	FBpp0086790	32.72	15	9	489	54.0	8.03	53.23
Mcm5	FBpp0081756	24.69	23	14	733	82.2	7.84	63.52
RpS8	FBpp0099686	60.58	27	12	208	23.7	10.48	106.65
Vha55	FBpp0082139	22.04	12	8	490	54.5	5.40	42.89
Amun	FBpp0073452	33.45	16	9	550	58.4	4.58	60.71
CG5384	FBpp0079650	28.84	18	10	475	53.7	6.25	62.65
CkIIalpha	FBpp0110433	62.50	35	15	336	39.9	7.24	129.16
Pp2A-29B	FBpp0099974	32.99	27	15	591	65.4	5.05	89.69
CG6904	FBpp0082494	27.43	32	13	689	79.2	6.60	121.84
Cand1	FBpp0079643	22.20	29	17	1248	139.3	5.87	83.54
Ald	FBpp0084367	44.88	23	12	361	39.0	7.40	82.73
Tpi	FBpp0084949	51.01	29	11	247	26.6	6.00	106.62
Lsd-2	FBpp0073784	55.40	53	17	352	38.2	8.41	198.08
CG7430	FBpp0074906	29.37	18	11	504	53.1	6.87	61.38
CG15100	FBpp0085809	29.06	30	18	1022	112.4	8.51	113.83
CG32165	FBpp0303948	16.24	22	12	1059	118.8	4.81	75.45
Idh	FBpp0076390	23.56	14	9	416	46.6	6.74	57.01
CG10602	FBpp0088356	30.83	22	11	613	68.4	5.50	84.07
Mdh2	FBpp0082985	64.88	27	14	336	35.3	9.11	101.44
HIP	FBpp0070575	24.67	12	9	377	41.0	5.35	42.34
Gp93	FBpp0084623	17.15	22	13	787	90.2	5.02	74.17
Fs(2)Ket	FBpp0294037	30.77	40	21	884	98.6	5.03	131.09
CG6287	FBpp0079809	54.52	31	16	332	35.2	7.36	107.30

Nat1	FBpp0074480	26.85	32	18	890	103.0	7.27	107.54
CG3590	FBpp0082816	40.75	28	15	481	53.8	7.49	107.78
26-29-p	FBpp0075508	30.78	29	11	549	62.1	6.74	92.85
mts	FBpp0079148	47.25	19	10	309	35.4	5.34	65.93
Aats-thr	FBpp0079913	30.72	22	15	690	79.3	6.90	63.62
Got2	FBpp0077536	43.40	26	11	424	47.2	8.78	88.12
Pros26.4	FBpp0083906	33.26	23	11	439	49.3	6.58	73.29
lost	FBpp0078562	54.21	54	24	535	58.9	9.32	196.40
Aats-ile	FBpp0078150	22.95	27	21	1229	141.0	7.52	83.62
Mcm2	FBpp0081317	20.86	23	13	887	100.4	5.11	74.88
Tsf1	FBpp0074331	22.00	16	9	641	71.8	7.06	51.26
Adh	FBpp0100045	48.83	17	8	256	27.7	7.43	64.85
RpS10b	FBpp0074500	60.62	32	9	160	17.9	9.80	97.35
CG17337	FBpp0085374	30.54	18	11	478	53.1	5.66	54.40
emb	FBpp0079278	20.60	23	15	1063	122.7	5.87	73.76
msk	FBpp0076408	33.37	48	27	1049	119.2	4.86	163.47
CG6084	FBpp0303937	56.33	33	15	316	35.9	6.80	116.93
Rpn3	FBpp0291495	34.01	22	15	494	56.0	9.00	63.38
Cas	FBpp0080484	24.92	34	19	975	110.1	5.74	103.00
CG1516	FBpp0087547	28.54	35	25	1181	130.8	6.81	117.24
mus209	FBpp0089395	46.15	20	9	260	28.8	4.81	75.16
Mcm7	FBpp0076312	24.58	26	12	720	81.2	6.99	92.64
Rpt1	FBpp0088021	32.33	23	12	433	48.5	6.04	76.10
Fdh	FBpp0081767	40.90	25	10	379	40.4	6.71	91.68
RpS4	FBpp0075618	57.85	51	16	261	29.1	10.18	165.52
grsm	FBpp0288712	39.21	25	14	533	57.2	6.21	88.10
Jafrac1	FBpp0073594	57.22	47	12	194	21.7	5.71	162.55
Argk	FBpp0076271	28.01	18	8	432	47.9	6.04	71.88
Aldh	FBpp0079406	48.08	33	16	520	57.0	6.80	130.88

ATPCL	FBpp0289825	18.90	16	12	1095	119.7	7.12	53.66
Gdh	FBpp0088988	42.35	37	21	562	62.5	8.27	128.51
Gdi	FBpp0079458	38.60	34	12	443	49.9	5.72	132.64
ERp60	FBpp0087164	50.10	35	17	489	55.3	5.87	141.85
Rack1	FBpp0079187	61.01	55	18	318	35.6	7.47	202.91
Rpn6	FBpp0086604	35.07	25	12	422	47.2	5.88	79.77
scu	FBpp0074285	77.25	52	13	255	26.9	9.03	203.75
Pros45	FBpp0076890	52.59	32	15	405	45.8	8.41	106.76
Nap1	FBpp0072128	40.54	65	16	370	42.7	4.79	228.16
ran	FBpp0073327	53.24	28	10	216	24.7	7.81	87.02
Tudor-SN	FBpp0072419	34.67	40	20	926	103.0	8.05	135.22
Aats-gly	FBpp0075385	32.55	25	14	679	75.8	6.40	82.22
Gdh	FBpp0088990	59.20	65	26	549	61.0	8.40	244.30
Cyp1	FBpp0074017	48.02	37	10	227	24.7	9.22	151.61
Vha68-2	FBpp0079999	29.97	22	14	614	68.3	5.34	71.61
Hop	FBpp0077790	33.67	23	10	490	55.7	6.81	84.60
CG7433	FBpp0074677	43.00	25	15	486	54.6	8.62	78.57
Rpt3	FBpp0073292	23.49	17	9	413	47.0	5.38	47.94
Pgi	FBpp0087760	45.34	32	16	558	62.3	7.12	114.50
CG12082	FBpp0072976	15.84	17	11	827	92.0	5.49	52.97
dpa	FBpp0088055	35.57	34	20	866	96.5	7.52	125.55
CG8223	FBpp0081398	42.89	36	13	492	51.9	4.39	154.13
Rpn1	FBpp0074662	30.69	40	20	919	102.2	5.85	143.57
blw	FBpp0071794	44.75	57	23	552	59.4	9.01	196.34
capt	FBpp0099390	44.81	28	14	424	45.6	7.06	92.06
l(1)G0156	FBpp0074549	60.73	79	23	354	38.6	7.36	255.66
Acon	FBpp0081002	30.88	24	16	787	85.3	8.24	86.94
Pxt	FBpp0082932	65.51	67	34	809	90.5	6.90	249.28
Ef1gamma	FBpp0084761	42.00	41	16	431	48.9	7.05	172.96

TppII	FBpp0086888	21.71	37	21	1354	148.8	6.99	120.15
GstD1	FBpp0082077	40.19	28	9	209	23.9	7.23	96.98
Tbp-1	FBpp0083843	37.15	31	13	428	47.8	5.34	114.35
GlyP	FBpp0077501	24.88	26	15	844	96.9	6.52	98.31
Trxr-1	FBpp0071115	54.99	32	15	491	53.2	6.33	134.40
FKBP59	FBpp0079468	46.01	42	16	439	48.8	5.41	137.18
betaTub60D	FBpp0072177	29.07	57	11	454	50.8	4.88	180.18
yip2	FBpp0079472	72.11	60	20	398	41.6	8.51	226.23
RpS27A	FBpp0079606	42.31	29	6	156	17.9	9.77	98.10
Pdi	FBpp0075401	60.28	38	22	496	55.7	4.82	122.37
RnrL	FBpp0079648	43.47	77	28	812	91.9	7.46	261.74
Aats-ala	FBpp0089366	16.98	17	10	966	107.7	6.13	70.36
Hsp27	FBpp0076182	59.62	40	10	213	23.6	7.44	158.17
14-3-3epsilon	FBpp0082989	47.66	21	12	256	29.2	4.84	69.76
Hsc70-3	FBpp0073445	47.10	76	25	656	72.2	5.36	267.16
Pgk	FBpp0077419	43.61	29	14	415	43.8	7.42	108.70
Karybeta3	FBpp0078500	33.03	57	26	1105	123.5	4.73	201.30
RpS3	FBpp0083802	69.51	84	21	246	27.5	9.39	296.04
Mtpalpha	FBpp0079453	52.02	84	34	744	79.6	8.85	286.79
Pen	FBpp0079527	56.70	70	23	522	57.8	5.35	253.88
T-cp1	FBpp0083683	52.60	86	28	557	59.5	6.39	299.55
Eno	FBpp0077575	49.88	52	18	433	46.6	6.55	184.58
CG5525	FBpp0079992	55.53	83	24	533	57.1	7.56	309.17
Rpn2	FBpp0084754	27.75	40	22	1020	113.1	5.19	123.77
CG8258	FBpp0087764	63.37	79	30	546	59.4	5.31	272.20
Act57B	FBpp0290168	55.00	72	18	360	40.2	5.83	261.71
CG8036	FBpp0081370	49.68	64	22	626	68.0	7.11	229.46
alphaTub85E	FBpp0081565	40.31	72	14	449	49.9	5.20	240.14
eIF-4a	FBpp0078806	44.91	43	16	403	45.8	5.66	158.25

Tcp-1zeta	FBpp0073902	61.91	117	26	533	58.2	6.62	419.88
Cct5	FBpp0087096	53.91	65	26	512	55.8	6.13	197.42
Cct5	FBpp0087095	68.82	76	31	542	59.2	6.25	273.22
Tcp-1eta	FBpp0081401	64.15	88	29	544	59.3	6.34	295.11
Cctgamma	FBpp0082787	51.29	91	25	544	59.4	6.80	324.80
Rfabg	FBpp0088252	26.65	95	62	3351	372.4	7.97	321.19
CG7033	FBpp0071226	69.16	70	31	535	58.0	5.85	247.68
ATPsyn-beta	FBpp0088250	54.85	62	19	505	54.1	5.27	192.16
Hsp60	FBpp0073290	58.64	68	27	573	60.8	5.49	240.20
Act5C	FBpp0070788	63.03	81	21	376	41.8	5.48	295.54
alphaTub67C	FBpp0076122	55.84	65	19	462	51.1	5.26	231.65
TER94	FBpp0087479	42.70	58	25	801	88.8	5.35	206.89
Uba1	FBpp0087583	36.94	77	33	1191	130.7	5.29	277.50
Hsc70Cb	FBpp0075513	51.12	70	29	804	88.4	5.43	253.84
Hsp26	FBpp0076224	70.19	56	11	208	23.0	7.56	205.41
betaTub85D	FBpp0081524	34.30	160	18	446	49.8	4.83	469.75
CG4893	FBpp0075183	57.81	120	20	192	20.6	9.20	385.22
CHORD	FBpp0083939	33.33	15	8	354	40.2	6.11	54.32
Hsc70-4	FBpp0082514	58.68	152	38	651	71.1	5.52	532.22
Ef2b	FBpp0085265	60.43	114	40	844	94.4	6.60	387.86
PyK	FBpp0083610	63.28	66	25	512	55.0	7.85	253.97
alphaTub84D	FBpp0081062	72.00	106	23	450	49.9	5.14	367.80
Yp2	FBpp0071359	79.41	212	32	442	49.6	7.96	818.23
Yp3	FBpp0073652	79.05	240	36	420	46.1	8.50	849.29
Yp1	FBpp0071354	85.88	360	30	439	48.7	7.69	1376.06
Ef1alpha48D	FBpp0087142	65.87	208	24	463	50.3	9.07	708.95
Gapdh1	FBpp0087977	87.95	430	27	332	35.3	8.18	1687.32
Hsp83	FBpp0072904	50.35	139	42	717	81.8	5.02	480.23
betaTub56D	FBpp0085720	75.39	273	31	447	50.1	4.86	926.19

Gapdh2	FBpp0073922	87.65	484	28	332	35.3	8.44	1896.16
msps	FBpp0290695	17.14	32	25	2042	225.8	8.06	101.90
CG8003	FBpp0075989	19.90	11	6	407	45.7	7.62	35.98
betaCop	FBpp0074348	16.08	13	9	964	107.3	6.32	49.27
CG4169	FBpp0075069	34.55	16	9	440	45.4	9.44	51.57
ferrochelatase	FBpp0085208	37.76	15	9	384	43.6	8.40	61.12
CG6412	FBpp0080547	26.42	9	6	318	35.4	6.87	30.78
RpII140	FBpp0082353	12.67	12	11	1176	134.0	7.05	43.19
CG3800	FBpp0071903	33.33	8	3	165	17.6	8.65	36.31
CG12264	FBpp0079874	22.51	9	8	462	51.0	8.21	33.76
Cul-2	FBpp0085254	11.16	10	6	753	87.3	6.92	30.89
CTPsyn	FBpp0075387	12.28	11	6	627	69.4	7.24	32.42
faf	FBpp0085203	4.76	11	9	2711	303.9	6.10	41.52
Sod2	FBpp0086226	29.03	12	4	217	24.6	7.93	50.86
Lsd-2	FBpp0110268	59.40	51	17	335	36.3	8.79	196.46
eIF5	FBpp0089137	21.77	17	11	464	51.7	5.26	58.54
alpha-Adaptin	FBpp0088490	16.91	16	12	940	105.6	7.18	57.76
Gl	FBpp0075498	12.02	16	10	1265	141.1	5.62	57.73
CG18190	FBpp0271746	41.53	18	9	248	28.7	5.55	55.62
rig	FBpp0085564	9.72	15	9	1235	137.7	6.77	55.30
sle	FBpp0081710	7.89	17	7	1420	158.6	5.24	53.60
CG9577	FBpp0070006	27.56	12	5	312	33.8	7.46	53.24
Pep	FBpp0089253	14.86	16	8	693	75.6	5.44	49.70
RpS13	FBpp0079328	30.46	12	5	151	17.2	10.55	49.61
nonA	FBpp0100043	11.03	10	5	698	77.0	9.25	49.18
Nop56	FBpp0083625	17.54	16	7	496	54.8	9.28	48.83
poe	FBpp0079167	3.12	17	13	5322	590.3	6.38	47.74
Bap	FBpp0074558	16.83	15	12	921	101.1	5.11	47.55
RpS15Aa	FBpp0073623	46.92	13	7	130	14.8	9.80	47.45

CG7546	FBpp0076506	10.03	15	6	1166	124.3	5.14	47.36
Pros26	FBpp0075119	31.49	15	6	235	25.8	6.54	46.83
CG30268	FBpp0271819	2.18	4	2	1240	144.8	6.90	45.91
Fib	FBpp0071892	37.79	13	7	344	34.6	10.29	45.23
CG2025	FBpp0073438	7.06	12	6	1147	132.7	5.15	44.16
yps	FBpp0297134	22.35	13	6	340	37.0	10.01	44.06
RfC3	FBpp0079609	27.71	12	6	332	37.4	7.20	43.57
Srp68	FBpp0076244	23.18	15	11	604	69.0	8.47	43.49
128up	FBpp0087084	27.45	14	8	368	41.1	8.60	43.30
RpS25	FBpp0081845	41.03	13	6	117	13.2	10.27	43.16
nocte	FBpp0305169	4.38	10	7	2305	234.9	9.25	42.41
RfC38	FBpp0079812	32.58	14	8	356	40.8	8.40	42.32
Rab11	FBpp0083414	48.60	12	7	214	24.2	5.73	41.94
rump	FBpp0081601	15.19	12	6	632	66.7	7.85	41.85
ncd	FBpp0084900	17.14	10	7	700	77.4	9.17	41.67
fax	FBpp0099912	19.52	11	6	415	46.6	5.25	41.33
CG6693	FBpp0081843	20.07	12	5	299	34.9	8.59	41.26
gammaCop	FBpp0290505	14.46	11	8	878	97.1	5.30	40.77
vig2	FBpp0084151	13.35	13	5	412	45.2	9.41	40.28
RpS17	FBpp0076207	37.40	14	5	131	15.3	9.94	40.24
Prp8	FBpp0087124	4.47	8	8	2396	279.4	8.88	39.79
Zn72D	FBpp0075249	12.44	10	7	884	96.0	8.44	39.58
RpS5b	FBpp0082465	20.43	12	5	230	25.7	8.54	39.05
Int6	FBpp0075104	21.15	9	7	435	51.1	6.79	37.76
AGO2	FBpp0075312	20.84	11	6	1214	136.8	9.61	37.68
RpS23	FBpp0086701	37.06	10	7	143	16.0	10.59	37.63
Klp61F	FBpp0072616	12.10	11	7	1066	121.1	6.43	37.59
BubR1	FBpp0085368	8.08	11	7	1460	165.0	6.33	37.28
Atx2	FBpp0082555	14.48	11	8	1084	117.5	8.91	36.82

mRpS29	FBpp0071753	19.13	12	6	392	44.3	8.41	36.66
RpS5a	FBpp0074180	20.61	11	5	228	25.4	8.63	34.71
DnaJ-1	FBpp0076830	25.45	9	5	334	37.0	8.85	34.63
Capr	FBpp0074865	9.89	11	7	961	103.5	6.51	34.41
bic	FBpp0086895	41.42	9	4	169	17.7	7.49	34.29
alt	FBpp0088431	11.52	10	6	842	95.0	7.09	34.18
Prosbeta3	FBpp0081488	28.29	11	5	205	23.2	5.44	33.72
His2B:CG33882	FBpp0091127	29.27	12	4	123	13.7	10.35	33.25
CG9769	FBpp0078532	29.64	10	6	280	31.1	6.52	33.07
RpS28b	FBpp0071295	60.00	11	5	65	7.5	10.37	33.02
CG12163	FBpp0078464	12.21	8	4	475	53.5	8.05	32.80
Trx-2	FBpp0079436	23.58	11	3	106	11.7	4.88	32.67
Klp10A	FBpp0073331	10.56	12	6	805	88.6	7.21	32.48
Imp	FBpp0073274	17.84	9	6	566	62.1	9.61	32.10
RpII215	FBpp0073387	4.66	8	6	1887	209.0	7.81	31.69
RpL6	FBpp0085165	27.57	10	6	243	27.7	10.68	31.58
RpL23	FBpp0071808	32.86	9	3	140	14.9	10.83	31.49
l(2)k09913	FBpp0071872	21.52	9	5	316	36.5	9.38	31.43
CG10863	FBpp0073028	18.04	10	4	316	36.3	7.08	31.18
lig	FBpp0303480	8.90	7	5	1079	107.7	6.33	31.07
RpL27	FBpp0084306	39.26	9	5	135	15.9	10.61	30.76
Vha36-1	FBpp0086468	23.98	10	6	246	27.6	9.54	30.36
Nacalpha	FBpp0086971	19.35	9	4	217	23.0	4.74	30.17
Act42A	FBpp0085365	80.59	292	29	376	41.8	5.48	1151.27
Act87E	FBpp0082253	63.03	164	24	376	41.8	5.48	566.18
Act88F	FBpp0082597	63.03	153	25	376	41.7	5.48	486.58
Act79B	FBpp0078131	55.32	90	17	376	41.8	5.48	327.75
CG9485	FBpp0071525	41.50	78	39	1542	173.5	6.47	279.37
l(2)s5379	FBpp0100140	53.74	39	14	294	33.5	5.44	137.28

CG11899	FBpp0084759	48.35	29	13	364	39.5	8.41	117.72
Aats-trp	FBpp0081460	49.07	30	16	430	47.9	6.86	110.33
CtBP	FBpp0306581	40.80	28	14	473	50.3	6.83	97.73
Sptr	FBpp0071184	72.03	30	17	261	29.2	7.03	96.41
CtBP	FBpp0099514	41.18	25	14	476	50.7	6.83	86.91
yl	FBpp0073714	7.69	17	9	1937	214.8	6.23	68.25
Arc42	FBpp0083227	32.84	19	10	405	43.6	7.88	64.45
shi	FBpp0110335	20.96	16	13	830	93.0	8.24	63.64
CG12018	FBpp0072658	27.15	19	9	431	48.0	6.52	61.48
Msh6	FBpp0075399	16.05	20	15	1190	133.1	7.28	59.67
Prx6005	FBpp0077368	46.40	17	9	222	24.8	5.36	59.39
CG1218	FBpp0078279	28.06	20	10	449	51.1	7.03	59.33
CG4390	FBpp0083246	20.63	15	3	286	31.4	5.83	57.77
Got1	FBpp0086371	43.75	16	12	416	46.1	7.53	57.41
CG31549	FBpp0078356	46.30	16	10	257	26.9	7.09	55.83
CG4752	FBpp0071743	15.46	17	13	1294	139.5	6.47	55.60
CG9547	FBpp0078893	26.97	14	7	419	45.7	8.51	54.07
CG3714	FBpp0306839	21.64	16	10	550	61.5	6.71	52.70
wmd	FBpp0072023	32.32	15	8	328	36.1	6.52	49.67
Aats-asn	FBpp0080705	32.08	14	12	558	63.9	5.96	49.49
CG7261	FBpp0077502	11.19	14	10	1189	134.4	5.92	48.68
Sam-S	FBpp0088446	26.47	15	9	408	44.7	6.54	48.21
alph	FBpp0084808	37.23	15	8	368	40.9	5.30	47.52
Oga	FBpp0083452	10.99	13	8	1019	113.8	4.88	47.06
CG2915	FBpp0087954	25.61	15	10	453	51.0	6.13	46.56
MAPk-Ak2	FBpp0070802	35.38	13	9	359	41.4	7.33	44.79
SdhA	FBpp0085736	21.48	15	11	661	72.3	7.12	44.69
CG3663	FBpp0072230	30.77	14	5	208	23.3	7.69	44.37
CG1236	FBpp0078370	28.82	13	9	347	38.4	8.27	43.44

CkIIbeta	FBpp0073403	37.67	13	7	215	24.8	5.55	43.09
ACC	FBpp0087948	3.70	11	5	2323	261.8	6.07	42.90
Gmd	FBpp0078685	20.25	11	5	395	44.8	7.30	42.53
CG8031	FBpp0082187	29.15	12	7	295	33.2	6.93	41.97
Sgt	FBpp0080535	25.08	13	8	331	36.2	4.64	41.60
B52	FBpp0082270	27.96	12	8	329	37.5	11.47	40.79
GstE12	FBpp0072341	36.32	13	5	223	25.2	6.29	40.76
GstZ1	FBpp0081522	35.37	15	9	246	27.9	8.90	40.65
Tbc1d15-17	FBpp0077784	13.85	14	7	715	81.8	5.19	40.47
CG8207	FBpp0086475	16.44	11	6	438	48.1	7.72	40.27
CG11334	FBpp0085169	37.36	13	11	364	39.1	7.01	40.13
CG11107	FBpp0088040	16.60	12	8	729	82.6	7.06	39.64
CG10237	FBpp0080748	24.25	12	7	301	35.1	8.51	39.40
smid	FBpp0304233	13.52	13	8	910	100.2	5.27	38.87
Rae1	FBpp0071600	25.72	12	8	346	38.6	7.58	38.78
Ranbp21	FBpp0074528	14.18	12	11	1241	139.1	6.86	38.77
PHGPx	FBpp0072931	47.93	12	7	169	18.7	7.83	38.33
SmD1	FBpp0075684	36.29	11	3	124	13.8	11.65	36.93
Pmm45A	FBpp0087711	18.94	12	10	623	69.9	5.45	36.41
CG6638	FBpp0076375	17.38	9	5	420	46.2	8.18	36.22
Dcp-1	FBpp0071971	25.39	10	6	323	35.9	7.74	36.07
CG4968	FBpp0079577	30.53	10	5	262	30.4	4.98	35.67
Exo84	FBpp0088351	12.80	11	7	742	83.3	6.06	34.61
Nup43	FBpp0082998	26.82	9	8	358	40.1	5.22	34.32
Dbp80	FBpp0112438	19.02	9	7	447	50.1	8.44	34.12
GS	FBpp0290194	13.17	9	6	562	61.8	6.18	33.90
CG4764	FBpp0077625	34.62	11	5	182	20.5	7.01	33.64
CG17746	FBpp0072955	23.72	10	8	371	41.0	4.89	33.40
Kap-alpha3	FBpp0081584	12.26	9	3	514	57.0	5.26	33.34

Nup205	FBpp0074568	3.39	10	5	2067	232.6	6.67	33.06
Bub3	FBpp0084831	20.55	9	6	326	37.4	6.54	32.90
Eb1	FBpp0089088	28.62	9	7	290	32.5	5.34	32.89
CG11241	FBpp0078188	20.28	11	8	508	55.8	8.29	32.68
CG5389	FBpp0075203	7.40	11	3	622	67.7	5.12	32.54
CG9578	FBpp0070007	27.57	10	6	272	32.3	6.57	32.36
CG31472	FBpp0081245	32.49	11	5	237	26.8	6.81	31.88
CG1749	FBpp0073354	28.47	10	6	404	44.1	4.73	31.67
Gale	FBpp0072455	20.29	9	5	350	38.7	7.46	31.45
Argk	FBpp0111703	15.20	9	6	375	42.2	6.32	30.83
CG5757	FBpp0085966	40.76	11	7	211	23.8	7.44	30.78
CG5168	FBpp0079610	21.32	10	5	408	46.2	7.37	30.51
CG3430	FBpp0078978	19.50	9	7	605	68.0	5.78	30.31
tral	FBpp0304882	46.95	53	16	656	69.7	9.48	203.79
l(3)72Ab	FBpp0075282	16.15	43	25	2142	244.4	6.04	152.26
Jabba	FBpp0303041	31.19	37	11	388	43.0	5.39	143.06
sta	FBpp0070279	67.09	35	13	313	35.3	6.55	124.73
Nup358	FBpp0084188	14.25	36	25	2695	296.2	6.09	121.67
RpS15	FBpp0086270	55.10	33	5	147	17.0	10.33	112.78
CG6084	FBpp0075870	51.58	25	13	316	35.9	6.65	91.97
shi	FBpp0089280	22.92	22	15	877	97.7	8.37	77.99
NAT1	FBpp0306828	12.77	21	12	1488	162.8	8.53	72.66
Fmr1	FBpp0081678	21.31	15	7	643	71.7	9.13	66.24
CG4849	FBpp0084626	16.41	18	10	975	110.6	5.03	64.66
Not1	FBpp0111586	8.20	18	13	2220	249.6	6.96	62.09
larp	FBpp0297480	14.41	17	14	1409	151.6	9.57	62.03
eIF-3p66	FBpp0088565	20.18	18	8	560	63.8	7.02	61.92
mod	FBpp0085233	34.69	19	13	542	60.3	5.43	61.20
gammaTub37C	FBpp0080769	23.41	14	7	457	51.3	6.32a	59.19