Transcription Factor Networks in *Drosophila melanogaster*

A dissertation presented

by

David Young Rhee

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Cell and Developmental Biology

Harvard University

Cambridge, Massachusetts

May 2013

Advisor: Spyros Artavanis-Tsakonas David Young Rhee

Transcription Factor Networks in *Drosophila melanogaster*

Abstract

Differential gene expression is an essential component of the programs that give rise to specific cellular fates and functions. This differential regulation occurs primarily at the transcriptional level and is controlled by complex regulatory networks governed by the action of transcription factors at specific DNA regulatory elements. Transcription factors rarely act alone, often functioning through combinatorial interactions with other transcription factors, co-factors and chromatin-remodeling proteins. Defining these protein-protein interactions is an essential component to understanding transcription factor function and consequently, the cell as an integrated network.

The core of this work encompasses a study of *Drosophila melanogaster* transcription factors, defining protein-protein interactions using a co-affinity purification mass spectrometry methodology, representing roughly half of the established catalog of transcription factors. These interactions were subsequently used to probe functional relationships *in vivo*, validating a number of physical interactions in the animal, while also demonstrating predictive value with regard to function, for the protein-protein interaction dataset as a whole.

Using these defined protein interactions, this work explores the biology of transcription factors from the perspective of the protein complex, integrating a variety of data including large-scale expression datasets, transcription factor occupancy studies and inferred gene regulatory networks. These datasets are used to build tissue-specific interaction networks, identifying prospective interactions in a variety of settings

throughout development. Next, shared physical targets of interacting transcription factors are identified, defining likely targets of combinatorial regulation by these interacting factors. Lastly, regulatory network inference models are combined with physical interaction data to define an integrated network, connecting transcription factor protein interactions directly to the gene regulatory network of the cell. This integrated network is subsequently used as a tool to connect the functional network of genetic modifiers related to *mastermind*, a transcriptional co-factor in the Notch signaling pathway. The fundamental goal of this work is to provide a framework from which to build hypotheses and to probe the mechanisms of gene regulation. Given the broad coverage of the data, and the degree of conservation of transcription complexes and regulatory programs, this study lays the foundation for a deeper understanding of transcriptional regulation in *Drosophila* and other metazoans.

Table of Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

List of Tables

Chapter 1:

Chapter 2:

Chapter 3:

Dedicated to my parents:

Bong Soo and Sung Ho Rhee, who have been my biggest supporters.

Acknowledgements

I am deeply grateful for the advice and support of my advisor Spyros Artavanis-Tsakonas, who has been an inspiration to me through his enthusiasm for science and overall generosity. I am not sure what possessed him to take on a struggling fourth year graduate student, but I will always be thankful to him for taking a chance on me and for helping me pursue my dreams. I am also greatly indebted to Davie Van Vactor, who has been a wonderful mentor and committee chair as I navigated my way through the difficulties of graduate school. He has been generous with his time and advice and I am certain I would be in a different place had it not been for his help. I would like to thank John Blenis and Marian Walhout for their valuable input and advice as a part of my dissertation advisory committee. Special thanks to Kate Hodgins, Maria Bollinger and Danny Gonzalez in the BBS program office for their help in all things. They always managed to put a smile on my face, even during the most frustrating parts of my graduate school experience.

I am extremely thankful for the members of the Artavanis-Tsakonas and Van Vactor laboratories, past and present, who I have interacted with on a daily basis for the last few years. In particular, Bob Obar for his mentoring, friendship and for our daily discussions ranging from science, to baseball, to homebrewing. Mark Kankel and Doug Dimlich for our spirited discussions (often irrelevant) and frequent coffee breaks. Jean-Francois Rual for his excellent advice and for taking time to painstakingly read through my manuscript drafts. Pujita Vaidya, Namita Vaidya, Chapman Beekman, and Christina Wong – who suffered through thousands of plasmid preps with me over the last few years. I could not have completed my work without your excellent help. Glenn Doughty for all

the Friday munchkins. My baymates Pallavi Kshetrapal and Kazuya Hori for their friendship, patience with my general messiness and my love of Lil' Wayne. In no particular order, thanks to Diana Ho, Anindya Sen, Jan Manent, Sanja Sale, Joseph Arboleda-Velasquez, Elizabeth McNeill, Jennifer Long, Laura-Anne Lowery, Harsha Kuthethur, Tamar Chobanyan, and Julia Yelick. You have all made my time on the 4th floor of LHRRB that much better, despite the noxious fumes coming from the walls.

I have had the pleasure of working with a number of fantastic collaborators, who not only made my projects possible, but have made the process more enjoyable along the way. Much thanks to Bo Zhai from the Gygi lab for shooting all my mass spec, Matt Slattery and Lijia Ma from Kevin White's group at the University of Chicago for helping me navigate all of their ChIP data, Sue Celniker from Lawrence Berkeley National Laboratory for taking the time to discuss my projects. Deepest thanks to Dong-Yeon Cho and Theresa Przytycka from the NIH, who have made much of the computational analyses in this work possible.

I would also like to thank the people who have taken the time to mentor and train me over the years, I would not be here if it weren't for your patience and understanding. In particular, my college advisors Jim Reynhout and Weldon Jones. My undergraduate research advisor Ross Johnson, who instilled in me a healthy appreciation for gap junctions. Jeff Essner at Iowa State and my NIH IRTA mentor, Cecilia Lo.

Last but not least, I would like to thank my brother Tim, and my parents Bong Soo and Sung Ho Rhee. The last few years have been difficult for us as a family, but we have persevered through all of it together and it is your love that has made all the hard work possible.

Chapter 1

Introduction

"An organism consists essentially of an integrated system of chemical reactions controlled in some manner by genes."

-George Beadle and Edward Tatum 1941

A Requirement for Systems

As one takes a moment to consider the basic necessities for life, four widely examined requirements come to light. First, *the cell as the fundamental unit of life*; second, *the gene as the basis for heredity*; third, that *life needs chemistry to exist*; and last, *evolution through natural selection* (Nurse 2003). A fifth requirement has been recently considered, the need for biological organization or "systems" (Carvunis et al., 2013, Vidal 2009). A theme present since the time of Beadle and Tatum, this fifth component addresses the fact that genes and molecules do not act alone, but rather exist in a dynamic environment involving a vast array of interactions to control the range of biological outcomes. Therefore the cell exists as a dynamic and complex integrated network.

As we have entered an era of systems level analysis, technology has enabled us to probe many components of the cellular network, evidenced by large-scale studies of protein-protein interaction (PPI), gene regulatory network (GRN) and metabolic network studies (Guruharsha et al., 2012, Herrgård et al., 2008, Marbach et al., 2012). Each of these shed light onto different aspects of the global organization of the cell, though much work remains to be done and the challenge remains as how to best combine these various data types into one coherent picture of cellular biology. The ultimate goal is to understand how various stimuli affect this dynamic network to influence the functional

output of the cell and the organism as a whole (i.e. a phenotype). This particular body of work focuses on Transcription Factors (TFs), proteins that directly bind DNA and thus, connect the protein interaction network of a cell to the regulatory network of the genome through protein-DNA interactions (Figure 1.1).

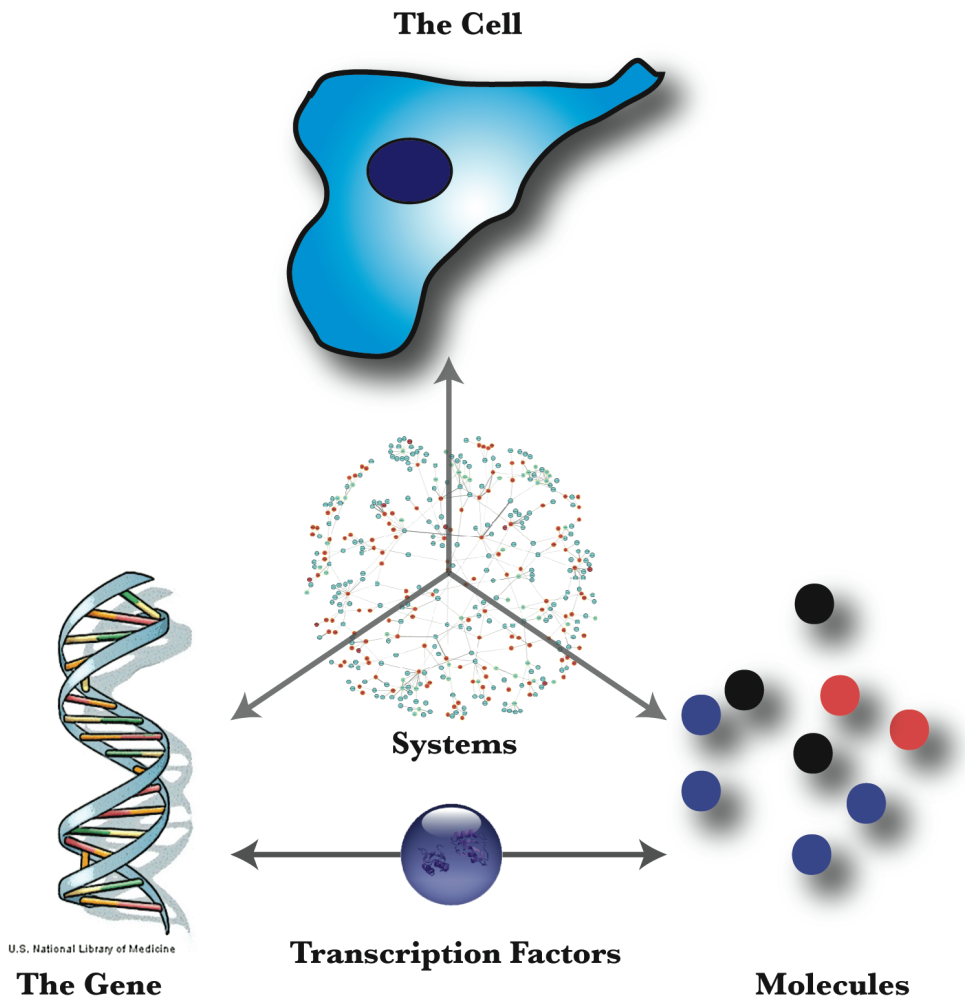

Figure 1.1 Systems Biology: An Essential Component of Life
Systems biology encompasses the organization of molecules and genes within a cell. Transcription Factors exist as a central component of this organization as they connect the interaction network of proteins directly to the genome.

Nucleic Acids and Gene Regulation

Since the initial discovery of deoxyribonucleic acid (DNA) in the 19th century, one of the biggest challenges in Biology and indeed, science as a whole has been to unravel the complexities and functions of DNA. Nucleic acids were first isolated from pus-derived leukocytes and later from sperm cells by Friederich Miescher, a Swiss physician (Dahm 2008). Originally named "nuclein," Miescher was able to demonstrate that this material was localized to the nucleus of cells, contained large amounts of phosphorus and was fundamentally different from any protein known at the time. Despite these early discoveries, the connection between DNA and heredity remained elusive for many years. It was not until the now famous Avery-MacLeod-McCarty experiment in 1944 that DNA was shown to carry genetic information, "transforming" bacteria from a non-virulent to a virulent form (Avery et al., 1944). This was corroborated a decade later in the Hershey-Chase experiments in 1952, where viral DNA was shown to enter bacteria during infection, while protein did not (Hershey and Chase 1952). Shortly thereafter, work by Francis Crick and James Watson established the double helix model of DNA (Watson and Crick 1953). Building on these discoveries, Francis Crick laid out the central dogma of molecular biology in 1958; in short, that DNA makes RNA makes Protein (Crick 1958).

While the transfer of information from nucleic acids to proteins was known for some time, messenger RNA was not uncovered until 1961, when Sydney Brenner, Matthew Meselson and Francois Jacob reported the presence of an "unstable intermediate carrying information from genes to ribosomes" (Brenner et al., 1961). That same year, Jacob and Jacques Monod published their interrogation of the *lac* operon, where the complexities of transcriptional regulation were first elucidated (Jacob and

Monod 1961). This represents the first gene regulatory network, introducing us to now common themes of transcriptional regulation (Figure 1.2).

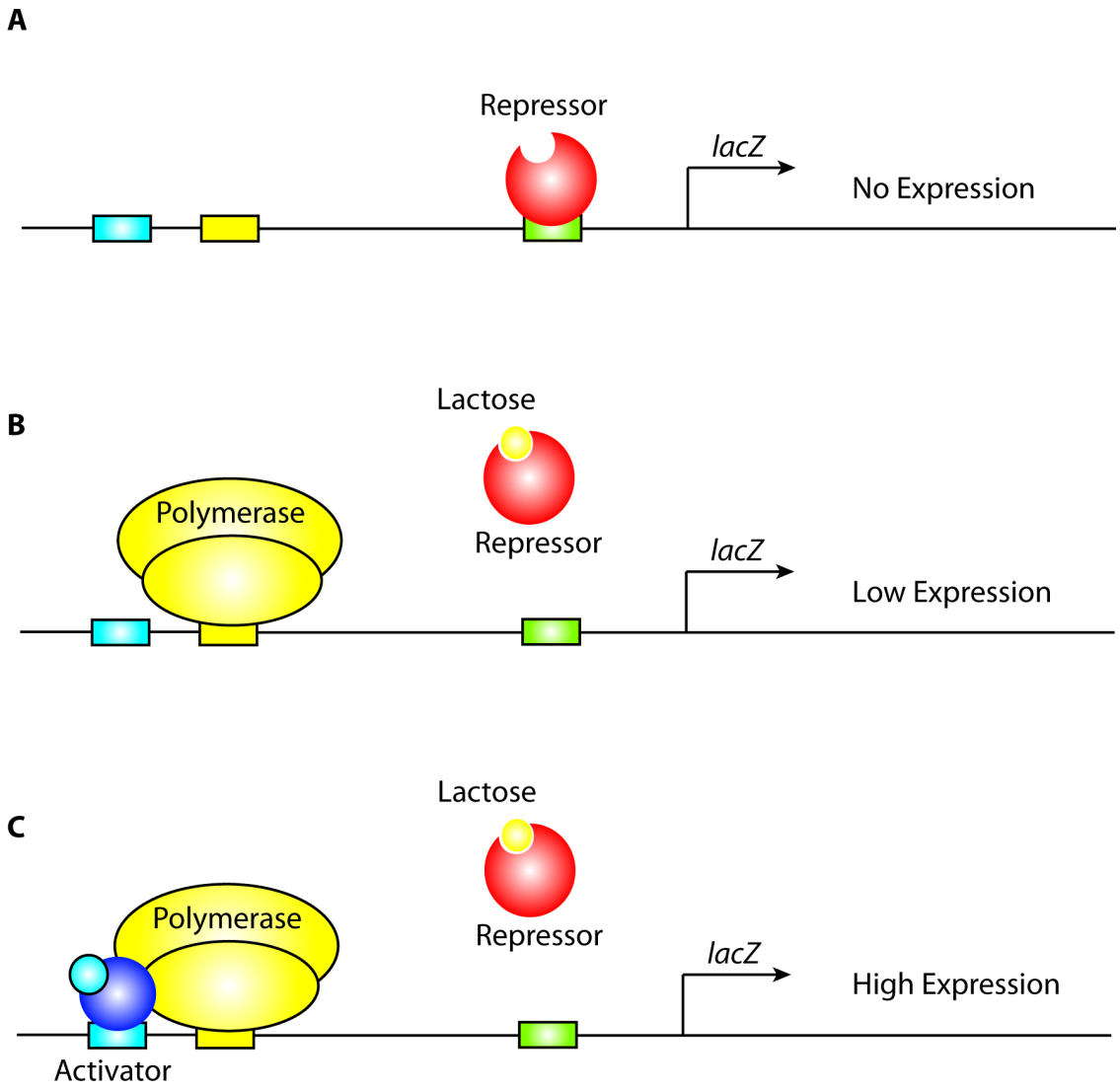

Figure 1.2 Transcriptional Regulation of the *lac* Operon
Simplified view of the transcriptional regulatory mechanisms of the lac operon. Many of the basic themes found in eukaryotic transcription were uncovered from work on the prokaryotic *lac* operon. These include (A) inhibition of transcription by repressor proteins, (B) De-repression and binding of polymerase to promoter elements, and (C) Activation of transcription by an activator protein.

While many discoveries were being made in prokaryotic systems, work in eukaryotes began with the discovery of RNA polymerase activity in the rat liver by Weiss and Gladstone (Weiss and Gladstone, 1959). About a decade later, the three eukaryotic RNA polymerases were identified by Robert Roeder, who went on to pioneer much of the work in eukaryotic transcription including the discovery of transcriptional co-activators (Roeder and Rutter 1969). What became apparent early on was that although the polymerases had been identified, other additional factors were necessary to activate transcription above a basal level, as DNA was bound by repressive factors or was inaccessibly packaged as chromatin. Robert Tjian discovered the first sequence-specific DNA-binding factor, SV40 T antigen, in 1978 (Tjian 1978). Shortly thereafter, Robert Roeder's group identified the first eukaryotic transcription factor, TFIIIA, opening the door to the discovery of hundreds of sequence-specific transcription factors (Engelke et al., 1980, Ginsberg et al., 1984).

Transcription Factors

Transcription factors are defined as proteins that bind specific sequences of DNA and either activate or repress transcription. Their genes comprise between 5-10% of the protein coding capacity of the genome (depending on the species) and are identified by the presence of a DNA-binding domain, falling into several families based on the type of domain (Adryan and Teichmann 2006, Babu et al., 2004). TFs are modular in nature and are typically composed of a DNA binding domain accompanied by an activating domain. They bind at specific enhancer elements, short DNA motifs that are modular in nature and function autonomously in most cases (Spitz and Furlong 2012). The *Drosophila melanogaster* genome encodes fifty different types of DNA binding domains; however, only

14 of these are appear in more than 5 proteins (Table 1.1) (Adryan and Teichmann 2010).

Table 1.1 Common TF Families in *Drosophila melanogaster*
The most commonly found *Drosophila* TF families organized by PFAM domain and the total number of proteins in each family (based on Adryan and Teichmann 2010).

DNA Binding Domain	Number of TFs
Zinc Finger-C2H2	249
Homeobox	99
HLH	55
Zinc Finger-C4	22
BESS	20
Forkhead	19
bZIP_2	11
HTH-psq	9
T-box	9
Myb	8
Ets	8
bZIP_1	7
GATA	6
zinc finger-BED	5

Upon TF binding, transcription is either activated or suppressed through interactions with the general transcription machinery, which directly controls the activity of preinitiation complexes at the promoters of target genes. These interactions may be direct or may be facilitated through the mediator complex or a variety of co-regulatory factors (Figure 1.3) (Roeder 2005). Recent work has also suggested that tissue-specific variants of the basal transcription machinery exist, allowing for further regulation of transcription in specific contexts (D'Alessio et al., 2009).

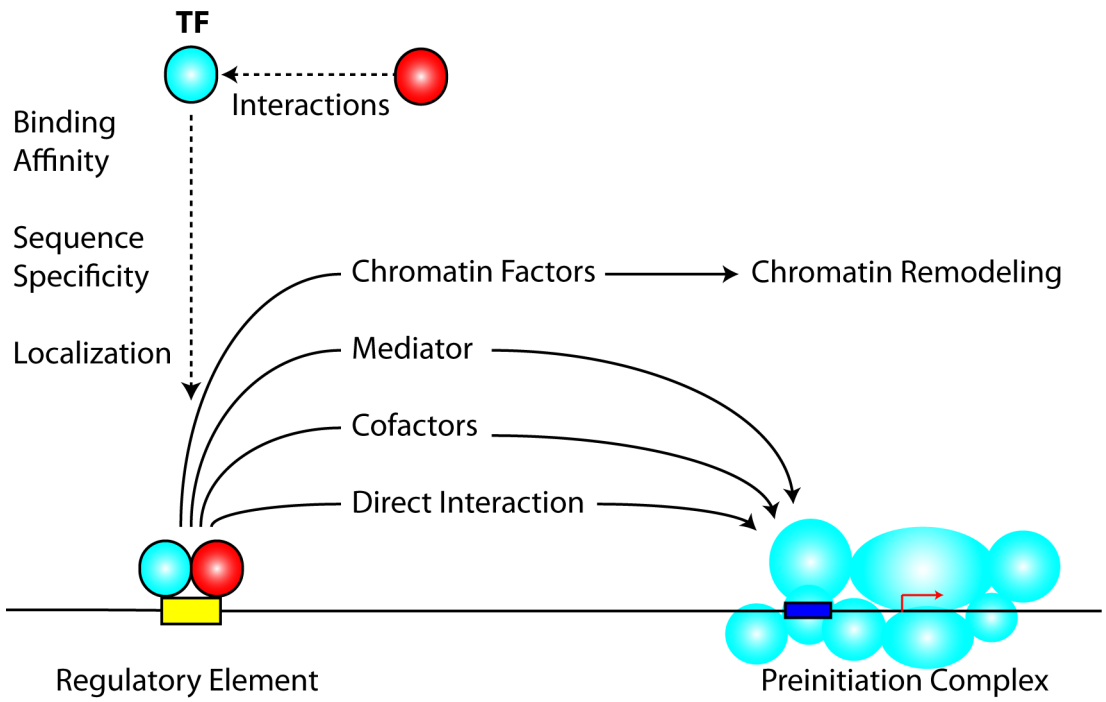

Figure 1.3 Transcription Factor Interactions Regulate TF Function
TF protein interactions play a central role in their function. These interactions mediate the localization of TFs, their binding affinity and binding sequence specificity, connections to chromatin as well as the basal transcriptional machinery.

In addition to the interactions with basal machinery components, TFs interact with a wide range of other proteins at enhancer sites, including other TFs, cofactors and chromatin modifiers (D'Alessio et al., 2009, Grove and Walhout 2008, Naar et al., 2001, Spitz and Furlong 2012). The biological activity of each TF depends on these protein interactions, which ultimately govern TF localization, DNA binding affinity and activation of chromatin remodeling, as well as DNA binding sequence specificity (Siggers et al., 2011, Slattery et al., 2011). This combinatorial nature is further reflected in the fact that TFs tend to bind the genome together, at high occupancy target (HOT) regions, defined as areas where 15 or more independent TFs are present (Gerstein et al., 2010). This collective binding of factors assembles both activators and repressors, allowing for precise control of transcription from a particular locus, as is the case during early

segmentation of the *Drosophila* embryo (Stanojevic et al., 1991). Given the importance of TF protein-protein interactions, defining these interactions is essential to understanding their function.

TFs play a role in a wide range of biological processes, but are most often discussed in the context of development as they are frequently expressed in specific spatiotemporal patterns; and there are clear cases where they act as "master regulators," where a particular TF can specify a distinctive tissue or an organ. One example of this is in muscle development, where MyoD, a basic helix-loop-helix transcription factor, activates transcriptional programs to give rise to muscle cell identity (Choi et al., 1990). Other examples include Pax6 in eye development and tinman in *Drosophila* heart development (Bodmer 1993, Halder et al., 1998). Recent work has also demonstrated that the use of just four TFs: Oct 4, Sox2, KLF3 and Myc can reprogram fibroblasts to become pluripotent stem cells (Takahashi and Yamanaka 2006).

Interestingly, the majority of TFs remain expressed in the adult, suggesting they are important for processes beyond their developmental roles. TFs are frequently the target of signaling pathways, are involved in the control of the cell cycle and can be induced by environmental signals, as is the case for example with the heat shock response (Lindquist 1986, Medema et al., 2000). They also play a significant role in disease, including disorders involving hormone response and in cancer (Latchman 1996). While much remains to be discovered with regard to TF biology, what is clear is that TF protein interactions are essential to understanding their functions and that TFs play a pivotal role in the life of the cell, through many mechanisms.

Protein Interactomes

As the majority of proteins (not just TFs) rely on interactions with other proteins, large-scale protein interaction networks have provided a valuable resource for predicting and understanding biological function. These networks are composed of nodes and edges, representing proteins and interactions, respectively. Most of these studies have relied on yeast two-hybrid methods (Y2H) to generate large datasets of binary protein-protein interactions in a variety of different systems (Giot et al., 2003, Ito et al., 2001, Li et al., 2004, Rual et al., 2005, Stanyon et al., 2004, Stelzl et al., 2005, Uetz et al., 2000). As the Y2H system uses a transcriptional read out, TFs pose a distinct problem, as many TFs are capable of activating transcription on their own. As a result, TFs are often underrepresented in Y2H-based interactome studies. Nonetheless, the system has been used previously to define binary TF-TF interactions via a matrix approach in *C. elegans* and in mammalian species (Grove et al., 2009, Ravasi et al., 2010). These reports have contributed a number of novel connections between TFs, but still only represent a small portion of the entire TF interactome as a whole, and by experimental design, only examine each TF pair in isolation and cannot take into account the large repertoire of protein interactions between TFs and other non-TF proteins.

An alternative to Y2H-based networks is the use of co-affinity purification followed by tandem mass spectrometry (Co-AP/MS). In this method, a protein is pulled-down and interacting proteins are subsequently identified through MS/MS analysis. By combining many individual pull-down experiments, one is able to construct large-scale protein interaction networks. This method has been used successfully to build proteome wide interaction networks (Gavin et al., 2006, Ho et al., 2002, Krogan et al., 2006), though only one study to date has examined a metazoan species at a large scale (Figure

1.4) (Appendix B, Guruharsha et al., 2011). The benefits of this approach include the examination of proteins at the level of the complex and in the case of TFs, the added advantage of avoiding a transcriptional readout for interaction.

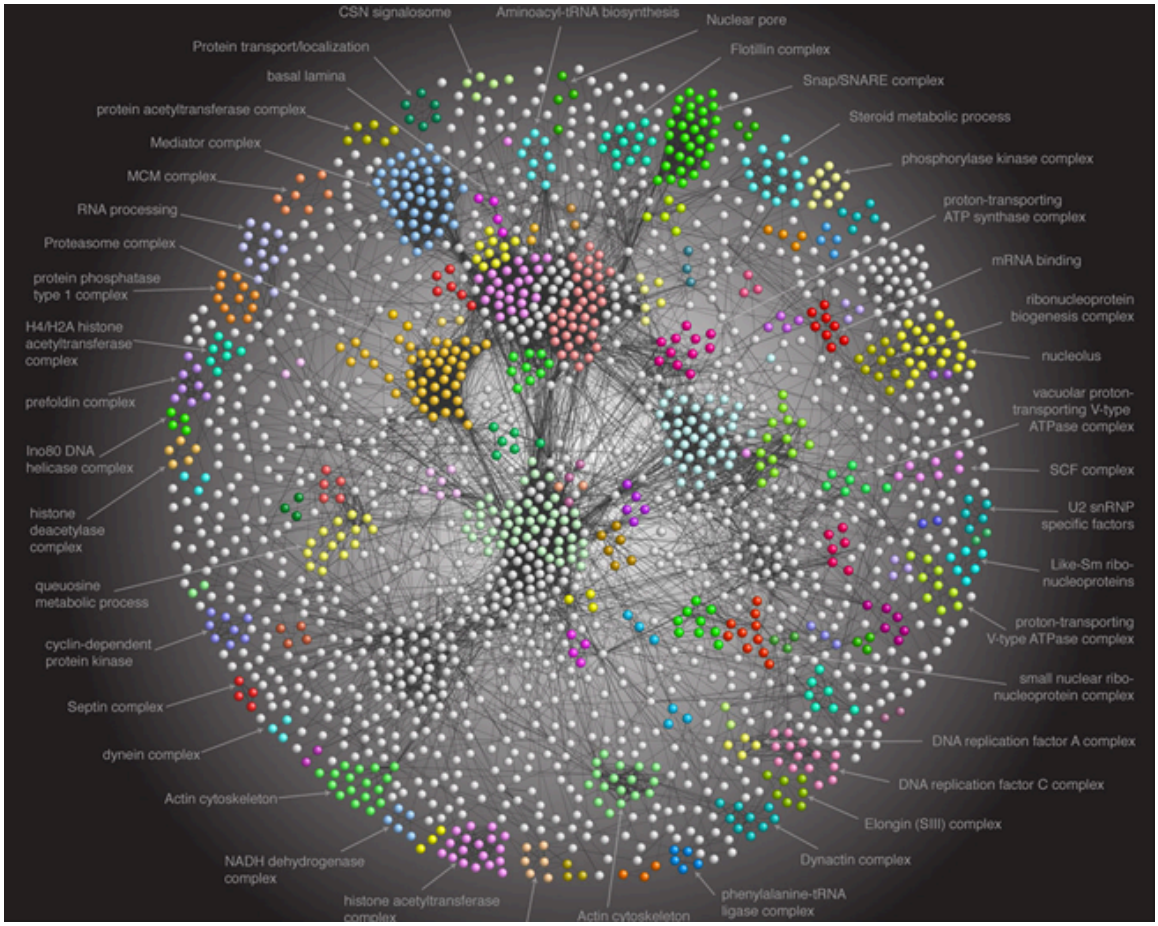

Figure 1.4 A *Drosophila* Protein Interaction Network Map
A co-AP/MS based protein interaction network encompassing ~2,300 proteins connected by ~11,000 connections (Appendix B, Guruharsha et al., 2011).

To date, no TF-specific protein interactome study has been published for *Drosophila*, nor has the co-AP/MS approach been used to specifically examine these relationships in any metazoan species at a large scale. Existing interaction data in *Drosophila* covers only a small proportion of known TFs, thus the majority of TF protein interactions have yet to be defined. This body of work uses a co-AP/MS approach to

systematically define these interactions, identifying interactions, in many cases novel ones, for nearly half of the TFs in *Drosophila*.

Gene Regulatory Networks (GRNs)

While protein interaction networks capture the physical relationships between the proteins in a cell, GRNs capture the connections between TFs and the DNA elements needed to regulate gene expression. These networks consist of two types of nodes, TFs and their binding sites. These nodes are connected through two types of edges, physical and regulatory, where physical edges are defined by the binding of TFs to DNA and regulatory edges capture the activation or suppression of a gene product, based on the overexpression or loss of a particular TF (Figure 1.5) (Capaldi et al., 2008). Not only are TFs the central focus of these networks, they represent a crucial interface between the protein interactome and the regulatory network of the cell, thus providing a link between two distinct spaces within the cellular network.

GRNs are often constructed from TF occupancy studies, using methods such as Chromatin Immunoprecipitation-sequencing (ChIP-seq), though recent work has suggested that direct physical binding of a factor does not always correspond to a functional output (Spitz and Furlong 2012). It would seem, given the combinatorial nature of TFs, that methods that examine these factors in isolation require more information to understand the complex regulatory mechanisms of differential gene expression. It is easy to imagine a scenario where many of these factors are simply bound to a repressive co-factor, thus inhibiting transcription. As such, integrating the protein interactome of TFs into regulatory network models will provide deeper insight into the function of these proteins.

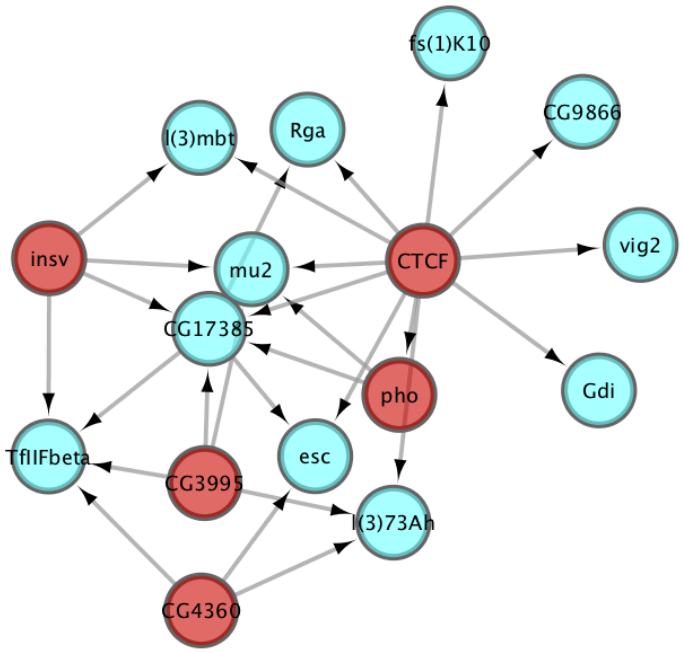

Figure 1.5 Gene Regulatory Networks
An example regulatory network from our integrated network analysis. TFs (red nodes) represent the central components of the network. Directional edges indicate transfer of information, in this case, regulatory relationships between TFs and their target proteins (blue nodes).

Alternative approaches to defining regulatory interactions have relied on computational methods, where a set of high-quality regulatory interactions is used to train datasets using machine learning approaches (Marbach et al., 2012, Roy et al., 2010). These models allow for the incorporation of many different data types, including large-scale expression studies, TF occupancy and chromatin marks. Although the majority of the regulatory edges inferred from these methods will need to be validated experimentally, these approaches provide a powerful tool for developing hypotheses and allow for the incorporation of many of the large-scale datasets currently available. We utilize such regulatory networks to connect our TF protein interaction network to the regulatory network of the cell, thus defining regulatory relationships from the perspective of the protein complex.

Notch Signaling

Notch signalling represents one of only a few fundamentally conserved metazoan-signalling mechanisms in development (Artavanis-Tsakonas et al., 1999). The Notch pathway was originally discovered through the occurrence of a spontaneous mutant in *Drosophila*, identified by a serrated wing phenotype (Morgan and Bridges 1916, Mohr 1919). Early work established the pleiotropic nature of the pathway, revealing embryonic "neurogenic" phenotypes, where neural tissue formed at the expense of epidermis, as well as roles in the development of the wing margin and bristles (Poulson 1937). The Notch locus was cloned in the early 1980's, revealing a transmembrane receptor containing EGF-like repeats in the extracellular domain (Artavanis-Tsakonas et al., 1983, Kidd et al., 1983, Wharton et al., 1985).

The fundamental pathway consists of the Notch receptor and membrane bound Delta-Serrate-LAG2 (DSL) ligands (Jagged in mammals). The interaction between the Notch receptor on one cell and its ligand on an adjacent cell triggers two proteolytic events, first involving ADAM-family metalloproteases and a second, mediated by gamma-secretase, which releases the Notch intracellular domain (NICD) from the membrane (Bray 2006). The NICD enters the nucleus and interacts with CSL [CBF1, Su(H) and LAG-1], a transcription factor, and *mastermind* (*mam*), a transcription co-factor, to activate transcription (Figure 1.6).

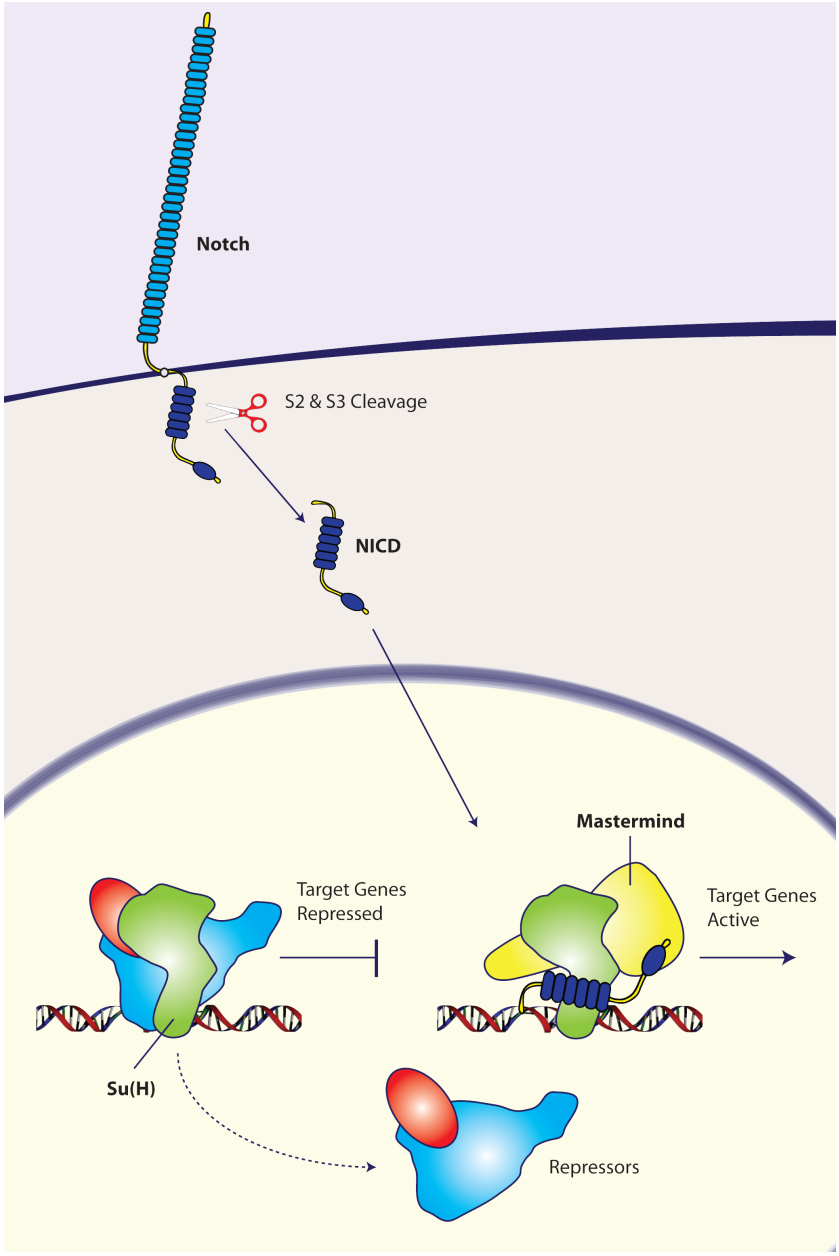

Figure 1.6 The Notch Signaling Pathway
Simplified view of the Notch signaling pathway. When Notch is activated, it undergoes to cleavage steps to release the Notch Intracellular Domain (NICD) from the cell membrane. The NICD enters the nucleus, where it interacts with Suppressor of Hairless (Su(H)), creating a binding interface for the transcriptional co-activator, *mastermind*.

The pathway is inherently simple in design in that the receptor itself includes a transcriptional activator, presumably circumventing the need for second messengers and

amplification steps. This straightforward design belies the underlying complexity, revealed largely through genetic screens performed in *Drosophila* and in *C. elegans* (Xu and Artavanis-Tsakonas 1990, Fortini and Artavanis-Tsakonas 1994, Go and Artavanis-Tsakonas 1998, Kankel et al., 2007, Shalaby et al. 2009). These various screens have uncovered hundreds of genes that functionally interact with Notch pathway components, though different screens show little overlap (Reviewed in Guruharsha et al., 2012). What is unclear is why there is such little overlap between studies and how these numerous genes are connected to the Notch signalling network at a mechanistic level.

Recent large-scale protein-protein interaction studies (Guruharsha et al., 2011), expression profiling and transcription factor occupancy studies (Celniker et al., 2009, Roy et al., 2010) have provided a framework from which to probe the connections between these Notch-connected proteins. While some functionally interacting proteins do interact with one another through direct physical edges (either protein-protein or protein-DNA), the majority do not. Capturing both physical and regulatory edges in a network allows us to explore the space among these functionally related proteins, and will hopefully provide insight into understanding the pleiotropic nature of the pathway. We use such an approach, building an integrated PPI-regulatory network to interrogate genetic modifiers related to Notch signaling.

Project Goals

The vast majority of TF protein interactions in metazoans have not yet been defined. Given that these interactions play a central role in TF function, we have constructed a large-scale TF protein interaction network, identifying novel interactions for a significant proportion of TFs in *Drosophila*. We use these physical interactions to

predict functional relationships and validate some of these *in vivo*, demonstrating predictive value for our PPI network.

As TFs are frequently discussed in the context of tissue specificity and development, we have integrated our interaction network with expression datasets to define both tissue-specific and broadly expressed proteins, identifying interactions that are likely to exist in specific tissues or across many different tissue types. We also define shared physical targets of interacting pairs of TFs, identifying examples where combinatorial regulation is likely to occur.

As TFs represent the connection between the protein interactome and GRNs of a cell, we integrate our protein interaction network with inferred regulatory network models to define transcription regulatory motifs, such as feedback loops between interacting proteins, and to build an integrated network, identifying shared targets of proteins that exist as a protein complex. This integrated network allows us to bridge the gap between physical interactions and functional genetic datasets, which we demonstrate by connecting known genetic modifiers of *mastermind*, a Notch transcription co-activator.

These data are intended to be used as hypothesis-generating tools, where a single protein, protein complex or a network of genetic interactions can be used as a starting point to probe biological mechanisms.

Bibliography

Adryan, B., and Teichmann, S.A. (2006). FlyTF: a systematic review of site-specific transcription factors in the fruit fly Drosophila melanogaster. Bioinformatics 22, 1532-1533.

Adryan, B., and Teichmann, S.A. (2010). The developmental expression dynamics of Drosophila melanogaster transcription factors. Genome Biol. 11(4):R40

Artavanis-Tsakonas, S., Muskavitch, M.A., and Yedvobnick, B. (1983). Molecular cloning of Notch, a locus affecting neurogenesis in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America 80, 1977-1981.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: cell fate control and signal integration in development. Science 284, 770-776.

Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. The Journal of experimental medicine 79, 137-158.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. Current opinion in structural biology 14, 283-291.

Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in Drosophila. Development 118, 719-729.

Bray, S. (2006) Notch Signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol. 7(9):678-89

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 190, 576-581.

Capaldi, A.P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nature genetics 40, 1300-1306.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., et al. (2009). Unlocking the secrets of the genome. Nature 459, 927-930.

Choi, J., Costa, M.L., Mermelstein, C.S., Chagas, C., Holtzer, S., and Holtzer, H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. Proceedings of the National Academy of Sciences of the United States of America 87, 7988-7992.

Crick, F.H. (1958). On protein synthesis. Symposia of the Society for Experimental Biology 12, 138-163.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. Molecular cell 36, 924-931.

Dahm, R. (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Human genetics 122, 565-581.

Engelke, D.R., Ng, S.Y., Shastry, B.S., and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell 19, 717-728.

Fortini, M.E., and Artavanis-Tsakonas, S. (1994). The suppressor of hairless protein participates in notch receptor signaling. Cell 79, 273-282.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631-636.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330, 1775-1787.

Ginsberg, A.M., King, B.O., and Roeder, R.G. (1984). Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. Cell 39, 479-489.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. (2003). A protein interaction map of Drosophila melanogaster. Science 302, 1727-1736.

Go, M.J., and Artavanis-Tsakonas, S. (1998). A genetic screen for novel components of the notch signaling pathway during Drosophila bristle development. Genetics 150, 211-220.

Grove, C.A., De Masi, F., Barrasa, M.I., Newburger, D.E., Alkema, M.J., Bulyk, M.L., and Walhout, A.J. (2009). A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell 138, 314-327.

Grove, C.A., and Walhout, A.J. (2008). Transcription factor functionality and transcription regulatory networks. Molecular bioSystems 4, 309-314.

Guruharsha, K.G., Rual, J.F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O., et al. (2011). A protein complex network of Drosophila melanogaster. Cell 147, 690-703.

Halder, G., Callaerts, P., Flister, S., Walldorf, U., Kloter, U., and Gehring, W.J. (1998). Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. Development 125, 2181-2191.

Herrgard, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Bluthgen, N., Borger, S., Costenoble, R., Heinemann, M., et al. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nature biotechnology 26, 1155-1160.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. The Journal of general physiology 36, 39-56.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415, 180-183.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (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, 4569-4574.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology 3, 318-356.

Kankel, M.W., Hurlbut, G.D., Upadhyay, G., Yajnik, V., Yedvobnick, B., and Artavanis-Tsakonas, S. (2007). Investigating the genetic circuitry of mastermind in Drosophila, a notch signal effector. Genetics 177, 2493-2505.

Kidd, S., Lockett, T.J., and Young, M.W. (1983). The Notch locus of Drosophila melanogaster. Cell 34, 421-433.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440, 637-643.

Latchman, D.S. (1996). Transcription-factor mutations and disease. The New England journal of medicine 334, 28-33.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. (2004). A map of the interactome network of the metazoan C. elegans. Science 303, 540-543.

Lindquist, S. (1986). The heat-shock response. Annual review of biochemistry 55, 1151-1191.

Marbach, D., Roy, S., Ay, F., Meyer, P.E., Candeias, R., Kahveci, T., Bristow, C.A., and Kellis, M. (2012). Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks. Genome research 22, 1334-1349.

Medema, R.H., Kops, G.J., Bos, J.L., and Burgering, B.M. (2000). AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27kip1. Nature 404, 782-787.

Mohr, O.L. (1919). Character Changes Caused by Mutation of an Entire Region of a Chromosome in Drosophila. Genetics 4, 275-282.

Morgan, T. H., and Bridges, C. B. (1919). THE INHERITANCE OF A FLUCTUATING CHARACTER. J Gen Physiol. 20;1(6):639-43

Naar, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annual review of biochemistry 70, 475-501.

Nurse, P. (2003). The great ideas of biology. Clin Med 3, 560-568.

Poulson, D.F. (1937). Chromosomal Deficiencies and the Embryonic Development of Drosophila Melanogaster. Proceedings of the National Academy of Sciences of the United States of America 23, 133-137.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140, 744-752.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS letters 579, 909-915.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature 224, 234-237.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330, 1787-1797.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature 437, 1173-1178.

Shalaby, N.A., Parks, A.L., Morreale, E.J., Osswalt, M.C., Pfau, K.M., Pierce, E.L., and Muskavitch, M.A. (2009). A screen for modifiers of notch signaling uncovers Amun, a protein with a critical role in sensory organ development. Genetics 182, 1061-1076.

Siggers, T., Duyzend, M.H., Reddy, J., Khan, S., and Bulyk, M.L. (2011). Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. Molecular systems biology 7, 555.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., et al. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell 147, 1270-1282.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics 13, 613-626.

Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science 254, 1385-1387.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell 122, 957-968.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-676.

Tjian, R. (1978). The binding site on SV40 DNA for a T antigen-related protein. Cell 13, 165-179.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403, 623-627.

Vidal, M. (2009). A unifying view of 21st century systems biology. FEBS letters 583, 3891-3894.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737-738.

Weiss, S., and Gladstone, L. A. (1959). A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J. Am. Chem. Soc. 81, 4118-4119

Wharton, K.A., Johansen, K.M., Xu, T., and Artavanis-Tsakonas, S. (1985). Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats. Cell 43, 567-581.

Xu, T., and Artavanis-Tsakonas, S. (1990). deltex, a locus interacting with the neurogenic genes, Notch, Delta and mastermind in Drosophila melanogaster. Genetics 126, 665-677.

Chapter 2

A Protein Interaction Network of *Drosophila* Transcription Factors

Attributions

Portions of this chapter were published as:

D.Y. Rhee, B. Zhai, C. Wong, C. Beekman, S. Gygi, R. Obar, S. Artavanis-Tsakonas. (2010). The *Drosophila* Transcription Factor Protein Interactome. Mol Biol Cell 22, 4705 (abstract #2043).

I carried out all experiments with the following exceptions: Bo Zhai[1] maintained and operated the mass spectrometer. Julian Mintseris[1] developed and executed the HGScore algorithm on my data. Christina Wong[1] and Chapman Beekman[1] aided in plasmid preps and transfections. Bob Obar[1] helped design experiments and in troubleshooting. The FLAG-HA tagged clone collection was generated in Sue Celniker's group[2].

1. Harvard Medical School, Department of Cell Biology, Boston, MA
2. Berkeley *Drosophila* Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

Summary

We established a co-AP/MS pipeline to specifically isolate and analyze TF protein complexes. To that end, we transfected, expressed and purified 499 unique tagged proteins, including TFs, putative TFs, transcription machinery components and chromatin remodeling proteins. We recovered interactions for 327 out of an estimated 711 TFs in *Drosophila*. These data were analyzed using the HyperGeometric Spectral Counts Score algorithm (HGScore) to construct a high-confidence protein interaction network containing 624 connections between 647 unique proteins, of which 229 are TFs. We compared this network to The Comprehensive *Drosophila* Interactions Database (DroID), demonstrating the recovery of a number of known protein complexes and revealing that the vast majority of interactions defined in our networks are novel. As proteins that interact often share function, we used our interaction data to identify proteins that directly interact with known functional modifiers of the Notch pathway. As a number of these relationships had not been tested before, we examined them *in vivo*, recovering functional genetic interactions for a significant fraction of our protein-protein interaction (PPI)-based predictions. This analysis demonstrates the utility of our PPI data for making functional predictions and validated a number of our physical interactions in the animal.

Introduction

From the earliest embryo to the adult, the spatiotemporal expression of genes is essential for normal development and physiology. At the very basis of this is the regulation of transcription via transcription factors, proteins that physically bind DNA to activate or suppress gene expression. As the target of signaling pathways and the first step in synthesis of proteins and regulatory RNAs, transcription factors represent a crucial point of regulation relating to the vast majority of cellular processes. The majority of TFs function through interactions with other proteins. Consequently, the characterization of these protein-protein interactions is essential for understanding how TFs function to regulate gene expression and in turn, the biology of the cell.

As we examined available resources for protein-protein interactions in *Drosophila*, it became apparent that in the majority of studies, TFs are either underrepresented or are completely absent from published interactomes. For instance, the recently published *Drosophila* Protein Interaction Map, representing the largest metazoan protein complex map to date, contains only 82 of an estimated ~700 *Drosophila* TFs, despite comprising 10,969 connections between 2,297 unique proteins (Appendix B, Guruharsha et al., 2011). Other studies have either focused specifically on non-TF proteins or utilized two-hybrid screening strategies, which are traditionally underrepresented for TFs due to the dependency on transcription as a read out for protein interaction (Formstecher et al., 2005, Friedman et al., 2011, Yu et al., 2011).

Alternative approaches to TF interactome construction have included interaction predictions based on co-expression (Adryan and Teichmann 2010, Suzuki et al., 2009) or on the combining of multiple TF occupancy studies (Cole et al., 2008, Lee et al., 2006. Mathur et al., 2008, Roy et al., 2010). In each case, direct interactions must still be

confirmed through additional experimental means. Given the combinatorial nature of TFs and the absence of general rules for their incorporation into protein complexes, systematically defining their interactions would represent a substantial leap forward in our understanding of gene regulation in the cell.

Toward this goal, we interrogated the protein interaction network of *Drosophila* TFs using a co-affinity purification/mass spectrometry (co-AP/MS) platform. This represents the first co-AP/MS based protein interaction network in any metazoan species to specifically address TF protein interactions. Although we recover a number of previously characterized TF protein interactions, the vast majority of edges in our network are novel, representing new avenues for investigation. We use this PPI framework to predict and validate proteins that function *in vivo* as a part of the Notch signalling network. Ultimately these data represent a resource for the community as a whole, providing a considerable framework for the exploration of the mechanisms of gene regulation.

Results

Literature Search to Establish a List of Drosophila Transcription Factors

As a first step, we sought to define a list of TFs and TF-related proteins, with which to begin our interactome study. A defining feature of TFs is their ability to bind DNA through the presence of a DNA binding domain. Based on this property, approaches to identifying TFs have included simple BLAST searches for DNA binding domains, or cross-species comparisons with previously characterized TFs and the DNA binding domains they contain (Adryan and Teichmann 2007). These methods have led to the availability of several *Drosophila* TF databases, most notably FlyTF.org, which

combines functional annotations from FlyBase, as well as domain-based TF predictions to identify putative TFs (Adryan and Teichmann 2006).

We built on several, partially overlapping TF prediction datasets: first, a list of 754 putative TFs from FlyTF.org, a list of 749 factors generated by Bart DePlacnke's group (Adryan and Teichmann 2006, Pfreundt et al., 2010), 433 factors from the Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/EST/TFweblist433.html) and a manually curated list of 711 factors from Susan Celniker's group (Celniker and Hammond, personal communication). The TFs on the list from Dr. Celniker's group are hereafter referred to as "characterized TFs," as they contain proteins that are definitively TFs, excluding computational predictions as well as other TF-related proteins (such as cofactors that do not bind to DNA). We combined all of these resources to generate a master list of 996 unique TF genes (Table 2.1).

Table 2.1 Transcription Factors in *Drosophila melanogaster*
A list of 996 putative TF genes, based manual curation and currently available motif based prediction models.

Table 2.1 TF Master List					
Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0264442	ab	FBgn0025463	Bap60	FBgn0043364	cbt
FBgn0000014	abd-A	FBgn0024251	bbx	FBgn0265574	Cdc5
FBgn0000015	Abd-B	FBgn0000166	bcd	FBgn0015618	Cdk8
FBgn0000018	abo	FBgn0015602	BEAF-32	FBgn0086697	Cenp-C
FBgn0000022	ac	FBgn0013753	Bgb	FBgn0000286	Cf2
FBgn0027620	Acf1	FBgn0000179	bi	FBgn0000289	cg
FBgn0033749	achi	FBgn0039509	bigmax	FBgn0035702	CG10147
FBgn0000028	acj6	FBgn0045759	bin	FBgn0033971	CG10209
FBgn0263738	Ada2a	FBgn0010520	Bka	FBgn0027514	CG1024
FBgn0037555	Ada2b	FBgn0035625	Blimp-1	FBgn0035690	CG10274
FBgn0000054	Adf1	FBgn0050077	blos1	FBgn0034643	CG10321
FBgn0005694	Aef1	FBgn0023097	bon	FBgn0034729	CG10344
FBgn0000061	al	FBgn0004893	bowl	FBgn0032707	CG10348
FBgn0261238	Alh	FBgn0000210	br	FBgn0032814	CG10366
FBgn0003270	amos	FBgn0033155	Br140	FBgn0032730	CG10431
FBgn0260642	Antp	FBgn0010300	brat	FBgn0032815	CG10462
FBgn0000097	aop	FBgn0039654	Brd8	FBgn0034570	CG10543
FBgn0000099	ap	FBgn0038499	Brf	FBgn0037051	CG10565
FBgn0015903	apt	FBgn0024250	brk	FBgn0032817	CG10631
FBgn0015904	ara	FBgn0013755	Bro	FBgn0036294	CG10654
FBgn0000137	ase	FBgn0259246	brp	FBgn0039329	CG10669
FBgn0005386	ash1	FBgn0004101	bs	FBgn0034945	CG10904
FBgn0000139	ash2	FBgn0000227	Bsg25A	FBgn0032858	CG10949
FBgn0031876	Atac1	FBgn0000529	bsh	FBgn0030010	CG10959
FBgn0032691	Atac2	FBgn0263108	BtbVII	FBgn0037379	CG10979
FBgn0039946	ATbp	FBgn0000233	btd	FBgn0030408	CG11085
FBgn0265193	Atf-2	FBgn0025679	Bteb2	FBgn0030266	CG11122
FBgn0028550	Atf3	FBgn0014949	btn	FBgn0037120	CG11247
FBgn0033010	Atf6	FBgn0259176	bun	FBgn0030058	CG11294
FBgn0010433	ato	FBgn0265598	Bx	FBgn0039816	CG11317
FBgn0019637	Atu	FBgn0011723	byn	FBgn0040366	CG11398
FBgn0013751	Awh	FBgn0004863	C15	FBgn0035024	CG11414
FBgn0025185	az2	FBgn0000250	cact	FBgn0037031	CG11456
FBgn0011758	B-H1	FBgn0000251	cad	FBgn0039733	CG11504
FBgn0004854	B-H2	FBgn0263979	Caf1	FBgn0036249	CG11560
FBgn0004870	bab1	FBgn0259234	Camta	FBgn0031232	CG11617
FBgn0025525	bab2	FBgn0004878	cas	FBgn0030316	CG11695
FBgn0004862	bap	FBgn0024249	cato	FBgn0030314	CG11696
FBgn0042085	Bap170	FBgn0015919	caup	FBgn0031115	CG11710
FBgn0025716	Bap55	FBgn0011571	caz	FBgn0031391	CG11723

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0037618	CG11762	FBgn0031037	CG14207	FBgn0038741	CG17186
FBgn0028647	CG11902	FBgn0039504	CG14260	FBgn0039369	CG17195
FBgn0034425	CG11906	FBgn0029899	CG14438	FBgn0039367	CG17197
FBgn0037645	CG11966	FBgn0029895	CG14441	FBgn0039366	CG17198
FBgn0027503	CG11970	FBgn0037183	CG14451	FBgn0034202	CG17287
FBgn0037655	CG11984	FBgn0037275	CG14655	FBgn0028895	CG17328
FBgn0039831	CG12054	FBgn0037317	CG14667	FBgn0036396	CG17359
FBgn0039808	CG12071	FBgn0037920	CG14710	FBgn0036395	CG17361
FBgn0030065	CG12075	FBgn0037922	CG14711	FBgn0033934	CG17385
FBgn0035238	CG12104	FBgn0037924	CG14712	FBgn0032763	CG17568
FBgn0029957	CG12155	FBgn0040777	CG14767	FBgn0031597	CG17612
FBgn0037329	CG12162	FBgn0038273	CG14860	FBgn0038550	CG17801
FBgn0043796	CG12219	FBgn0035407	CG14962	FBgn0038549	CG17802
FBgn0029822	CG12236	FBgn0035414	CG14965	FBgn0038547	CG17803
FBgn0038057	CG12267	FBgn0035518	CG15011	FBgn0038548	CG17806
FBgn0032295	CG12299	FBgn0034379	CG15073	FBgn0025635	CG17829
FBgn0035137	CG1233	FBgn0028878	CG15269	FBgn0032600	CG17912
FBgn0033581	CG12391	FBgn0028531	CG15286	FBgn0039860	CG1792
FBgn0035481	CG12605	FBgn0031144	CG1529	FBgn0033491	CG18011
FBgn0030952	CG12609	FBgn0030009	CG15336	FBgn0030012	CG18262
FBgn0040929	CG12659	FBgn0030077	CG15365	FBgn0036725	CG18265
FBgn0033459	CG12744	FBgn0031608	CG15435	FBgn0032979	CG1832
FBgn0037206	CG12768	FBgn0031610	CG15436	FBgn0033458	CG18446
FBgn0033252	CG12769	FBgn0032493	CG15479	FBgn0037931	CG18476
FBgn0033569	CG12942	FBgn0039712	CG15514	FBgn0038592	CG18599
FBgn0032150	CG13123	FBgn0030673	CG15601	FBgn0032202	CG18619
FBgn0033667	CG13183	FBgn0038833	CG15696	FBgn0042205	CG18764
FBgn0033627	CG13204	FBgn0034120	CG15710	FBgn0037466	CG1965
FBgn0035643	CG13287	FBgn0036538	CG15715	FBgn0030003	CG2116
FBgn0035687	CG13296	FBgn0033186	CG1602	FBgn0030005	CG2120
FBgn0036479	CG13458	FBgn0033185	CG1603	FBgn0030008	CG2129
FBgn0039201	CG13617	FBgn0033183	CG1620	FBgn0035213	CG2199
FBgn0039209	CG13624	FBgn0039602	CG1647	FBgn0030240	CG2202
FBgn0031874	CG13775	FBgn0033449	CG1663	FBgn0032871	CG2611
FBgn0035157	CG13894	FBgn0003715	CG16778	FBgn0025838	CG2652
FBgn0035160	CG13897	FBgn0037698	CG16779	FBgn0014931	CG2678
FBgn0031718	CG14014	FBgn0032490	CG16813	FBgn0024975	CG2712
FBgn0040392	CG14050	FBgn0032491	CG16815	FBgn0027599	CG2790
FBgn0033474	CG1407	FBgn0028931	CG16863	FBgn0029672	CG2875
FBgn0036331	CG14117	FBgn0032291	CG17118	FBgn0030206	CG2889
FBgn0031023	CG14200	FBgn0035144	CG17181	FBgn0050020	CG30020

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0029928	CG3032	FBgn0085405	CG34376	FBgn0037876	CG4820
FBgn0050401	CG30401	FBgn0085451	CG34422	FBgn0038766	CG4854
FBgn0050403	CG30403	FBgn0028887	CG3491	FBgn0038768	CG4936
FBgn0050431	CG30431	FBgn0040355	CG3526	FBgn0039370	CG4956
FBgn0034946	CG3065	FBgn0029824	CG3726	FBgn0028744	CG5033
FBgn0051224	CG31224	FBgn0029861	CG3815	FBgn0043457	CG5180
FBgn0051365	CG31365	FBgn0032130	CG3838	FBgn0038039	CG5196
FBgn0051388	CG31388	FBgn0029867	CG3847	FBgn0032473	CG5204
FBgn0051441	CG31441	FBgn0027524	CG3909	FBgn0038047	CG5245
FBgn0051460	CG31460	FBgn0036423	CG3919	FBgn0032248	CG5343
FBgn0051510	CG31510	FBgn0038472	CG3995	FBgn0038046	CG5641
FBgn0051612	CG31612	FBgn0031257	CG4133	FBgn0032587	CG5953
FBgn0051627	CG31627	FBgn0260390	CG42516	FBgn0038339	CG6118
FBgn0034961	CG3163	FBgn0260953	CG42585	FBgn0036155	CG6163
FBgn0051642	CG31642	FBgn0261641	CG42724	FBgn0036152	CG6175
FBgn0051835	CG31835	FBgn0261680	CG42727	FBgn0037794	CG6254
FBgn0051875	CG31875	FBgn0261705	CG42741	FBgn0036126	CG6272
FBgn0051955	CG31955	FBgn0261802	CG42748	FBgn0038316	CG6276
FBgn0052006	CG32006	FBgn0034114	CG4282	FBgn0030933	CG6470
FBgn0052105	CG32105	FBgn0030455	CG4318	FBgn0038301	CG6654
FBgn0052121	CG32121	FBgn0036274	CG4328	FBgn0035902	CG6683
FBgn0052264	CG32264	FBgn0263048	CG43343	FBgn0032388	CG6686
FBgn0052532	CG32532	FBgn0263072	CG43347	FBgn0037877	CG6689
FBgn0052700	CG32700	FBgn0038787	CG4360	FBgn0033889	CG6701
FBgn0052719	CG32719	FBgn0265275	CG43674	FBgn0035903	CG6765
FBgn0052767	CG32767	FBgn0263772	CG43689	FBgn0030878	CG6769
FBgn0052772	CG32772	FBgn0039078	CG4374	FBgn0037918	CG6791
FBgn0260741	CG3281	FBgn0264489	CG43897	FBgn0037921	CG6808
FBgn0052982	CG32982	FBgn0264744	CG44002	FBgn0037923	CG6813
FBgn0053017	CG33017	FBgn0030432	CG4404	FBgn0036810	CG6885
FBgn0037980	CG3313	FBgn0031296	CG4415	FBgn0031711	CG6907
FBgn0053178	CG33178	FBgn0038765	CG4424	FBgn0038978	CG7045
FBgn0053213	CG33213	FBgn0265182	CG44247	FBgn0038979	CG7046
FBgn0034985	CG3328	FBgn0031894	CG4496	FBgn0038852	CG7056
FBgn0053557	CG33557	FBgn0039336	CG4553	FBgn0032517	CG7099
FBgn0052831	CG33695	FBgn0037841	CG4565	FBgn0030963	CG7101
FBgn0054031	CG34031	FBgn0037844	CG4570	FBgn0031947	CG7154
FBgn0031573	CG3407	FBgn0029936	CG4617	FBgn0036791	CG7271
FBgn0083985	CG34149	FBgn0033815	CG4676	FBgn0038551	CG7357
FBgn0085369	CG34340	FBgn0035036	CG4707	FBgn0036179	CG7368
FBgn0085396	CG34367	FBgn0039355	CG4730	FBgn0036522	CG7372

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0035691	CG7386	FBgn0029939	CG9650	FBgn0004198	ct
FBgn0038108	CG7518	FBgn0036661	CG9705	FBgn0020496	CtBP
FBgn0030990	CG7556	FBgn0037445	CG9727	FBgn0035769	CTCF
FBgn0038626	CG7691	FBgn0027866	CG9776	FBgn0262707	CTPsyn
FBgn0033616	CG7745	FBgn0037621	CG9797	FBgn0259938	cwo
FBgn0038564	CG7785	FBgn0038146	CG9799	FBgn0023094	cyc
FBgn0034096	CG7786	FBgn0034821	CG9876	FBgn0022936	CycH
FBgn0032016	CG7818	FBgn0031435	CG9883	FBgn0000411	D
FBgn0036124	CG7839	FBgn0034814	CG9890	FBgn0000412	D1
FBgn0039740	CG7928	FBgn0262160	CG9932	FBgn0022935	D19A
FBgn0037584	CG7963	FBgn0035721	CG9948	FBgn0022699	D19B
FBgn0038244	CG7987	FBgn0029504	CHES-1-like	FBgn0033015	d4
FBgn0033993	CG8089	FBgn0013764	Chi	FBgn0000413	da
FBgn0027567	CG8108	FBgn0000307	chif	FBgn0005677	dac
FBgn0030663	CG8117	FBgn0086758	chinmo	FBgn0020493	Dad
FBgn0030664	CG8119	FBgn0028387	chm	FBgn0030093	dalao
FBgn0037617	CG8145	FBgn0015371	chn	FBgn0039286	dan
FBgn0037619	CG8159	FBgn0043002	Chrac-14	FBgn0039283	danr
FBgn0033358	CG8216	FBgn0043001	Chrac-16	FBgn0263239	dar1
FBgn0035824	CG8281	FBgn0004859	ci	FBgn0261723	Dbx
FBgn0026573	CG8290	FBgn0262582	cic	FBgn0013799	Deaf1
FBgn0037717	CG8301	FBgn0015025	CkIIalpha-i1	FBgn0026533	Dek
FBgn0034057	CG8314	FBgn0023076	Clk	FBgn0000439	Dfd
FBgn0037722	CG8319	FBgn0262975	cnc	FBgn0033744	Dh44-R2
FBgn0037634	CG8359	FBgn0034667	comr	FBgn0011274	Dif
FBgn0034062	CG8388	FBgn0263240	Coop	FBgn0023091	dimm
FBgn0037746	CG8478	FBgn0261573	CoRest	FBgn0040467	Dip1
FBgn0030699	CG8578	FBgn0000283	Cp190	FBgn0040466	Dip2
FBgn0033762	CG8632	FBgn0053221	CR33221	FBgn0040465	Dip3
FBgn0036900	CG8765	FBgn0263745	CR43670	FBgn0000459	disco
FBgn0031476	CG8813	FBgn0263746	CR43671	FBgn0042650	disco-r
FBgn0030706	CG8909	FBgn0029920	CR4575	FBgn0260632	dl
FBgn0030710	CG8924	FBgn0000370	crc	FBgn0000157	Dll
FBgn0030680	CG8944	FBgn0005585	Crc	FBgn0262656	dm
FBgn0034186	CG8950	FBgn0004396	CrebA	FBgn0030477	dmrt11E
FBgn0030659	CG9215	FBgn0014467	CrebB-17A	FBgn0038851	dmrt93B
FBgn0032512	CG9305	FBgn0021738	Crg-1	FBgn0039683	dmrt99B
FBgn0026582	CG9418	FBgn0000376	crm	FBgn0027453	Dnz1
FBgn0032485	CG9426	FBgn0014143	croc	FBgn0028789	Doc1
FBgn0034599	CG9437	FBgn0020309	crol	FBgn0035956	Doc2
FBgn0030787	CG9609	FBgn0001994	crp	FBgn0035954	Doc3

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0020306	dom	FBgn0000575	emc	FBgn0040372	G9a
FBgn0011763	Dp	FBgn0000576	ems	FBgn0031495	GABPI
FBgn0010109	dpn	FBgn0000577	en	FBgn0032223	GATAd
FBgn0000492	Dr	FBgn0031375	erm	FBgn0038391	GATAe
FBgn0015664	Dref	FBgn0035849	ERR	FBgn0261703	gce
FBgn0024244	drm	FBgn0000588	esc	FBgn0014179	gcm
FBgn0015381	dsf	FBgn0001981	esg	FBgn0019809	gcm2
FBgn0011764	Dsp1	FBgn0263934	esn	FBgn0004868	Gdi
FBgn0000504	dsx	FBgn0005660	Ets21C	FBgn0050011	gem
FBgn0020307	dve	FBgn0005658	Ets65A	FBgn0250732	gfzf
FBgn0000520	dwg	FBgn0039225	Ets96B	FBgn0004618	gl
FBgn0039411	dys	FBgn0004510	Ets97D	FBgn0263097	Glut4EF
FBgn0000541	E(bx)	FBgn0005659	Ets98B	FBgn0001133	grau
FBgn0002609	E(spl)m3-HLH	FBgn0000606	eve	FBgn0259211	grh
FBgn0002631	E(spl)m5-HLH	FBgn0005427	ewg	FBgn0001138	grn
FBgn0002633	E(spl)m7-HLH	FBgn0000611	exd	FBgn0001139	gro
FBgn0000591	E(spl)m8-HLH	FBgn0041156	exex	FBgn0261278	grp
FBgn0002733	E(spl)mbeta-HLH	FBgn0005558	ey	FBgn0001148	gsb
FBgn0002734	E(spl)mdelta-HLH	FBgn0000625	eyg	FBgn0001147	gsb-n
FBgn0002735	E(spl)mgamma-HLH	FBgn0039937	fd102C	FBgn0010323	Gsc
FBgn0260243	E(var)3-9	FBgn0031086	fd19B	FBgn0001150	gt
FBgn0000617	e(y)1	FBgn0264954	fd3F	FBgn0010825	Gug
FBgn0000618	e(y)2	FBgn0004896	fd59A	FBgn0001168	h
FBgn0087008	e(y)3	FBgn0004895	fd64A	FBgn0016660	H15
FBgn0000629	E(z)	FBgn0036134	fd68A	FBgn0001170	H2.0
FBgn0011766	E2f	FBgn0004897	fd96Ca	FBgn0032812	Hakai
FBgn0024371	E2f2	FBgn0004898	fd96Cb	FBgn0045852	ham
FBgn0008646	E5	FBgn0025832	Fen1	FBgn0032209	Hand
FBgn0000543	ecd	FBgn0037475	Fer1	FBgn0026575	hang
FBgn0000546	EcR	FBgn0038402	Fer2	FBgn0001180	hb
FBgn0000560	eg	FBgn0037937	Fer3	FBgn0008636	hbn
FBgn0086908	egg	FBgn0000659	fkh	FBgn0025825	Hdac3
FBgn0263740	eIF-2gamma	FBgn0038197	foxo	FBgn0026428	HDAC6
FBgn0000567	Eip74EF	FBgn0262477	FoxP	FBgn0001185	her
FBgn0000568	Eip75B	FBgn0004652	fru	FBgn0030899	Her
FBgn0004865	Eip78C	FBgn0004656	fs(1)h	FBgn0027788	Hey
FBgn0264490	Eip93F	FBgn0000927	fs(1)Ya	FBgn0040318	HGTX
FBgn0004858	elB	FBgn0001077	ftz	FBgn0261434	hkb
FBgn0023212	Elongin-B	FBgn0001078	ftz-f1	FBgn0261283	HLH106
FBgn0023211	Elongin-C	FBgn0029173	fu2	FBgn0011276	HLH3B
FBgn0031604	Elp3	FBgn0039932	fuss	FBgn0011277	HLH4C

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0022740	HLH54F	FBgn0001320	kni	FBgn0036761	MED19
FBgn0004362	HmgD	FBgn0001323	knrl	FBgn0013531	MED20
FBgn0010228	HmgZ	FBgn0001325	Kr	FBgn0040339	MED22
FBgn0001206	Hmr	FBgn0028420	Kr-h1	FBgn0038760	MED25
FBgn0264005	Hmx	FBgn0002561	l(1)sc	FBgn0037359	MED27
FBgn0004914	Hnf4	FBgn0016970	l(2)k10201	FBgn0039337	MED28
FBgn0030082	HP1b	FBgn0033029	l(2)NC136	FBgn0035149	MED30
FBgn0039019	HP1c	FBgn0002283	l(3)73Ah	FBgn0037262	MED31
FBgn0261456	hpo	FBgn0002441	l(3)mbt	FBgn0035754	MED4
FBgn0014859	Hr38	FBgn0265276	l(3)neo38	FBgn0051390	MED7
FBgn0261239	Hr39	FBgn0002522	lab	FBgn0011656	Mef2
FBgn0264562	Hr4	FBgn0011278	lbe	FBgn0025874	Meics
FBgn0000448	Hr46	FBgn0008651	lbl	FBgn0035357	MEP-1
FBgn0034012	Hr51	FBgn0034217	Lhr	FBgn0037207	Mes2
FBgn0015239	Hr78	FBgn0031759	lid	FBgn0034726	Mes4
FBgn0037436	Hr83	FBgn0041111	lilli	FBgn0034240	MESR4
FBgn0015240	Hr96	FBgn0026411	Lim1	FBgn0002723	Met
FBgn0015949	hrg	FBgn0002023	Lim3	FBgn0262519	Mi-2
FBgn0001222	Hsf	FBgn0035626	lin-28	FBgn0261963	mid
FBgn0001235	hth	FBgn0262636	Lin29	FBgn0032940	Mio
FBgn0025776	ind	FBgn0039039	lmd	FBgn0033846	mip120
FBgn0031434	insv	FBgn0261565	Lmpt	FBgn0014343	mirr
FBgn0001269	inv	FBgn0034520	lms	FBgn0263112	Mitf
FBgn0011774	Irbp	FBgn0005630	lola	FBgn0263490	mld
FBgn0011604	Iswi	FBgn0022238	lolal	FBgn0014863	Mlp84B
FBgn0036004	Jarid2	FBgn0263667	Lpt	FBgn0023215	Mnt
FBgn0039350	jigr1	FBgn0040765	luna	FBgn0002780	mod
FBgn0027339	jim	FBgn0002576	lz	FBgn0002781	mod(mdg4)
FBgn0086655	jing	FBgn0011648	Mad	FBgn0002783	mor
FBgn0039777	Jon99Fii	FBgn0034534	maf-S	FBgn0027378	MRG15
FBgn0001291	Jra	FBgn0264981	mamo	FBgn0261109	mrn
FBgn0015396	jumu	FBgn0024956	Mat1	FBgn0052296	Mrtf
FBgn0001297	kay	FBgn0017578	Max	FBgn0002775	msl-3
FBgn0037659	Kdm2	FBgn0027950	MBD-like	FBgn0027951	MTA1-like
FBgn0033233	Kdm4A	FBgn0038016	MBD-R2	FBgn0040305	MTF-1
FBgn0011236	ken	FBgn0262732	mbf1	FBgn0032904	Mtp
FBgn0041205	key	FBgn0011655	Med	FBgn0028530	mTTF
FBgn0024887	kin17	FBgn0036811	MED11	FBgn0085444	mute
FBgn0015721	king-tubby	FBgn0027592	MED15	FBgn0260789	mxc
FBgn0013469	klu	FBgn0038578	MED17	FBgn0002914	Myb
FBgn0001319	kn	FBgn0026873	MED18	FBgn0002922	nau

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0034650	NC2alpha	FBgn0003042	Pc	FBgn0014018	Rel
FBgn0028926	NC2beta	FBgn0003044	Pcl	FBgn0011701	repo
FBgn0261617	nej	FBgn0004394	pdm2	FBgn0004795	retn
FBgn0027553	NELF-B	FBgn0261588	pdm3	FBgn0020379	Rfx
FBgn0017430	Nelf-E	FBgn0016694	Pdp1	FBgn0017550	Rga
FBgn0028999	nerfin-1	FBgn0003053	peb	FBgn0033310	rgr
FBgn0041105	nerfin-2	FBgn0004401	Pep	FBgn0003254	rib
FBgn0002931	net	FBgn0003071	Pfk	FBgn0259172	rn
FBgn0037085	Neu2	FBgn0035405	pfk	FBgn0003267	ro
FBgn0035993	Nf-YA	FBgn0004861	ph-p	FBgn0039152	Rootletin
FBgn0032816	Nf-YB	FBgn0025334	PHDP	FBgn0033998	row
FBgn0029905	Nf-YC	FBgn0002521	pho	FBgn0263757	Rpb4
FBgn0030505	NFAT	FBgn0035997	phol	FBgn0051155	Rpb7
FBgn0042696	NfI	FBgn0028579	phtf	FBgn0037121	Rpb8
FBgn0024321	NK7.1	FBgn0261015	Pif1A	FBgn0015805	Rpd3
FBgn0005771	noc	FBgn0046874	Pif1B	FBgn0003275	RpII18
FBgn0085436	Not1	FBgn0034878	pita	FBgn0026373	RpII33
FBgn0085424	nub	FBgn0032401	Plzf	FBgn0004463	RpIII128
FBgn0036812	Nufip	FBgn0003117	pnr	FBgn0003300	run
FBgn0005636	nvy	FBgn0003118	pnt	FBgn0083981	RunxA
FBgn0261613	Oaz	FBgn0039227	polybromo	FBgn0259162	RunxB
FBgn0004102	oc	FBgn0003129	Poxm	FBgn0020617	Rx
FBgn0038063	Octbeta2R	FBgn0003130	Poxn	FBgn0034763	RYBP
FBgn0002985	odd	FBgn0023489	Pph13	FBgn0002842	sa
FBgn0026058	OdsH	FBgn0027945	ppl	FBgn0037672	sage
FBgn0032651	Oli	FBgn0011474	pr-set7	FBgn0261648	salm
FBgn0028996	onecut	FBgn0003145	prd	FBgn0000287	salr
FBgn0003002	opa	FBgn0004595	pros	FBgn0030788	Sap30
FBgn0050443	Opbp	FBgn0005624	Psc	FBgn0016754	sba
FBgn0025360	Optix	FBgn0263102	psq	FBgn0010575	sbb
FBgn0021767	org-1	FBgn0020912	Ptx1	FBgn0004170	sc
FBgn0261885	osa	FBgn0022361	Pur-alpha	FBgn0003330	Sce
FBgn0015524	otp	FBgn0259785	pzg	FBgn0040918	schlank
FBgn0003028	ovo	FBgn0262937	Rabex-5	FBgn0003334	Scm
FBgn0039044	p53	FBgn0020618	Rack1	FBgn0003339	Scr
FBgn0038418	pad	FBgn0037620	ranshi	FBgn0028993	scro
FBgn0085432	pan	FBgn0038128	Ravus	FBgn0004880	scrt
FBgn0051481	pb	FBgn0015799	Rbf	FBgn0003345	sd
FBgn0038371	Pbp45	FBgn0038390	Rbf2	FBgn0002573	sens
FBgn0260398	Pbp49	FBgn0261064	Rbsn-5	FBgn0051632	sens-2
FBgn0037540	Pbp95	FBgn0264493	rdx	FBgn0028991	seq

Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0030486	Set2	FBgn0036248	ssp	FBgn0004110	tin
FBgn0003396	shn	FBgn0010278	Ssrp	FBgn0028979	tio
FBgn0032741	Side	FBgn0016917	Stat92E	FBgn0026080	Tip60
FBgn0004666	sim	FBgn0001978	stc	FBgn0000964	tj
FBgn0015542	sima	FBgn0020249	stck	FBgn0003720	tll
FBgn0022764	Sin3A	FBgn0003459	stwl	FBgn0026160	tna
FBgn0003411	sisA	FBgn0004837	Su(H)	FBgn0036285	toe
FBgn0027364	Six4	FBgn0003567	su(Hw)	FBgn0037751	topi
FBgn0005638	slbo	FBgn0003612	Su(var)2-10	FBgn0033636	tou
FBgn0261477	slim	FBgn0026427	Su(var)2-HP2	FBgn0019650	toy
FBgn0002941	slou	FBgn0003607	Su(var)205	FBgn0038767	trem
FBgn0003430	slp1	FBgn0003598	Su(var)3-7	FBgn0010287	Trf
FBgn0004567	slp2	FBgn0263755	Su(var)3-9	FBgn0261793	Trf2
FBgn0025800	Smox	FBgn0020887	Su(z)12	FBgn0262139	trh
FBgn0265523	Smr	FBgn0265623	Su(z)2	FBgn0013263	Trl
FBgn0003448	sna	FBgn0033782	sug	FBgn0023518	trr
FBgn0085450	Snoo	FBgn0005561	sv	FBgn0003862	trx
FBgn0011715	Snr1	FBgn0003651	svp	FBgn0003866	tsh
FBgn0003460	so	FBgn0086358	Tab2	FBgn0030502	tth
FBgn0004892	sob	FBgn0010355	Taf1	FBgn0003870	ttk
FBgn0024288	Sox100B	FBgn0028398	Taf10	FBgn0003896	tup
FBgn0039938	Sox102F	FBgn0011290	Taf12	FBgn0039530	Tusp
FBgn0005612	Sox14	FBgn0011836	Taf2	FBgn0003900	twi
FBgn0005613	Sox15	FBgn0010280	Taf4	FBgn0263118	tx
FBgn0036411	Sox21a	FBgn0010356	Taf5	FBgn0003944	Ubx
FBgn0042630	Sox21b	FBgn0022724	Taf8	FBgn0024184	unc-4
FBgn0029123	SoxN	FBgn0041092	tai	FBgn0015561	unpg
FBgn0020378	Sp1	FBgn0015550	tap	FBgn0263352	Unr
FBgn0039169	Spps	FBgn0040071	tara	FBgn0029711	Usf
FBgn0037981	Spt3	FBgn0003687	Tbp	FBgn0003963	ush
FBgn0028683	spt4	FBgn0086350	tef	FBgn0003964	usp
FBgn0040273	Spt5	FBgn0003683	term	FBgn0259789	vfl
FBgn0010768	sqz	FBgn0038805	TFAM	FBgn0033748	vis
FBgn0003499	sr	FBgn0261953	TfAP-2	FBgn0261930	vnd
FBgn0003507	srp	FBgn0033929	Tfb1	FBgn0016076	vri
FBgn0003511	Sry-beta	FBgn0015828	TfIIEalpha	FBgn0263511	Vsx1
FBgn0003512	Sry-delta	FBgn0010421	TfIIFbeta	FBgn0263512	Vsx2
FBgn0003513	ss	FBgn0010422	TfIIS	FBgn0086680	vvl
FBgn0015299	Ssb-c31a	FBgn0264075	tgo	FBgn0005642	wdn
FBgn0011481	Ssdp	FBgn0010416	TH1	FBgn0001990	wek
FBgn0037202	Ssl1	FBgn0026869	Thd1	FBgn0010328	woc

Table 2.1 Continued

Flybase ID	**Gene Symbol**				
FBgn0001983	wor				
FBgn0021872	Xbp1				
FBgn0261850	Xpd				
FBgn0261113	Xrp1				
FBgn0005596	yemalpha				
FBgn0034970	yki				
FBgn0032321	YL-1				
FBgn0004050	z				
FBgn0083919	Zasp52				
FBgn0004053	zen				
FBgn0004054	zen2				
FBgn0022720	zf30C				
FBgn0004606	zfh1				
FBgn0004607	zfh2				
FBgn0037446	Zif				

We obtained FLAG-HA tagged inducible expression clones encoding 668 proteins from the Universal Proteomics Resource (Yu et al., 2011; http://fruitfly.org/EST/proteomics/shtml), a part of the Berkeley *Drosophila* Genome Project. These clones contain a metallothionein promoter, which allows for conditional expression upon the addition of copper to the cell culture media. In addition to TFs, we included clones for proteins that are related to transcription, such as basal transcriptional machinery components and chromatin-remodeling proteins. These 668 clones were used in subsequent experiments (Table 2.2).

Table 2.2 Transcription Factor Clones Used in Protein Interactome Study
A list of the 668 FLAG-HA tagged expression clones used in this study.

Table 2.2 Clones used in TF Interactome Study				
Gene Symbol	**Flybase ID**	**TAP ID**	**CG-X_ID**	**Clone ID**
ac	FBgn0000022	FH0553	CG3796-RA	BO05060
achi	FBgn0033749	FH3711	CG8819-RA	BO01514
Ada2b	FBgn0037555	FH3707	CG9638-RA	BO01115
Adf1	FBgn0000054	FH3657	CG15845-RB	BO01438
Aef1	FBgn0005694	FH3675	CG5683-RB	BO01058
AGO1	FBgn0262739	FH7379	CG6671-RB	BO27756
al	FBgn0000061	FH3706	CG3935-RA	BO01508
Antp	FBgn0260642	FH3674	CG1028-RH	BO01056
aop	FBgn0000097	FH5324	CG3166-RA	BO08607
ap	FBgn0000099	FH0305	CG8376-RA	BO01541
AP-1sigma	FBgn0039132	FH0015	CG5864-PA	BO04073
AP-2	FBgn0261953	FH0301	CG7807-RB	BO01534
apt	FBgn0015903	FH4679	CG5393-RA	BO01159
ase	FBgn0000137	FH3578	CG3258-RA	BO12113
ash2	FBgn0000139	FH0338	CG6677-RC	BO01193
Atac1	FBgn0031876	FH4119	CG9200-RA	BO06001
ATbp	FBgn0039946	FH5317	CG40145-RA	BO05253
Atf-2	FBgn0050420	FH5328	CG30420-RA	BO08649
Atf6	FBgn0033010	FH4252	CG3136-RA	BO10638
Atu	FBgn0019637	FH5331	CG1433-RA	BO08746
Awh	FBgn0013751	FH5738	CG1072-RB	BO20573
az2	FBgn0025185	FH4017	CG1605-RA	BO01233
B-H1	FBgn0011758	FH5268	CG5529-RA	BO07915
B-H2	FBgn0004854	FH5251	CG5488-RA	BO05067
bap	FBgn0004862	FH4672	CG7902-RA	BO01488
Bap55	FBgn0025716	FH1314	CG6546-RA	BO08473
Bap60	FBgn0025463	FH5819	CG4303-RA	BO04441
bbx	FBgn0024251	FH3571	CG1414-RB	BO08722
bcd	FBgn0000166	FH5966	CG1034-RG	BO22429
BEAF-32	FBgn0015602	FH5889	CG10159-RB	BO09254
Bgb	FBgn0013753	FH3648	CG7959-RA	BO01028
bigmax	FBgn0039509	FH3658	CG3350-RA	BO01037
bin	FBgn0045759	FH7327	CG18647-RA	BO19661
Bka	FBgn0010520	FH5040	CG4539-RA	BO13518
Blimp-1	FBgn0035625	FH4413	CG5249-RA	BO12273
bowl	FBgn0004893	FH7339	CG10021-RB	BO18764
br	FBgn0000210	FH6059	CG11491-PB	BO22733
Br140	FBgn0033155	FH3572	CG1845-RA	BO08626
Brf	FBgn0038499	FH5270	CG31256-RA	BO08822
brk	FBgn0024250	FH0367	CG9653-RA	BO03983

Table 2.2 Continued

Bro	FBgn0013755	FH6005	CG7960-PA	BO23519
bs	FBgn0004101	FH0362	CG3411-RA	BO03627
Bsg25A	FBgn0000227	FH2187	CG12205-RA	BO12252
bsh	FBgn0000529	FH4737	CG10604-RB	BO05014
BtbVII	FBgn0263108	FH5284	CG11494-RA	BO10974
bun	FBgn0259176	FH3644	CG5461-RC	BO01018
Bx	FBgn0000242	FH3677	CG6500-RA	BO01462
C15	FBgn0004863	FH3432	CG7937-RA	BO18195
cact	FBgn0000250	FH0370	CG5848-RD	BO05426
cad	FBgn0000251	FH3712	CG1759-RA	BO01124
Caf1	FBgn0015610	FH1340	CG4236-RA	BO08946
cas	FBgn0004878	FH5283	CG2102-RA	BO10966
cato	FBgn0024249	FH5715	CG7760-RA	BO21325
caup	FBgn0015919	FH5325	CG10605-RA	BO08619
caz	FBgn0011571	FH0975	CG3606-RB	BO07087
cbt	FBgn0043364	FH3475	CG4427-RB	BO18407
Cdk8	FBgn0015618	FH7244	CG10572-RA	BO27522
Cdk9	FBgn0019949	FH2827	CG5179-RA	BO14857
Cf2	FBgn0000286	FH5924	CG11924-RB	BO20005
cg	FBgn0000289	FH5895	CG8367-RA	BO15450
CG10209	FBgn0033971	FH1130	CG10209-RA	BO08157
CG1024	FBgn0027514	FH3902	CG1024-RA	BO17594
CG10263	FBgn0032812	FH4231	CG10263-RA	BO10225
CG10267	FBgn0037446	FH3697	CG10267-RA	BO01108
CG10274	FBgn0035690	FH0333	CG10274-RA	BO01177
CG10321	FBgn0084603	FH5287	CG10321-RA	BO11018
CG10344	FBgn0034729	FH0318	CG10344-RA	BO02425
CG10348	FBgn0032707	FH6829	CG10348-RA	BO08637
CG10366	FBgn0032814	FH5915	CG10366-RA	BO01220
CG10414	FBgn0032691	FH4046	CG10414-RA	BO04846
CG10431	FBgn0032730	FH5305	CG10431-RA	BO05013
CG10462	FBgn0032815	FH7575	CG10462-RA	BO20690
CG10565	FBgn0037051	FH1869	CG10565-RA	BO10151
CG10654	FBgn0036294	FH5614	CG10654-RA	N/A
CG10669	FBgn0039329	FH6679	CG10669-RA	BO17672
CG10904	FBgn0034945	FH2222	CG10904-RA	BO12523
CG10949	FBgn0032858	FH1143	CG10949-RA	BO08183
CG10959	FBgn0030010	FH0371	CG10959-RA	BO05444
CG11247	FBgn0037120	FH3473	CG11247-RA	BO18404
CG11294	FBgn0030058	FH6663	CG11294-RA	BO05118
CG11317	FBgn0039816	FH1125	CG11317-RA	BO07678
CG11398	FBgn0040366	FH3682	CG11398-RA	BO01470

Table 2.2 Continued

CG11504	FBgn0039733	FH1210	CG11504-RB	BO08529
CG11617	FBgn0031232	FH3673	CG11617-RA	BO01458
CG11641	FBgn0033288	FH5294	CG11641-RA	BO01337
CG11696	FBgn0030314	FH3544	CG11696-RA	BO01589
CG11710	FBgn0031115	FH5062	CG11710-RA	BO14236
CG11723	FBgn0031391	FH1313	CG11723-RA	BO08572
CG11906	FBgn0034425	FH3542	CG11906-RA	BO01583
CG12029	FBgn0263239	FH5306	CG12029-RB	BO05019
CG12054	FBgn0039831	FH4286	CG12054-RA	BO11026
CG12071	FBgn0039808	FH3649	CG12071-RA	BO01030
CG12075	FBgn0030065	FH6518	CG12075-RA	BO25077
CG12104	FBgn0035238	FH1619	CG12104-RA	BO10373
CG12162	FBgn0037329	FH1500	CG12162-RA	BO09837
CG12190	FBgn0034763	FH3640	CG12190-RA	BO01412
CG12219	FBgn0043796	FH0339	CG12219-RA	BO01203
CG12236	FBgn0029822	FH5846	CG12236-RA	BO18091
CG12267	FBgn0038057	FH1005	CG12267-RA	BO07253
CG12299	FBgn0032295	FH4683	CG12299-RA	BO01289
CG1233	FBgn0035137	FH6223	CG1233-RB	BO17791
CG12370	FBgn0033744	FH5880	CG12370-RA	BO21748
CG12391	FBgn0068463	FH0342	CG12391-RA	BO01228
CG12605	FBgn0035481	FH6220	CG12605-RA	BO18186
CG12659	FBgn0040929	FH4705	CG12659-RB	BO04454
CG12744	FBgn0033459	FH5259	CG12744-RA	BO06060
CG12768	FBgn0037206	FH6993	CG12768-RA	BO25158
CG12769	FBgn0033252	FH6728	CG12769-RA	BO25974
CG12942	FBgn0033569	FH0352	CG12942-RA	BO01275
CG13123	FBgn0032150	FH3681	CG13123-RA	BO01468
CG13183	FBgn0033667	FH7025	CG13183-RA	BO25411
CG13204	FBgn0033627	FH5420	CG13204-RB	BO17129
CG13458	FBgn0036479	FH4160	CG13458-RA	BO08055
CG13624	FBgn0039209	FH3540	CG13624-RA	BO03548
CG14014	FBgn0031718	FH5798	CG14014-RB	BO20854
CG1407	FBgn0033474	FH3685	CG1407-RB	BO01074
CG14207	FBgn0031037	FH0004	CG14207-PA	BO06709
CG14260	FBgn0039504	FH6177	CG14260-PA	BO24114
CG14451	FBgn0037183	FH2792	CG14451-RA	BO16289
CG14655	FBgn0037275	FH0335	CG14655-RA	BO02582
CG14667	FBgn0037317	FH3661	CG14667-RA	BO01444
CG14710	FBgn0037920	FH0325	CG14710-RA	BO02547
CG14712	FBgn0037924	FH5286	CG14712-RA	BO11001
CG14767	FBgn0040777	FH1623	CG14767-RA	BO10286

Table 2.2 Continued

CG14860	FBgn0038273	FH5603	CG14860-RA	N/A
CG14962	FBgn0035407	FH0321	CG14962-RA	BO02505
CG15011	FBgn0035518	FH5319	CG15011-RA	BO05301
CG15086	FBgn0034374	FH6535	CG15086-RE	BO24395
CG15286	FBgn0028531	FH6562	CG15286-RA	BO25623
CG1529	FBgn0031144	FH5783	CG1529-RA	BO20791
CG15433	FBgn0031604	FH2242	CG15433-RA	BO12568
CG15435	FBgn0031608	FH5074	CG15435-RA	BO14510
CG15436	FBgn0031610	FH3686	CG15436-RA	BO01078
CG15479	FBgn0032493	FH3517	CG15479-RA	BO18565
CG15514	FBgn0039712	FH0593	CG15514-RA	BO06031
CG15601	FBgn0030673	FH5716	CG15601-RA	BO19601
CG15710	FBgn0034120	FH0535	CG15710-RA	BO05027
CG15835	FBgn0033233	FH1337	CG15835-RA	BO08938
CG1603	FBgn0033185	FH3966	CG1603-RA	BO18502
CG1620	FBgn0033183	FH7383	CG1620-RA	BO27781
CG16778	FBgn0003715	FH5303	CG16778-RA	BO05005
CG16815	FBgn0032491	FH5708	CG16815-RA	BO19017
CG16863	FBgn0028931	FH3994	CG16863-RA	BO12517
CG16899	FBgn0037735	FH0536	CG16899-RA	BO05031
CG16975	FBgn0032475	FH3547	CG9495-RA	BO01619
CG17118	FBgn0032291	FH3667	CG17118-RA	BO01047
CG17186	FBgn0038741	FH0537	CG17186-RA	BO05132
CG17195	FBgn0039369	FH0538	CG17195-RA	BO05033
CG17197	FBgn0039367	FH0539	CG17197-RA	BO05135
CG17198	FBgn0039366	FH0540	CG17198-RA	BO05136
CG17328	FBgn0028895	FH0541	CG17328-RA	BO05038
CG17359	FBgn0036396	FH4668	CG17359-RA	BO01474
CG17361	FBgn0036395	FH2184	CG17361-RA	BO12243
CG17385	FBgn0068384	FH0542	CG17385-RA	BO05039
CG17568	FBgn0032763	FH0543	CG17568-RA	BO05040
CG17802	FBgn0038549	FH0299	CG17802-RA	BO01520
CG17803	FBgn0038547	FH6115	CG17803-PA	BO05041
CG17806	FBgn0038548	FH4673	CG17806-RA	BO01120
CG17829	FBgn0025635	FH0302	CG17829-RB	BO01536
CG17912	FBgn0032600	FH4670	CG17912-RA	BO01084
CG1792	FBgn0039860	FH3692	CG1792-RA	BO01092
CG18011	FBgn0033491	FH5313	CG18011-RA	BO05242
CG18446	FBgn0033458	FH4484	CG18446-RA	BO13202
CG18476	FBgn0037931	FH5311	CG18476-RA	BO05088
CG18619	FBgn0032202	FH3586	CG18619-RA	BO14043
CG18764	FBgn0042205	FH5624	CG18764-RA	N/A

Table 2.2 Continued

CG1965	FBgn0037466	FH3909	CG1965-RA	BO17926
CG2116	FBgn0030003	FH5901	CG2116-RA	BO18024
CG2129	FBgn0030008	FH0328	CG2129-RA	BO01155
CG2199	FBgn0035213	FH6709	CG2199-RA	BO26446
CG2611	FBgn0032871	FH3636	CG2611-RA	BO01008
CG2652	FBgn0025838	FH0671	CG2652-RA	BO06350
CG2678	FBgn0014931	FH5278	CG2678-RA	BO10070
CG2702	FBgn0037540	FH4039	CG2702-RA	BO04620
CG2712	FBgn0024975	FH4680	CG2712-RA	BO01170
CG2790	FBgn0027599	FH3920	CG2790-RA	BO17965
CG2875	FBgn0029672	FH5898	CG2875-RA	BO17492
CG2889	FBgn0030206	FH3589	CG2889-RA	BO14278
CG30084	FBgn0083919	FH5307	CG30084-RG	BO05045
CG30417	FBgn0050417	FH0546	CG30417-RA	BO05046
CG30431	FBgn0050431	FH6114	CG30431-PA	BO05147
CG31367	FBgn0051367	FH1278	CG31367-RA	BO08126
CG31388	FBgn0051388	FH0527	CG31388-RA	BO05007
CG31441	FBgn0051441	FH0570	CG31441-RA	BO05090
CG31460	FBgn0051460	FH5290	CG31460-RA	BO15642
CG31612	FBgn0051612	FH5308	CG31612-RA	BO05052
CG3163	FBgn0034961	FH0596	CG3163-RA	BO06134
CG31835	FBgn0051835	FH5576	CG31835-RA	N/A
CG31955	FBgn0051955	FH5602	CG31955-RA	N/A
CG32264	FBgn0052264	FH5891	CG32264-RB	BO10673
CG3227	FBgn0031434	FH5936	CG3227-RA	BO22890
CG32700	FBgn0052700	FH1296	CG32700-RA	BO08054
CG32721	FBgn0027553	FH5206	CG32721-RA	BO05938
CG3281	FBgn0260741	FH0336	CG3281-RA	BO01185
CG32982	FBgn0052982	FH4370	CG32982-RB	BO11843
CG33097	FBgn0053097	FH4916	CG33097-RA	BO09943
CG3313	FBgn0037980	FH0965	CG3313-RA	BO07070
CG33178	FBgn0053178	FH2479	CG33178-RA	BO13764
CG33213	FBgn0053213	FH0517	CG33213-RA	BO04625
CG33980	FBgn0053980	FH5248	CG33980-RA	BO05128
CG3407	FBgn0031573	FH5279	CG3407-RA	BO10190
CG34149	FBgn0083985	FH6351	CG34149-RA	BO24546
CG34360	FBgn0250818	FH3111	CG32469-RA	BO15786
CG34376	FBgn0085405	FH6105	CG34376-PA	BO22791
CG3526	FBgn0040355	FH6197	CG3526-PC	BO24159
CG3711	FBgn0040344	FH6528	CG3711-RB	BO24957
CG3726	FBgn0029824	FH0351	CG3726-RA	BO01273
CG3815	FBgn0029861	FH1612	CG3815-RA	BO07423

Table 2.2 Continued

CG3838	FBgn0032130	FH3799	CG3838-RB	BO05482
CG3909	FBgn0027524	FH0372	CG3909-RA	BO05676
CG3919	FBgn0036423	FH3699	CG3919-RA	BO01492
CG3995	FBgn0038472	FH0935	CG3995-RA	BO07126
CG40196	FBgn0058196	FH5654	CG40196-RA	BO18614
CG4042	FBgn0037018	FH7001	CG4042-RA	BO25396
CG4133	FBgn0031257	FH4495	CG4133-RA	BO13228
CG4282	FBgn0034114	FH0347	CG4282-RA	BO01263
CG4318	FBgn0030455	FH0555	CG4318-RA	BO05063
CG4328	FBgn0036274	FH6861	CG4328-RA	BO25934
CG4360	FBgn0038787	FH7164	CG4360-RB	BO26452
CG4415	FBgn0031296	FH0526	CG4415-RA	BO05106
CG4424	FBgn0038765	FH5261	CG4424-RA	BO07023
CG4565	FBgn0037841	FH0556	CG4565-RA	BO05064
CG4617	FBgn0029936	FH4835	CG4617-RA	BO09291
CG4707	FBgn0035036	FH0350	CG4707-RA	BO01271
CG4730	FBgn0039355	FH3581	CG4730-RA	BO12628
CG4756	FBgn0030788	FH5275	CG4756-RA	BO08618
CG4854	FBgn0038766	FH3678	CG4854-RA	BO01464
CG4882	FBgn0025336	FH0600	CG4882-RA	BO06038
CG4936	FBgn0038768	FH4724	CG4936-RA	BO05406
CG4956	FBgn0039370	FH0557	CG4956-RA	BO05065
CG5033	FBgn0028744	FH5299	CG5033-RA	BO01307
CG5196	FBgn0038039	FH6004	CG5196-PB	BO23694
CG5204	FBgn0032473	FH0354	CG5204-RA	BO01293
CG5343	FBgn0032248	FH3580	CG5343-RA	BO12439
CG5641	FBgn0038046	FH3700	CG5641-RA	BO01110
CG5687	FBgn0035293	FH1700	CG5687-RA	BO07566
CG5708	FBgn0032196	FH5656	CG5708-RB	BO18616
CG5846	FBgn0032171	FH3654	CG5846-RA	BO01034
CG6118	FBgn0038339	FH5310	CG6118-RB	BO05083
CG6254	FBgn0037794	FH7318	CG6254-RA	BO18688
CG6272	FBgn0036126	FH3634	CG6272-RA	BO01006
CG6276	FBgn0038316	FH4853	CG6276-RA	BO10434
CG6470	FBgn0030933	FH1164	CG6470-RA	BO08335
CG6654	FBgn0038301	FH5267	CG6654-RA	BO07550
CG6683	FBgn0035902	FH0389	CG6683-RA	BO04058
CG6689	FBgn0037877	FH5908	CG6689-RA	BO08605
CG6769	FBgn0030878	FH3794	CG6769-RA	BO05430
CG6808	FBgn0037921	FH5928	CG6808-RA	BO20002
CG6812	FBgn0036843	FH7026	CG6812-RA	BO25317
CG6854	FBgn0036478	FH3025	CG6854-RC	BO15940

Table 2.2 Continued

CG6885	FBgn0036810	FH5701	CG6885-RA	BO18886
CG6902	FBgn0035899	FH6962	CG6902-RA	BO27776
CG6905	FBgn0035136	FH3901	CG6905-RA	BO17491
CG6907	FBgn0031711	FH2672	CG6907-RA	BO14655
CG6930	FBgn0037947	FH3693	CG6930-RA	BO01486
CG7101	FBgn0030963	FH0320	CG7101-RA	BO01080
CG7154	FBgn0031947	FH0358	CG7154-RA	BO01319
CG7271	FBgn0036791	FH5703	CG7271-RA	BO18891
CG7357	FBgn0038551	FH0322	CG7357-RA	BO01131
CG7372	FBgn0036522	FH6804	CG7372-RA	BO26424
CG7386	FBgn0035691	FH0348	CG7386-RA	BO01265
CG7556	FBgn0030990	FH1042	CG7556-RA	BO07529
CG7745	FBgn0033616	FH1332	CG7745-RA	BO08931
CG7785	FBgn0038564	FH2251	CG7785-RA	BO12593
CG7818	FBgn0032016	FH3701	CG7818-RA	BO01502
CG7928	FBgn0039740	FH0368	CG7928-RA	BO04488
CG8108	FBgn0027567	FH3907	CG8108-RA	BO17917
CG8117	FBgn0030663	FH0563	CG8117-RA	BO05180
CG8145	FBgn0037617	FH0569	CG8145-RA	BO05087
CG8281	FBgn0035824	FH2965	CG8281-RA	BO15421
CG8301	FBgn0037717	FH5916	CG8301-RA	BO01243
CG8314	FBgn0034057	FH3670	CG8314-RA	BO01054
CG8319	FBgn0064278	FH5258	CG8319-RA	BO05251
CG8359	FBgn0037634	FH5301	CG8359-RA	BO04255
CG8388	FBgn0034062	FH6546	CG8388-RA	BO24392
CG8478	FBgn0037746	FH0341	CG8478-RB	BO01209
CG8578	FBgn0030699	FH1551	CG8578-RA	BO03607
CG8765	FBgn0036900	FH5314	CG8765-RA	BO05245
CG8813	FBgn0031476	FH3668	CG8813-RA	BO01050
CG8924	FBgn0030710	FH0518	CG8924-RB	BO04633
CG8944	FBgn0030680	FH5320	CG8944-RB	BO05308
CG8950	FBgn0034186	FH5300	CG8950-RA	BO01323
CG9215	FBgn0030659	FH5894	CG9215-RA	BO13855
CG9416	FBgn0034438	FH4418	CG9416-RA	BO12292
CG9418	FBgn0026582	FH1451	CG9418-RA	BO09503
CG9426	FBgn0032485	FH0344	CG9426-RA	BO01249
CG9597	FBgn0038371	FH4000	CG9597-RA	BO13345
CG9609	FBgn0030787	FH3704	CG9609-RA	BO01113
CG9797	FBgn0037621	FH3708	CG9797-RA	BO01118
CG9799	FBgn0038146	FH5329	CG9799-RA	BO08655
CG9890	FBgn0034814	FH4836	CG9890-RA	BO09402
CG9948	FBgn0035721	FH0482	CG9948-RA	BO04428

Table 2.2 Continued

Chi	FBgn0013764	FH3975	CG3924-RA	BO08505
chinmo	FBgn0086758	FH5316	CG31666-RA	BO05252
chn	FBgn0015371	FH5930	CG11798-RC	BO22241
Chrac-16	FBgn0043001	FH3638	CG15736-RA	BO01408
cic	FBgn0262582	FH6773	CG5067-RA	BO26426
CkIIalpha-i1	FBgn0015025	FH2409	CG6215-RA	BO13419
cnc	FBgn0262975	FH7295	CG17894-RX	BO27754
Cog7	FBgn0051040	FH3875	CG31040-RA	BO17731
corto	FBgn0010313	FH6530	CG2530-RA	BO24953
crc	FBgn0000370	FH3705	CG9429-RA	BO01505
CrebA	FBgn0004396	FH5855	CG7450-RA	BO18882
CrebB-17A	FBgn0014467	FH5709	CG6103-RE	BO19025
CREG	FBgn0025456	FH4600	CG5413-RA	BO14976
crm	FBgn0000376	FH1943	CG2714-RB	BO10922
croc	FBgn0014143	FH0307	CG5069-RA	BO01551
crp	FBgn0001994	FH7233	CG7664-RA	BO26862
ct	FBgn0004198	FH0003	CG11387-RB	BO01441
CtBP	FBgn0020496	FH4851	CG7583-RA	BO10420
CTCF	FBgn0035769	FH5281	CG8591-RA	BO10663
cyc	FBgn0023094	FH3474	CG8727-RA	BO18505
CycH	FBgn0022936	FH0319	CG7405-RA	BO01065
D	FBgn0000411	FH6550	CG5893-RA	BO25591
D1	FBgn0000412	FH2782	CG9745-RC	BO16265
D19A	FBgn0022935	FH5312	CG10269-RA	BO05220
D19B	FBgn0022699	FH5902	CG10270-RA	BO18082
d4	FBgn0033015	FH0330	CG2682-RA	BO02565
dac	FBgn0005677	FH3679	CG4952-RF	BO01466
Dad	FBgn0020493	FH5322	CG5201-RA	BO05349
dalao	FBgn0030093	FH5922	CG7055-RA	BO17857
dan	FBgn0039286	FH5893	CG11849-RA	BO11229
danr	FBgn0039283	FH4480	CG13651-RA	BO13089
Deaf1	FBgn0013799	FH6806	CG8567-RB	BO26459
debcl	FBgn0029131	FH0549	CG33134-RA	BO05156
dei	FBgn0263118	FH3695	CG5441-RA	BO01102
Dek	FBgn0026533	FH5323	CG5935-RB	BO05424
Dif	FBgn0011274	FH7331	CG6794-RA	BO18870
dimm	FBgn0023091	FH5712	CG8667-RA	BO21311
Dip1	FBgn0040467	FH5410	CG15367-RA	BO16745
Dip2	FBgn0040466	FH6012	CG9771-PA	BO23520
Dip3	FBgn0040465	FH3687	CG12767-RA	BO01477
disco	FBgn0000459	FH0340	CG9908-RA	BO01205
dl	FBgn0260632	FH4023	CG3619-RA	BO03472

Table 2.2 Continued

dmrt99B	FBgn0039683	FH0534	CG15504-RA	BO05026
Dnz1	FBgn0027453	FH3664	CG6627-RA	BO01447
Doc1	FBgn0028789	FH5700	CG5133-RA	BO18877
Doc2	FBgn0035956	FH0304	CG5187-RA	BO01540
Doc3	FBgn0035954	FH3710	CG5093-RA	BO01512
Dp	FBgn0011763	FH7059	CG4654-RB	BO26871
dpn	FBgn0010109	FH0298	CG8704-RA	BO02217
Dr	FBgn0000492	FH4720	CG1897-RA	BO04881
Dref	FBgn0015664	FH7296	CG5838-RA	BO27191
drm	FBgn0024244	FH3632	CG10016-RB	BO01004
Dsp1	FBgn0011764	FH7073	CG12223-RF	BO27648
dsx	FBgn0000504	FH5766	CG11094-RA	BO20688
dve	FBgn0020307	FH5923	CG5799-RA	BO19655
dwg	FBgn0000520	FH0343	CG2711-RA	BO01232
dys	FBgn0039411	FH5309	CG32474-RB	BO05077
E(spl)	FBgn0000591	FH5714	CG8365-RA	BO21319
e(y)1	FBgn0000617	FH3656	CG6474-RA	BO01035
e(y)2	FBgn0000618	FH0533	CG15191-RA	BO05023
E(z)	FBgn0000629	FH5293	CG6502-RA	BO01296
E2f	FBgn0011766	FH6794	CG6376-RA	BO22975
E2f2	FBgn0024371	FH3691	CG1071-RA	BO01090
E5	FBgn0008646	FH6978	CG9930-RA	BO25289
ecd	FBgn0000543	FH1713	CG5714-RA	BO07647
EcR	FBgn0000546	FH5920	CG1765-RB	BO08711
egg	FBgn0086908	FH7406	CG30426-RA	BO05001
Eip63E	FBgn0005640	FH0565	CG10579-RB	BO05182
Eip75B	FBgn0000568	FH6801	CG8127-RD	BO26551
elB	FBgn0004858	FH5256	CG4220-RC	BO05229
EloA	FBgn0039066	FH1874	CG6755-RB	BO10187
Elongin-B	FBgn0023212	FH0293	CG4204-RA	BO18365
Elongin-C	FBgn0023211	FH0237	CG9291-RA	BO14135
emc	FBgn0000575	FH0001	CG1007-RA	BO01423
ems	FBgn0000576	FH5253	CG2988-RA	BO05317
en	FBgn0000577	FH0369	CG9015-RA	BO04957
ERR	FBgn0035849	FH0528	CG7404-RA	BO05009
esc	FBgn0000588	FH6071	CG14941-PA	BO10575
esn	FBgn0263934	FH7321	CG12833-RA	BO19814
Ets21C	FBgn0005660	FH5615	CG2914-RA	N/A
Ets65A	FBgn0005658	FH3659	CG7018-RB	BO01040
Ets96B	FBgn0039225	FH5318	CG6892-RB	BO05254
Ets98B	FBgn0005659	FH6481	CG5583-RA	BO26075
exd	FBgn0000611	FH4671	CG8933-RA	BO01483

Table 2.2 Continued

ey	FBgn0005558	FH5917	CG1464-RB	BO01245
fd102C	FBgn0039937	FH4686	CG11152-RA	BO05017
fd59A	FBgn0004896	FH5182	CG3668-RA	BO18406
Fen1	FBgn0025832	FH7261	CG8648-RA	BO27387
Fer3	FBgn0037937	FH3647	CG6913-RA	BO01422
fs(1)h	FBgn0004656	FH5326	CG2252-RA	BO08641
fs(1)Ya	FBgn0000927	FH1870	CG2707-RA	BO10161
ftz	FBgn0001077	FH0544	CG2047-RA	BO05143
ftz-f1	FBgn0001078	FH6823	CG4059-RB	BO23925
fu2	FBgn0029173	FH7333	CG9233-RA	BO18884
FucTB	FBgn0032117	FH6579	CG4435-RA	BO26406
G9a	FBgn0040372	FH4678	CG2995-RA	BO01637
GATAe	FBgn0038391	FH0316	CG10278-RA	BO01611
gcl	FBgn0005695	FH4150	CG8411-RA	BO07307
gcm2	FBgn0019809	FH5250	CG3858-RA	BO05162
Gdi	FBgn0004868	FH0323	CG4422-RA	BO01141
gem	FBgn0050011	FH5302	CG30011-RA	BO05002
gfzf	FBgn0250732	FH5811	CG33546-RD	BO20906
gl	FBgn0004618	FH5282	CG7672-RB	BO10771
gol	FBgn0004919	FH0005	CG2679-RB	BO01151
grau	FBgn0001133	FH5264	CG33133-RA	BO07402
grh	FBgn0259211	FH3641	CG5058-RF	BO01012
grn	FBgn0001138	FH0306	CG9656-RA	BO02243
gro	FBgn0001139	FH4089	CG8384-RA	BO05661
grp	FBgn0011598	FH5168	CG17161-RA	BO17862
gsb	FBgn0001148	FH0552	CG3388-RA	BO05059
gsb-n	FBgn0001147	FH5276	CG2692-RA	BO08620
gt	FBgn0001150	FH7277	CG7952-RB	BO21404
h	FBgn0001168	FH3684	CG6494-RB	BO01472
H15	FBgn0016660	FH5905	CG6604-RA	BO05071
H2.0	FBgn0001170	FH5698	CG11607-RA	BO18874
hb	FBgn0001180	FH5912	CG9786-RA	BO20286
hbn	FBgn0008636	FH0550	CG33152-RA	BO05057
Hdac3	FBgn0025825	FH4791	CG2128-RA	BO07312
HDAC6	FBgn0026428	FH4297	CG6170-RA	BO11104
Her	FBgn0030899	FH0558	CG5927-RA	BO05068
HGTX	FBgn0040318	FH5254	CG13475-RA	BO05326
HLH106	FBgn0261283	FH0361	CG8522-RA	BO01347
HLH3B	FBgn0011276	FH0545	CG2655-RA	BO05044
HLH4C	FBgn0011277	FH0547	CG3052-RA	BO05049
HLH54F	FBgn0022740	FH5713	CG5005-RA	BO21313
HLHm3	FBgn0002609	FH3652	CG8346-RA	BO01432

Table 2.2 Continued

HLHm5	FBgn0002631	FH0364	CG6096-RA	BO03637
HLHm7	FBgn0002633	FH0566	CG8361-RA	BO05084
HLHmbeta	FBgn0002733	FH3646	CG14548-RA	BO01022
HLHmdelta	FBgn0002734	FH0561	CG8328-RA	BO05075
HLHmgamma	FBgn0002735	FH0564	CG8333-RA	BO05181
HmgD	FBgn0004362	FH0602	CG17950-RA	BO06040
HmgZ	FBgn0010228	FH3579	CG17921-RA	BO12396
Hmx	FBgn0264005	FH7581	CG34419-RC	BO20654
hpo	FBgn0034453	FH4880	CG11228-RA	BO07565
Hr39	FBgn0261239	FH0356	CG8676-RD	BO01311
Hr46	FBgn0000448	FH3574	CG33183-RB	BO10774
Hr78	FBgn0015239	FH3561	CG7199-RA	BO07537
hrg	FBgn0015949	FH3543	CG9854-RA	BO01586
Hsf	FBgn0001222	FH0349	CG5748-RA	BO01267
hth	FBgn0001235	FH5696	CG17117-RC	BO18860
inv	FBgn0001269	FH5906	CG17835-RA	BO05095
Irbp	FBgn0011774	FH6695	CG5247-RA	BO25410
Iswi	FBgn0011604	FH7313	CG8625-RA	BO19895
jigr1	FBgn0039350	FH0366	CG17383-RA	BO03975
jim	FBgn0027339	FH6747	CG11352-RB	BO26423
jing	FBgn0086655	FH6402	CG9397-RH	BO22152
Jra	FBgn0001291	FH3669	CG2275-RA	BO01052
jumu	FBgn0015396	FH4682	CG4029-RA	BO01283
kay	FBgn0001297	FH0337	CG15509-RB	BO02592
ken	FBgn0011236	FH5295	CG5575-RA	BO05856
kin17	FBgn0024887	FH7014	CG5649-RA	BO25382
king-tubby	FBgn0015721	FH1223	CG9398-RA	BO07802
klu	FBgn0013469	FH5968	CG12296-RA	BO22451
kni	FBgn0001320	FH4675	CG4717-RA	BO01127
knrl	FBgn0001323	FH0346	CG4761-RA	BO01262
Kr-h1	FBgn0028420	FH5327	CG18783-RB	BO08648
l(1)sc	FBgn0002561	FH0296	CG3839-RA	BO01439
l(2)NC136	FBgn0033029	FH4057	CG8426-RA	BO05462
l(3)73Ah	FBgn0002283	FH3650	CG4195-RA	BO01427
l(3)mbt	FBgn0002441	FH3549	CG5954-RA	BO01636
l(3)neo38	FBgn0086910	FH3602	CG31364-RA	BO15871
lab	FBgn0002522	FH0313	CG1264-RA	BO02281
lbl	FBgn0008651	FH0562	CG6570-RA	BO05078
Lhr	FBgn0034217	FH0393	CG18468-RA	BO04065
lid	FBgn0031759	FH5927	CG9088-RA	BO11422
Lim1	FBgn0026411	FH4685	CG11354-RA	BO05650
Lim3	FBgn0002023	FH5857	CG10699-RB	BO19913

Table 2.2 Continued

lin-28	FBgn0035626	FH2046	CG17334-RA	BO11664
Lin29	FBgn0262636	FH7334	CG2052-RA	BO23111
lmd	FBgn0039039	FH5288	CG4677-RB	BO11030
Lmpt	FBgn0036672	FH5315	CG32171-RB	BO05248
lola	FBgn0005630	FH3545	CG12052-RA	BO01604
lolal	FBgn0022238	FH0006	CG5738-PA	BO01404
luna	FBgn0040765	FH5896	CG33473-RB	BO17282
Mad	FBgn0011648	FH0365	CG12399-RA	BO03940
maf-S	FBgn0034534	FH0317	CG9954-RA	BO01010
Mat1	FBgn0024956	FH5702	CG7614-RA	BO18888
Max	FBgn0017578	FH5711	CG9648-RA	BO21305
mbf1	FBgn0026208	FH3838	CG4143-RA	BO05835
MED11	FBgn0036811	FH4586	CG6884-RA	BO14801
MED15	FBgn0027592	FH4083	CG4184-RA	BO05644
MED17	FBgn0038578	FH0345	CG7957-RA	BO01255
MED18	FBgn0026873	FH7427	CG14802-RA	BO27188
MED19	FBgn0036761	FH1541	CG5546-RA	BO10336
MED20	FBgn0013531	FH5031	CG18780-RA	BO12994
MED22	FBgn0040339	FH2127	CG3034-RA	BO12135
MED25	FBgn0038760	FH4034	CG12254-RA	BO04431
MED27	FBgn0037359	FH3671	CG1245-RA	BO01454
MED28	FBgn0039337	FH2848	CG5121-RA	BO14994
MED30	FBgn0035149	FH0851	CG17183-RA	BO06806
MED31	FBgn0037262	FH0772	CG1057-RA	BO06595
MED4	FBgn0035754	FH4780	CG8609-RA	BO06989
MED6	FBgn0024330	FH3655	CG9473-RA	BO01434
MED7	FBgn0051390	FH3240	CG31390-RA	BO16655
MED9	FBgn0034311	FH4687	CG5134-RC	BO04059
Mef2	FBgn0011656	dMef2	CG1429-RE	N/A
Meics	FBgn0025874	FH0310	CG8474-RA	BO02274
Mes2	FBgn0037207	FH1488	CG11100-RB	BO09785
Mes4	FBgn0034726	FH0567	CG11301-RA	BO05185
Met	FBgn0002723	FH7594	CG1705-RA	BO21609
mid	FBgn0261963	FH5900	CG6634-RA	BO17893
Mio	FBgn0032940	FH7343	CG18362-RA	BO21428
Mitf	FBgn0263112	FH7304	CG17469-RX	BO27133
Mlp84B	FBgn0014863	FH4676	CG1019-RA	BO01547
Mnt	FBgn0023215	FH5296	CG13316-RA	BO10110
mod	FBgn0002780	FH1603	CG2050-RA	BO07076
mod(mdg4)	FBgn0002781	FH6929	CG32491-RD	BO27343
mor	FBgn0002783	FH6398	CG18740-RA	BO22139
MRG15	FBgn0027378	FH1194	CG6363-RA	BO08391

Table 2.2 Continued

msl-3	FBgn0002775	FH4980	CG8631-RA	BO12101
mtTFB2	FBgn0037778	FH1495	CG3910-RA	BO09911
Myb	FBgn0002914	FH5910	CG9045-RA	BO11256
nau	FBgn0002922	FH0529	CG10250-RA	BO05111
NC2beta	FBgn0028926	FH5897	CG4185-RA	BO17349
Nelf-E	FBgn0017430	FH2109	CG5994-RA	BO11985
nerfin-1	FBgn0028999	FH0303	CG13906-RA	BO01537
net	FBgn0002931	FH6830	CG11450-RB	BO21832
Neu2	FBgn0037085	FH5913	CG7204-RA	BO20006
Nf-YA	FBgn0035993	FH3703	CG3891-RA	BO01112
Nf-YB	FBgn0032816	FH4665	CG10447-RA	BO01414
Nf-YC	FBgn0029905	FH0312	CG3075-RA	BO01577
NfI	FBgn0042696	FH6802	CG2380-RB	BO26537
nub	FBgn0085424	FH0311	CG6246-RA	BO01576
nvy	FBgn0005636	FH3555	CG3385-RA	BO05579
Octbeta2R	FBgn0038063	FH7154	CG33976-RA	BO25951
odd	FBgn0002985	FH0297	CG3851-RA	BO01495
OdsH	FBgn0026058	FH5904	CG6352-RA	BO05070
Oli	FBgn0032651	FH3653	CG5545-RA	BO01032
opa	FBgn0003002	FH7579	CG1133-RA	BO20689
Opbp	FBgn0050443	FH5330	CG30443-RA	BO08734
Optix	FBgn0025360	FH3662	CG18455-RB	BO01044
otp	FBgn0015524	FH5787	CG10036-RE	BO20810
ovo	FBgn0003028	FH5503	CG6824-PB	BO11435
p53	FBgn0039044	FH3696	CG33336-RA	BO01104
pad	FBgn0038418	FH5304	CG10309-RA	BO05012
pan	FBgn0019664	FH3431	CG17964-RI	BO18085
Pc	FBgn0003042	FH5255	CG32443-RA	BO05228
Pcl	FBgn0003044	FH0359	CG5109-RA	BO01339
Pdp1	FBgn0016694	FH3663	CG17888-RA	BO01446
peb	FBgn0003053	FH3833	CG2668-RA	BO05809
Pep	FBgn0004401	FH3946	CG6143-RA	BO18063
pfk	FBgn0035405	FH6256	CG15812-RA	BO23535
PHDP	FBgn0025334	FH6410	CG11182-RA	BO22266
pho	FBgn0002521	FH5272	CG17743-RA	BO08708
phol	FBgn0035997	FH4681	CG3445-RA	BO01269
phtf	FBgn0028579	FH4116	CG3268-RA	BO05926
pita	FBgn0034878	FH0314	CG3941-RA	BO01592
pnr	FBgn0003117	FH5940	CG3978-RB	BO22933
pnt	FBgn0003118	FH5958	CG17077-RB	BO22343
polybromo	FBgn0039227	FH6404	CG11375-RA	BO22156
Poxm	FBgn0003129	FH6333	CG9610-RC	BO24616

Table 2.2 Continued

Poxn	FBgn0003130	FH0560	CG8246-RA	BO05174
ppl	FBgn0027945	FH0828	CG7758-RA	BO06675
pr-set7	FBgn0011474	FH0315	CG3307-RA	BO01595
pros	FBgn0004595	FH7407	CG17228-RL	BO08650
psq	FBgn0263102	FH0360	CG2368-RB	BO01341
Pur-alpha	FBgn0022361	FH0498	CG1507-RA	BO04476
pzg	FBgn0259785	FH5292	CG7752-RA	BO01627
Rack1	FBgn0020618	FH0009	CG7111-PA	BO13592
Rbf	FBgn0015799	FH5860	CG7413-RA	BO05453
rdx	FBgn0086364	FH5273	CG9924-RB	BO08612
Rel	FBgn0014018	FH4684	CG11992-RA	BO01335
repo	FBgn0011701	FH5265	CG31240-RA	BO07508
Rga	FBgn0017550	FH3951	CG2161-RA	BO18093
rib	FBgn0003254	FH5291	CG7230-RA	BO01588
rn	FBgn0259172	FH5249	CG32466-RC	BO05054
row	FBgn0033998	FH5277	CG8092-RB	BO10129
Rpb10	FBgn0039218	FH3631	CG13628-RA	BO01001
Rpb4	FBgn0053520	FH0548	CG33520-RD	BO05150
Rpb7	FBgn0051155	FH5143	CG31155-RA	BO16653
Rpb8	FBgn0037121	FH4698	CG11246-RA	BO04311
Rpd3	FBgn0015805	FH0946	CG7471-RA	BO07140
RpII140	FBgn0003276	FH6792	CG3180-RA	BO22865
RpII18	FBgn0003275	FH3097	CG1163-RA	BO15668
RpII33	FBgn0026373	FH0636	CG7885-RA	BO06090
RpIII128	FBgn0004463	FH4058	CG8344-RA	BO05464
run	FBgn0003300	FH7574	CG1849-RA	BO20687
Rx	FBgn0020617	FH5298	CG10052-RA	BO12609
sa	FBgn0002842	FH6165	CG11308-PA	BO24264
sage	FBgn0037672	FH0568	CG12952-RA	BO05086
sc	FBgn0004170	FH0554	CG3827-RA	BO05161
Sce	FBgn0003330	FH5263	CG5595-RA	BO07220
schlank	FBgn0040918	FH3731	CG3576-RB	BO04729
scro	FBgn0028993	FH6985	CG17594-RB	BO25109
sens	FBgn0002573	FH6789	CG32120-RA	BO22983
Side	FBgn0032741	FH0332	CG10446-RA	BO01176
sim	FBgn0004666	FH0363	CG7771-RA	BO03734
Sin3A	FBgn0022764	FH5929	CG8815-RB	BO21939
Sirt4	FBgn0029783	FH5102	CG3187-RC	BO14821
sisA	FBgn0003411	FH5919	CG1641-RA	BO08632
skd	FBgn0003415	FH7404	CG9936-RX	BO01331
slbo	FBgn0005638	FH0300	CG4354-RA	BO01526
slim	FBgn0026173	FH4677	CG5186-RA	BO01579

Table 2.2 Continued

slp1	FBgn0003430	FH3680	CG16738-RA	BO01064
Smox	FBgn0025800	FH5252	CG2262-RA	BO05207
sna	FBgn0003448	FH3698	CG3956-RA	BO01490
SNCF	FBgn0036349	FH5021	CG14112-RA	BO13264
Snr1	FBgn0011715	FH0002	CG1064-RA	BO01087
so	FBgn0003460	N/A	N/A	N/A
sob	FBgn0004892	FH0309	CG3242-RA	BO02272
Sox15	FBgn0005613	FH5438	CG8404-RA	BO18096
Sox21a	FBgn0036411	FH0559	CG7345-RA	BO05072
SoxN	FBgn0029123	FH5297	CG18024-RA	BO11012
Spt3	FBgn0037981	FH5699	CG3169-RA	BO18875
spt4	FBgn0028683	FH3393	CG12372-RA	BO17870
Spt5	FBgn0040273	FH3600	CG7626-RA	BO15466
srp	FBgn0003507	FH6380	CG3992-RD	BO21888
Sry-beta	FBgn0003511	FH3689	CG7938-RA	BO01480
Sry-delta	FBgn0003512	FH5280	CG17958-RA	BO10462
Ssb-c31a	FBgn0015299	FH3245	CG8396-RA	BO16765
Ssl1	FBgn0037202	FH5266	CG11115-RA	BO07542
ssp	FBgn0036248	FH1643	CG17153-RA	BO10446
Stat92E	FBgn0016917	FH5918	CG4257-RF	BO05316
stck	FBgn0020249	FH3683	CG7954-RA	BO01070
Su(H)	FBgn0004837	FH5289	CG3497-RA	BO11079
su(Hw)	FBgn0003567	FH3554	CG8573-RA	BO05465
Su(var)2-10	FBgn0003612	FH4633	CG8068-RD	BO15754
Su(var)205	FBgn0003607	FH0471	CG8409-RA	BO04357
Su(var)3-9	FBgn0003600	FH3608	CG6476-RC	BO16554
Su(z)12	FBgn0020887	FH1983	CG8013-RA	BO11056
sug	FBgn0033782	FH3694	CG3850-RA	BO01096
sv	FBgn0005561	FH0530	CG11049-RG	BO05015
svp	FBgn0003651	FH3666	CG11502-RC	BO01046
Tab2	FBgn0034431	FH0357	CG7417-RA	BO01315
Taf10	FBgn0028398	FH3642	CG2859-RA	BO01416
Taf10b	FBgn0026324	FH3639	CG3069-RA	BO01410
Taf12	FBgn0011290	FH4666	CG17358-RA	BO01024
Taf13	FBgn0032847	FH3637	CG10756-RA	BO01406
Taf2	FBgn0011836	FH3573	CG6711-RA	BO08754
Taf4	FBgn0010280	FH7572	CG5444-RC	BO20693
Taf5	FBgn0010356	FH0353	CG7704-RA	BO01277
Taf8	FBgn0022724	FH5649	CG7128-RA	N/A
tap	FBgn0015550	FH3702	CG7659-RA	BO01503
tara	FBgn0040071	FH3879	CG6889-RA	BO17757
Tbp	FBgn0003687	FH3688	CG9874-RA	BO01082

Table 2.2 Continued

term	FBgn0003683	FH3713	CG4216-RA	BO01515
Tfb1	FBgn0033929	FH5890	CG8151-RA	BO09644
Tfb2	FBgn0036513	FH0331	CG7764-RA	BO02567
TfIIA-S	FBgn0013347	FH7021	CG5163-RA	BO25301
TfIIEalpha	FBgn0015828	FH5260	CG10415-RA	BO06824
TfIIEbeta	FBgn0015829	FH4738	CG1276-RA	BO05020
TfIIFbeta	FBgn0010421	FH3665	CG6538-RA	BO01450
TfIIS	FBgn0010422	FH3676	CG3710-RA	BO01060
TH1	FBgn0010416	FH1838	CG9984-RA	BO09323
Tif-IA	FBgn0032988	FH1569	CG3278-RA	BO04451
Tip60	FBgn0026080	FH1801	CG6121-RA	BO08805
tll	FBgn0003720	FH0532	CG1378-RA	BO05022
toe	FBgn0036285	FH3568	CG10704-RA	BO08087
tou	FBgn0033636	FH1978	CG10897-RC	BO11038
toy	FBgn0019650	FH3550	CG11186-RA	BO01188
Trf	FBgn0010287	FH3651	CG7562-RA	BO01430
Trf2	FBgn0261793	FH7408	CG18009-RE	BO11453
trh	FBgn0262139	FH1740	CG9122-RA	BO07959
Trl	FBgn0013263	FH0334	CG33261-RE	BO01180
tsh	FBgn0003866	FH5962	CG1374-RA	BO22356
tth	FBgn0030502	FH4674	CG12175-RB	BO01125
ttk	FBgn0003870	FH6096	CG1856-PC	BO22972
tup	FBgn0003896	FH0327	CG10619-RB	BO01153
twi	FBgn0003900	FH5931	CG2956-RA	BO22815
Ubx	FBgn0003944	FH5274	CG10388-RE	BO08616
Usf	FBgn0029711	FH4669	CG17592-RA	BO01476
usp	FBgn0003964	FH3789	CG4380-RA	BO05420
vis	FBgn0033748	FH3709	CG8821-RA	BO01121
vri	FBgn0016076	FH5710	CG14029-RA	BO19049
vvl	FBgn0086680	FH5868	CG10037-RA	BO19909
wdn	FBgn0005642	FH5285	CG1454-RA	BO10988
wek	FBgn0001990	FH5262	CG4148-RA	BO07084
wor	FBgn0001983	FH0308	CG4158-RA	BO01566
Xbp1	FBgn0021872	FH5257	CG9415-RA	BO05246
Xpd	FBgn0015844	FH3570	CG9433-RB	BO08702
yki	FBgn0034970	FH0177	CG4005-RA	BO07155
z	FBgn0004050	FH5911	CG7803-RA	BO01216
zen	FBgn0004053	FH5925	CG1046-RA	BO22267
zf30C	FBgn0022720	FH0355	CG3998-RA	BO01303
zfh1	FBgn0004606	FH3546	CG1322-RA	BO01614

TF Specific Co-AP/MS Pipeline

Building on methods from our previous protein interaction network study (Appendix B, Guruharsha et al., 2011), we developed an experimental pipeline to specifically isolate TF protein complexes (Figure 2.1). Each expression clone was transiently transfected into *Drosophila* S2R+ cells, an embryonically derived cell line (Yanagawa et al., 1998), and nuclear extracts were generated, allowing us to address TF interactions specifically in the context of the nucleus. This additional step removes the abundant membrane and cytoplasmic proteins, increasing the sensitivity of the subsequent mass spectrometry analysis. Protein complexes were isolated using single-step affinity purification utilizing anti-HA affinity resin, fragmented with trypsin and analyzed by high-pressure liquid chromatography followed by tandem mass spectrometry (LC-MS/MS). The raw MS results were searched against the *Drosophila* genome to identify specific peptides and proteins, also providing peptide quantification via spectral counts.

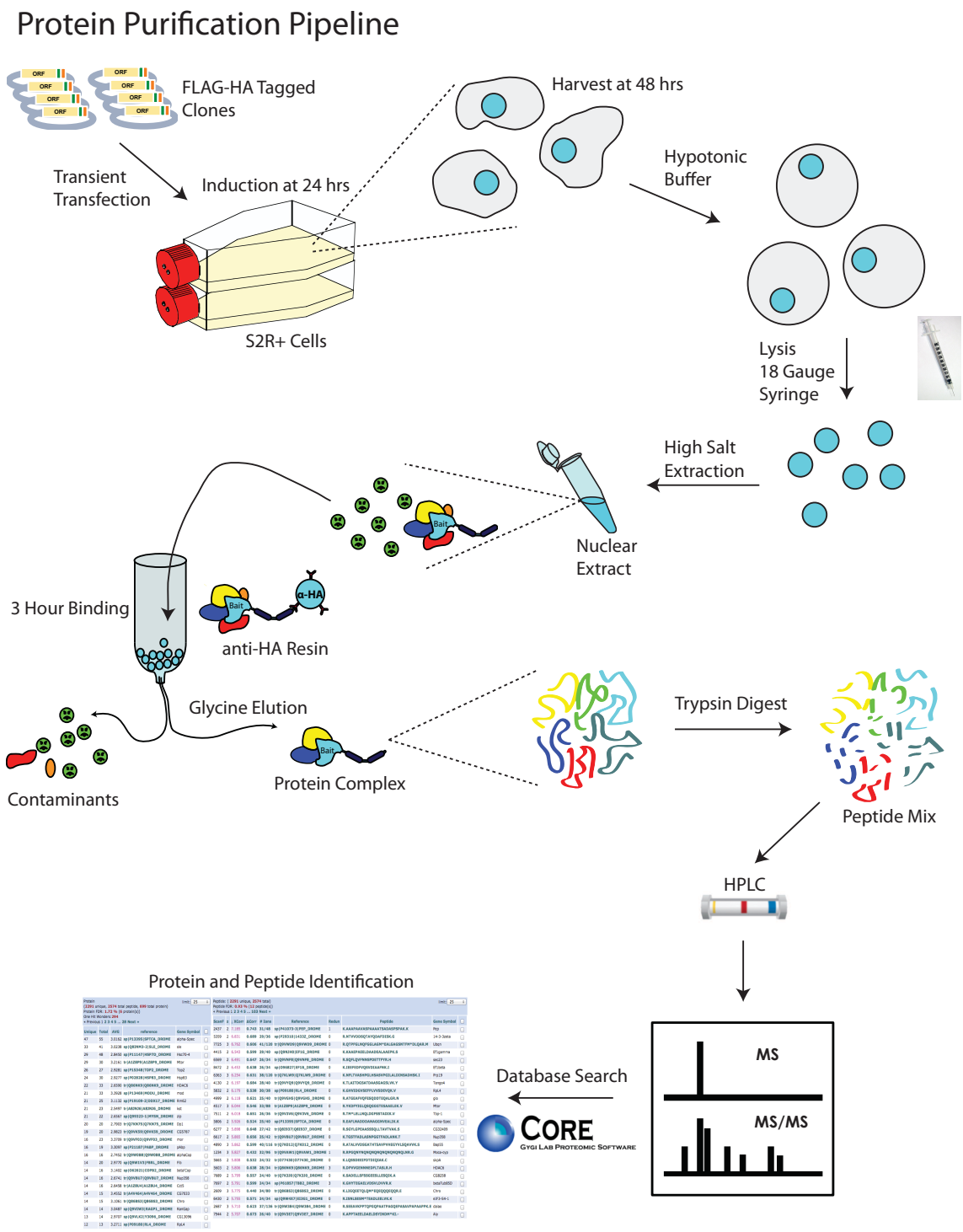

Figure 2.1: TF Protein Purification Experimental Pipeline
Experimental pipeline established for TF protein purification, diagram details the process from transfection of expression constructs through protein and peptide identification.

Approximately 80% of the transfected clones were expressed successfully, as their unique cognate peptides were detected by LC/MS/MS. A number of these individual MS experiments were removed from the subsequent analysis as we applied a manual filtering step, where either the number of total peptides in an experiment was well below or well above average, or in clear cases of contamination. Across these filtered experiments, we recovered 2,065 proteins from 468 individual purifications with a 2.27% FDR (Supplemental Table 2.1). This represents recovery of approximately 1/3rd of the S2R+ proteome, based on transcriptome and whole proteome analyses (Cherbas et al., 2011; Appendix B, Guruharsha et al., 2011). We next examined the protein functional classes of our MS data, using the PANTHER classification system (Thomas et al., 2003). This analysis demonstrated relative enrichment for both nucleic acid binding proteins and TFs, while extracellular matrix proteins, receptors and cell adhesion molecules were underrepresented in our MS results, consistent with the notion that our experimental pipeline successfully addresses TFs and related proteins (Figure 2.2). From these data, we identified 3407 binary TF-TF interactions between characterized TFs (Supplemental Table 2.2), as well as interaction data for 72 chromatin-related proteins and 327 characterized TFs.

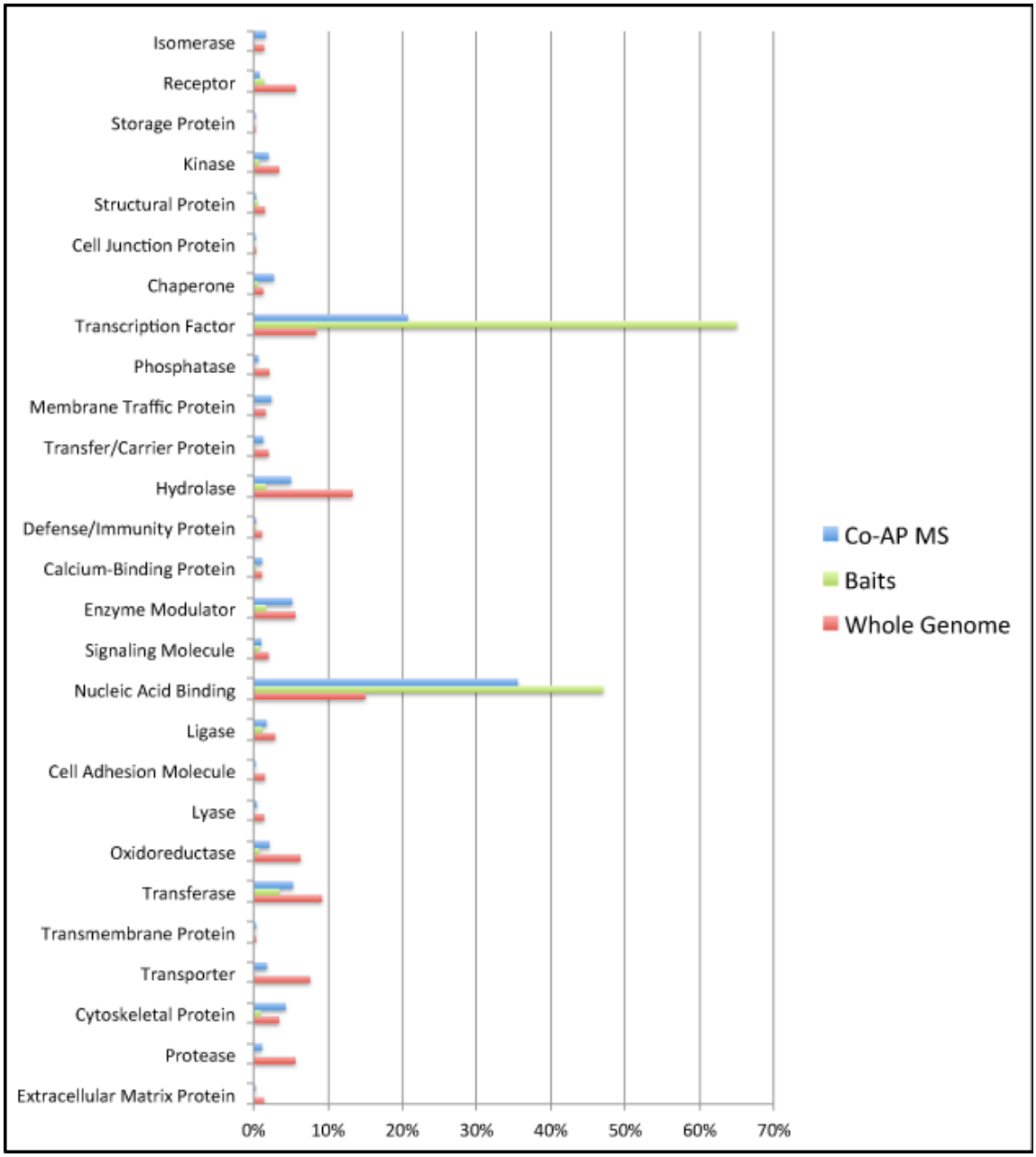

Figure 2.2: Protein Functional Class Analysis of MS Results
Analysis was performed using online resources at pantherdb.org (Thomas et al., 2003). Proteins used as bait (green) were compared to all proteins identified via MS across all experiments (blue) and the entire *Drosophila* genome (red). As expected, the proteins identified in our experiments or proteins used as bait were enriched for transcription factors and nucleic acid binding terms compared to the whole genome. Non-nuclear proteins such as extracellular matrix proteins, cell adhesion molecules and receptor proteins are underrepresented when compared to the distribution across the whole genome.

Construction of a High-confidence Interaction Network

We subsequently filtered our data using the HGScore method, which integrates quantitative data (spectral counts) into the analysis and was shown to recover more previously described interactions compared to other existing published methods and thus a higher quality interaction map (Appendix B, Guruharsha et al., 2011). Though HGScore was originally designed to include prey-prey interactions (interactions defined only by co-occurrence), we examined only bait-prey relationships to reduce network noise and to focus the network specifically on TF-protein interactions. In total, 174,561 interactions between the 2,065 identified proteins were analyzed and scored (Supplemental Table 2.3). These scored interactions were filtered to a false discovery rate (FDR) of 2%, based on the use of random datasets, leading to a high-confidence network containing 624 connections between 647 proteins, of which, 229 (35%) are characterized TFs (Figure 2.3, Supplemental Figure 2.1, and Supplemental Table 2.4). This interaction network shows a group of 406 proteins (63%) as the giant interconnected component of the network with a second group of 241 proteins in smaller, independent protein complexes. Interestingly, 39% (253) of the proteins in the high-confidence network have no previous functional annotation or are annotated only in silico (by inferred electronic annotation) thus our map provides direct physical evidence for the functions of these previously uncharacterized proteins (Marygold et al., 2013). It is important to note that we purposefully applied an extremely stringent statistical filter so as to remove false positives from our final high-confidence interaction network map. A number of previously characterized protein complexes fell below this cut-off, suggesting that there are significant data below this severe statistical limit. Though the level of noise may

increase, it may be useful to employ a more inclusive statistical cut-off when searching the network for interactions of interest.

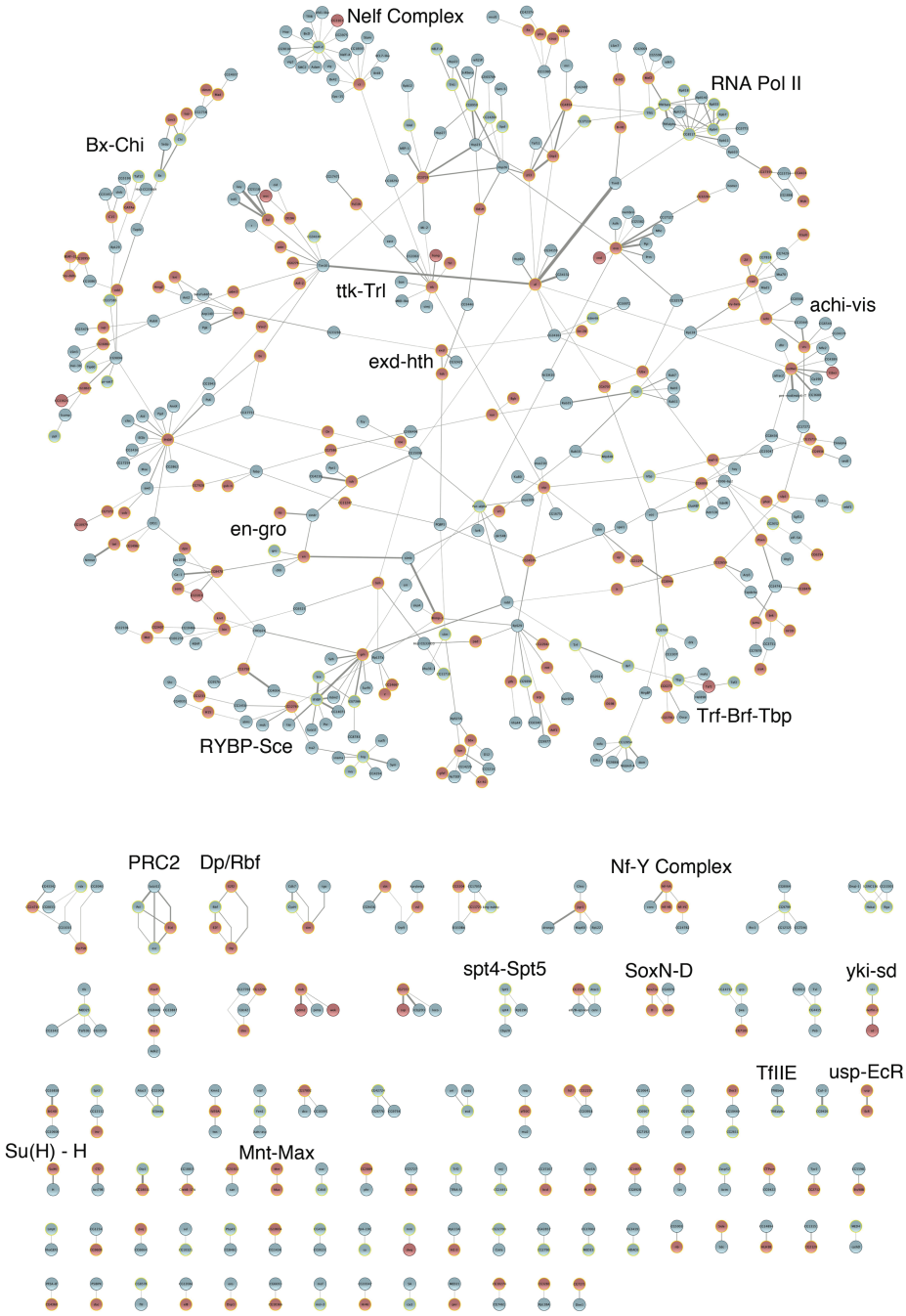

Figure 2.3 A High-Confidence TF protein-protein interaction network
High-confidence interaction network representing high-confidence interactions involving 229 characterized transcription factors (Red nodes). The network contains 647 proteins

(Figure 2.3 Continued)
connected by 624 edges. Protein interactions are shown as grey lines, with the thickness proportional to the HGScore for the interacting pair of proteins. A number of previously characterized protein complexes have been labeled.

TF Network Quality Assessment

As with previous large-scale protein interaction studies, defining a reference set of positive interactions has been difficult due to the small degree of overlap between existing data sets and the lack of a high-quality manually curated set of interactions, such as in yeast (Yu et al., 2008). We utilized the *Drosophila* Interactions Database (DroID, Murali et al., 2011), which contains protein interaction data from nine discrete sources, including recently published large-scale data sets (Friedman et al., 2011, Guruharsha et al., 2011). A direct comparison shows that 21% of edges in our high-confidence TF interaction network are present in the DroID database (Supplemental Table 2.5). Our experiments provide an additional experimental validation for these previously described interactions and validate our approach. Amongst these previously identified protein interactions, we recovered a number of well-characterized protein complexes such as the extradenticle-homothorax (exd-hth) transcriptional cofactor, Polycomb Repressive Complex 2 (PRC2), and the Dp-E2F TF (dREAM) complex, among others (Figure 2.3, Figure 2.4).

It is important to note that demonstrating the high quality of our data presents a unique challenge due to the lack of a "gold standard" reference set of PPI interactions in *Drosophila* to compare our data with, and the fact that 39% of the proteins in our network are currently unstudied. As such, we have used rigorous, established statistical methods to define interactions, leaning heavily on strict statistical cutoffs to limit the number of false-positive interactions in our high-confidence interaction network. The recovery of well-characterized protein complexes and, as I described further below, our ability to

functionally validate in vivo relationships predicted by our proteomic data, indicate the network we generate is reliable.

Recovery of Characterized TF Protein Complexes

As evidence of the quality of our PPI network analysis, we successfully recovered a number of previously identified, well-characterized protein complexes. We focus on several examples, while also highlighting the biological implications for some of our findings. The first complex, exd-hth, is a dimeric cofactor that can interact with all Hox family members (Fig 2.4A). Recent work has suggested that these interactions mediate the sequence binding specificity of the Hox genes, directly impacting TF binding as well as subsequent functions (Joshi et al., 2007, Slattery et al., 2011). In our interaction network, we recover the dimer as an interacting pair (Figure 2.4A). Interestingly, our network identifies interactions with five previously unstudied proteins (CG33260, CG34163, CG32425, CG5446, and PQBP1). One of these, CG33260, connects the exd-hth cofactor to the Hox gene, Ultrabithorax (Ubx) and the Ubx interactor, aristaless (al). Given previously characterized interactions between exd-hth and Ubx, this strongly suggests a role for CG33260 in Hox function.

We also recovered the polycomblike-polycomb repressive complex 2 (Pcl-PRC2, Fig 2.4B). The PRC2 complex trimethylates Lysine 27 of Histone 3 at Polycomb target genes, a modification that typically characterizes suppressed chromatin. The Pcl containing variant of this complex has been shown to result in high levels of H3K27 trimethylation at target genes, resulting in inhibition of target gene expression (Nekrasov et al., 2007).

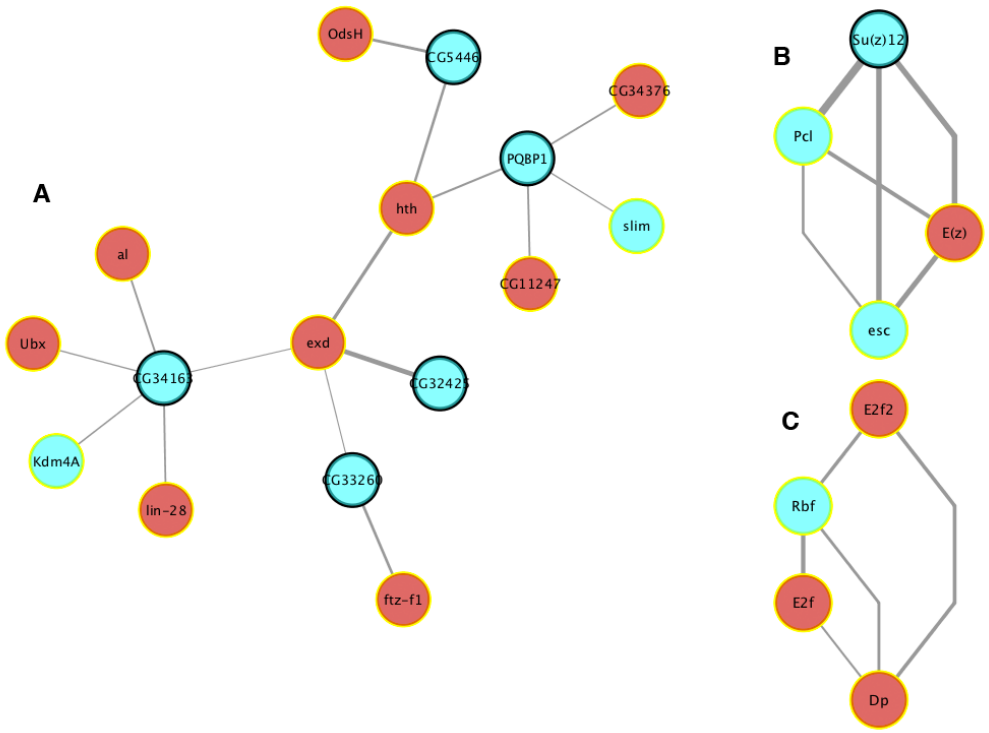

Figure 2.4 TF protein complexes
A subset of protein complexes from our high-confidence interaction network. Red nodes represent characterized transcription factors, blue nodes represent non-TF proteins. Protein interactions are connected using gray lines, with the thickness proportional to the HGScore for the pair of interacting proteins. Nodes outlined in green represent proteins used as a bait protein. (A) extradenticle-homothorax transcription cofactor. (B) Polycomblike-Polycomb Repressive Complex 2. (C) Dp-E2F dimeric transcription factor (dREAM complex).

Likewise, we recovered the dREAM complex, composed of the TF-TF dimer Dp-E2F and the TF Rbf (Figure 2.4C). dREAM is conserved in most eukaryotes and plays multiple roles including the regulation of development, cell division and apoptosis (van den Heuvel and Dyson 2008). Dp and E2f comprise a dimeric transcription factor that is important in the G1/S phase transition during the cell cycle, where E2f levels are rate-limiting for cell proliferation (Johnson et al., 1993). It has been shown previously that both E2F and E2F2 interact with DP and RBF in *Drosophila*, confirming the protein-protein interactions in our network (Fig 2.4, Frolov et al., 2001).

Functional Validation of TF Interaction Network

An essential aspect of PPI networks is their utility in predicting biological function and in generating hypotheses. We tested predictions from our interactions *in vivo*, specifically focusing our efforts on the Notch pathway, a conserved fundamental signalling mechanism broadly controlling cell fates in development (Artavanis-Tsakonas et al., 1999). In a previous report, genome-wide genetic modifier studies of a dominant-negative allele of *mastermind* (*mam*), a Notch transcriptional co-activator (Kankel et al., 2007), identified 408 genes that genetically interact with *mam*, recovering genetic modifiers in ~4% of genes screened. This particular screen utilized the Exelixis collection, a transposon-induced mutant collection with insertions in just over half of all genes in the *Drosophila* genome (Parks et a., 2004; Thibault et al., 2004).

With a simple guilt-by-association hypothesis that proteins that interact often share function, we mapped these previously identified genetic modifiers onto our interaction data and identified 88 proteins that physically interact with *mam* modifiers that had not been tested in the aforementioned genetic screen (Table 2.3). To interrogate these 88 genes functionally, we obtained transgenic RNAi alleles under UAS control for these genes and crossed them to a dominant-negative C-terminal *mastermind* truncation specifically expressed in the developing wing 1/2C96-GAL4, UAS-MamN (C96-MamN) (Helms et al., 1999, , Kankel et al., 2007, Kitagawa et al., 2001, Wu et al., 2000).

Table 2.3 Genetic Screen of Proteins that Physically Interact with Known *mastermind* Modifiers.

A table containing the 88 proteins that physically interact with previously identified *mastermind* modifiers that were tested in our genetic screen. The specific RNAi TRiP alleles used in the screen are listed, along with the phenotype seen when each was crossed to the dominant negative *mastermind* allele (C96-mamN x TRiP Stock) and for the control

(Table 2.3 Continued)
cross between the RNAi allele and the C96-Gal4 (wing) driver. For crosses that yielded an interaction, the number of flies with a modifier phenotype and percentage are listed.

Table 2.3 *mastermind* Genetic Screen

Gene Symbol	TRiP Stock ID	Bloomington ID	c96-MamN x TRiP Stock	C96-Trip Stock Alone Control	Total # Modifiers	Total # Flies	% Modifiers
ct	JF03304	29625	Enhancer	Anterior notching	**147**	**147**	**100.00**
CG10321	JF02328	26764	Suppressor	No Phenotype	**14**	**14**	**100.00**
EloA	HMS01255	37017	Suppressor	Mostly Lethal, escapers have slightly wrinkled wings	**103**	**103**	**100.00**
HLH3B	JF02098	26324	Suppressor	No Phenotype	**35**	**35**	**100.00**
TfIIS	HMS01117	34642	Suppressor	No Phenotype	**102**	**104**	**98.08**
lolal	GLV21087	35722	Suppressor	No Phenotype	**39**	**40**	**97.50**
NELF-B	HMS00165	34847	Suppressor	No Phenotype	**129**	**134**	**96.27**
Cdk12	HMS00155	34838	Enhancer	No Phenotype	**95**	**100**	**95.00**
maf-S	JF02008	25986	Suppressor	No Phenotype	**12**	**17**	**70.59**
CG13183	HMS01444	35031	Suppressor	No Phenotype	**41**	**62**	**66.13**
Ku80	JF02790	27710	Suppressor	No Phenotype	**31**	**47**	**65.96**
wdn	GLV21019	35654	Suppressor	No Phenotype	**50**	**77**	**64.94**
nub	JF02973	28338	Suppressor	No Phenotype	**20**	**31**	**64.52**
RpS9	HMS00271	33394	Suppressor	No Phenotype	**37**	**58**	**63.79**
Kdm4A	HMS01304	34629	Suppressor	No Phenotype	**32**	**55**	**58.18**
Poxn	JF02136	26238	Suppressor	No Phenotype	**71**	**128**	**55.47**
geminin	GLV21038	35673	Suppressor	No Phenotype	**48**	**88**	**54.55**
nerfin-1	JF02956	28324	Suppressor	No Phenotype	**53**	**120**	**44.17**
Cyp1	HMS00902	33950	Suppressor	No Phenotype	**44**	**109**	**40.37**
Sgf29	GL00597	36637	Suppressor	No Phenotype	**51**	**133**	**38.35**
Arp66B	HMS00711	32921	Suppressor	No Phenotype	**41**	**111**	**36.94**
C15	JF02824	27649	Suppressor	No Phenotype	**13**	**46**	**28.26**
CG6364	GL00263	35351	Suppressor	No Phenotype	**18**	**68**	**26.47**
gfzf	HMS00394	32399	Enhancer	No Phenotype	**16**	**66**	**24.24**
SCAR	HMS01536	36121	Enhancer	No Phenotype	**27**	**113**	**23.89**
CG1218	HMS01383	34389	Enhancer	No Phenotype	**28**	**123**	**22.76**
bin	HMS01197	34718	Suppressor	No Phenotype	**27**	**126**	**21.43**
CG3226	HMS00662	32875	Enhancer	No Phenotype	**19**	**101**	**18.81**
ATPsyn-gamma	JF03150	28723	Additive phenotype	Notching			
CG7839	JF02014	25992	Additive Phenotype	Low Penetrance Anterior Notch			
Orc4	HMS00404	32409	Additive Phenotype	Loss of Bristles			
gro	HMS01506	35759	gro phenotype dominant	Rounded, Cupped Wings			
CG2469	HMS00619	33736	Lethal	Lethal			
l(2)NC136	HMS00802	33002	Lethal	Lethal			
MED21	HMS01211	34731	Lethal	Lethal			
RpS30	HMS00636	32851	Lethal	Lethal			
bcd	HM05074	28586	No Change	No Phenotype			
BCL7-like	GLV21079	35714	No Change	No Phenotype			
Cdk7	GL00073	35199	No Change	No Phenotype			
CG10274	JF02137	26239	No Change	No Phenotype			
CG11180	JF03044	28629	No Change	No Phenotype			
CG11448	JF02940	28309	No Change	No Phenotype			
CG11999	HMS00565	34604	No Change	No Phenotype			
CG12219	JF02834	28000	No Change	No Phenotype			
CG14657	GL00519	36862	No Change	No Phenotype			
CG15710	JF02337	26773	No Change	No Phenotype			
CG1832	JF02426	27080	No Change	No Phenotype			
CG2021	HM05066	28579	No Change	No Phenotype			
CG32486	GL00391	35465	No Change	No Phenotype			
CG42358	HMS00438	32440	No Change	No Phenotype			
CG4415	GL00456	35613	No Change	No Phenotype			
CG4747	HMS00568	33696	No Change	No Phenotype			
CG5343	GL00600	36640	No Change	No Phenotype			
CG8243	HMS00876	33927	No Change	No Phenotype			
CG9344	HM05211	29532	No Change	No Phenotype			
CG9588	HM05013	28527	No Change	No Phenotype			
CycH	HMS01212	34732	No Change	No Phenotype			
D19A	HMS00244	33371	No Change	No Phenotype			
Doc3	JF02223	31932	No Change	No Phenotype			
Elp3	GL00417	35488	No Change	No Phenotype			
fd59A	JF02228	31937	No Change	No Phenotype			
Fer3	JF01996	25974	No Change	No Phenotype			
hbn	JF02195	31906	No Change	No Phenotype			
His2Av	HM05177	28966	No Change	No Phenotype			
HLH54F	JF03114	28698	No Change	No Phenotype			
hpo	JF02740	27661	No Change	No Phenotype			
kuk	JF03178	28750	No Change	No Phenotype			
l(3)73Ah	HM05190	28979	No Change	No Phenotype			
lid	JF02683	28944	No Change	No Phenotype			
Mad1	GLV21088	35723	No Change	No Phenotype			
mago	HM05142	28931	No Change	No Phenotype			
Map60	HMS00457	32458	No Change	No Phenotype			
Med	JF02218	31928	No Change	No Phenotype			
Met	JF02103	26205	No Change	No Phenotype			
Myb	JF02135	26237	No Change	No Phenotype			
Nup153	HM05248	30504	No Change	No Phenotype			
Nup75	JF02946	28315	No Change	No Phenotype			
Pgk	JF02889	28053	No Change	No Phenotype			
Pros28.1	GL00341	36063	No Change	No Phenotype			
Rab10	JF02058	26289	No Change	No Phenotype			
RpS27	HMS01581	36692	No Change	No Phenotype			
run	JF03088	28673	No Change	No Phenotype			
sbr	HM05135	28924	No Change	No Phenotype			
Sce	GL00371	35446	No Change	No Phenotype			
Smox	JF02320	26756	No Change	No Phenotype			
snf	HMS01067	34593	No Change	No Phenotype			
toe	JF02507	29345	No Change	No Phenotype			
Zif	JF02524	29360	No Change	No Phenotype			

The C96-MamN fly exhibits a wing nicking phenotype, similar to phenotypes seen with the loss of function of other Notch pathway components (Figure 2.5B). We screened the RNAi x C96-MamN crosses for modifiers of this wing phenotype, identifying both enhancers and suppressors (Figure 2.5). From the 88 crosses tested, we recovered genetic modifiers in 35% of our crosses (Table 2.3), representing a seven-fold increase when compared to the 4% recovered in the previously reported genome-wide unbiased screen, demonstrating clear predictive power for our protein-protein interaction data.

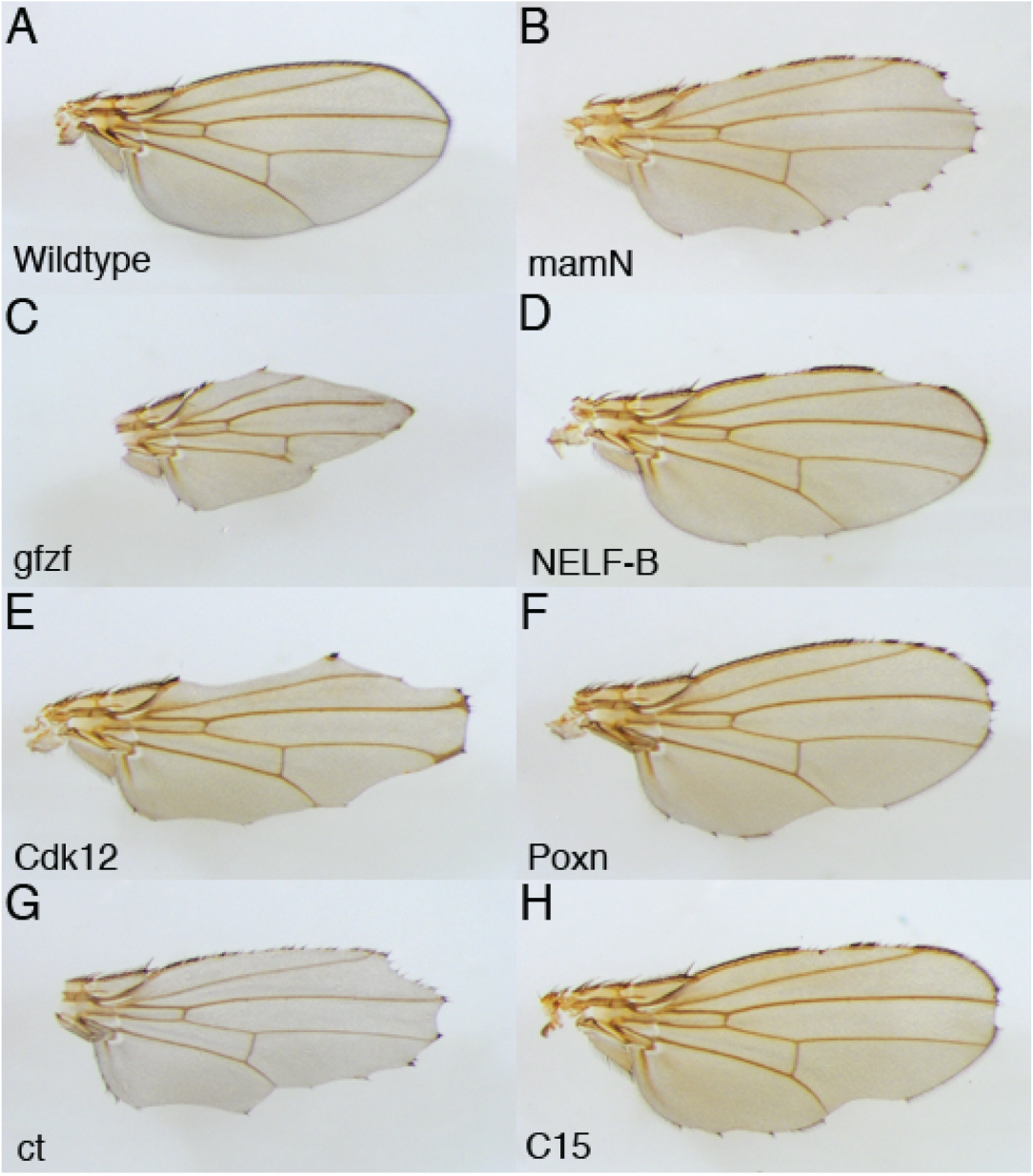

Figure 2.5 *mastermind* Genetic Screen Phenotypes
Examples of phenotypes identified in the genetic screen for modifiers of the *mastermind* phenotype. (A) Wild type *Drosophila* wing. (B) Dominant-negative *mastermind* (C96-mamN) phenotype. (C, E, G) Enhancer phenotypes seen when RNAi alleles of gfzf, Cdk12 and ct are crossed to C96-mamN. (D, F, H) Suppressor phenotypes seen when NELF-B, Poxn, and C15 are crossed to the C96-mamN.

One of the biggest challenges with interpreting genetic screens is in understanding how various genes that modify the same pathway are related to one another at a mechanistic level. As an example of the utility of our protein interaction data for this purpose, we found that five previously characterized modifiers of the *mam* phenotype: simj, Lim1, CG11334, fd68A and CG34417 — though previously unlinked to one another in the literature (Figure 2.6A), physically interact with cut (ct), a transcriptional target of the Notch pathway (Figure 2.6B). ct itself is a TF that was also demonstrated to genetically interact with *mam* in our genetic screen (Figure 2.5G). As three of the interacting proteins are TFs (the other two are unstudied), this strongly suggests their functional connection to the Notch signaling pathway may be mediated through regulation of transcription via TF-TF interactions with ct.

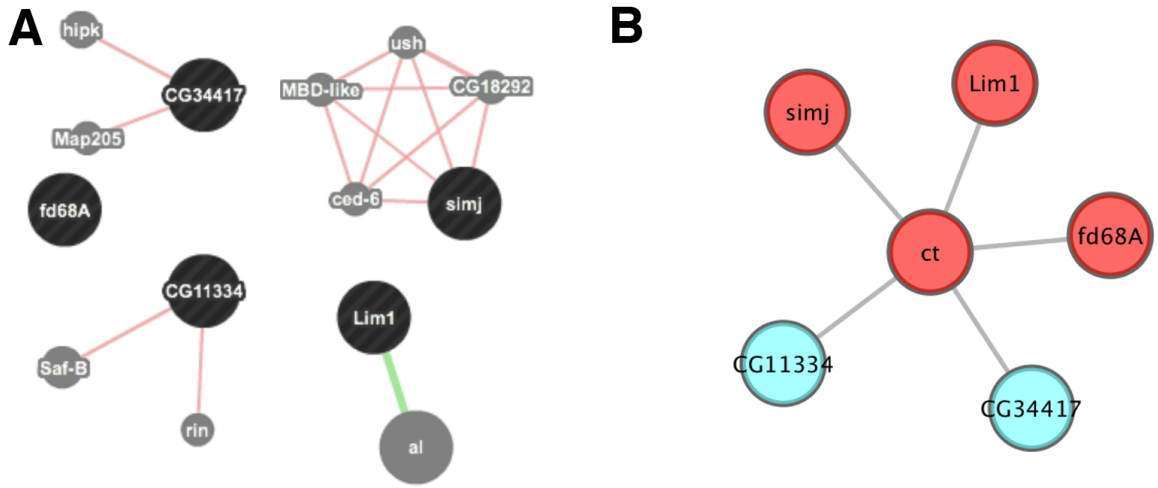

Figure 2.6 *mastermind* Modifier Protein Interactions
(A) Protein interactions for five previously identified *mastermind* genetic interactors from the GeneMANIA database (Warde-Farley et al., 2010). Though all five genetically interact with *mastermind*, they share no physical connections in the literature. (B) Protein interactions identified in our TF interaction data. All five of the previously identified *mam* modifiers interact with cut, a known target of Notch signaling. Red nodes are TF, while blue nodes are non-TF proteins.

Discussion

We present here a protein-protein interaction study of TFs in *Drosophila melanogaster*, defining interactions for nearly half of the characterized TFs in the species. As defining these interactions is essential to understanding TF function, we expect this body of work to be a valuable resource for probing the mechanisms of differential gene expression, relevant to the vast majority of biological processes. A considerable fraction of our interaction results are novel, which demonstrate biological hypotheses to be tested. The predictive value of our network is indicated by the recovery of known interactions, as well as through functional validation *in vivo* of interactions indicated by the network.

We acknowledge several limitations in our methods, in particular, the use of epitope-tagged proteins expressed at non-physiological levels. While we cannot ignore that epitope tags in some cases will perturb protein folding and function, the recovery of previously characterized protein complexes, including those identified via alternative methods such as two-hybrid screening, provide additional evidence of the validity of our experimental pipeline. Furthermore, similar methods have been used successfully to identify confirmed interactions in a number of settings, including the human autophagy system and in a proteome-wide analysis in *Drosophila* (Behrends et al., 2010; Appendix B, Guruharsha et al., 2011; Sowa et al., 2009). Ultimately, the protein-protein interactions defined in this body of work represent a starting point for further inquiry and will need to be validated through other additional experimental means.

TF protein interactions represent an essential component to understanding the combinatorial regulation of gene expression by TFs. Nevertheless, physical interactions characterize just one parameter of TF biology. In the following chapter, I focus on integrating these data with other data types to provide insight into tissue-specific

regulation by TFs, shared physical targets of interacting factors and ultimately gene regulatory network models that allow us to further probe the regulatory mechanisms within functional networks.

Materials and Methods

Protein Expression and Purification

C-terminal FLAG-HA tagged transcription factor clones in the pMK33-CFH-BD vector were acquired from the Berkeley *Drosophila* Genome Project (Yu et al., 2011). Each clone was transiently transfected into two 54 ml cultures of *Drosophila* S2R+ cells using Effectene (Qiagen), and subsequently cultured in Schneider's media with 10% Fetal Bovine Serum. 24 hours post-transfection, gene expression was induced with 0.35 mM $CuSO_4$ and cells were harvested 24 hours after induction (Veraksa et al., 2005). Nuclear extracts were prepared as previously described with the exception that cells were lysed using an 18-gauge syringe (Dignam et al. 1983). Nuclear extracts were diluted 1:1 with dialysis buffer (20 mM HEPES pH 7.6, 20% glycerol, 100 mM KCl, 2mM $MgCl_2$, 0.1 mM EDTA, 1mM DTT, 0.25mM PMSF, and Roche mini complete protease inhibitor) to reduce the overall salt concentration. Each extract was incubated with 40 uL of dimethyl pimelimidate cross-linked HA immunoaffinity resin (Sigma) for three hours at 4°. Following incubation, the resin was washed 2x with dialysis buffer followed by 2x PBS washes. Bound proteins were eluted using IgG Elution Buffer (Thermo Scientific), 400 uL total divided into two separate five minute incubations performed at room temperature with gentle shaking. The elution was then neutralized with 52 uL 1M Tris pH 8.0.

Mass Spectrometry

Co-purified proteins were subsequently precipitated with Tricholoroacetic acid (TCA), followed by a 10% TCA wash and two acetone washes. The samples were then dried, digested overnight with trypsin, cleaned with c18 Stage Tips (Thermo Scientific), and analyzed by LC-MS/MS on a linear trap quadrupole (Thermo Scientific) instrument. MS/MS spectra were searched with SEQUEST (Eng et al., 2008) against FlyBase release 5.41 and filtered to 2.27% protein FDR for the entire data set with the reverse database approach (Elias and Gygi, 2007). Column carry-over between experiments was corrected with a statistical approach, incorporating peptide abundance and probability of consecutive observations.

Network Construction

Following processing and filtering, the high-confidence TF interaction map was generated using the HGSCore method to distinguish specific interactions as described previously, but filtering out indirect prey-prey interactions to focus the network on the TF-interacting subspace. To draw the cut-off for interaction specificity and determine false discovery rate, we ran HGScore on 40 simulated datasets, randomly sampled from the real dataset until convergence on a cut-off score resulting in a 2% FDR.

Genetic Screen

Flies were cultured on standard media and crosses were carried out at 23°. The

C96-Gal4, UASMamN (C96-MamN) stocks were previously described (Helms et al., 1999). UAS-RNAi fly stocks were obtained from the TRiP collection at Harvard Medical School (NIH/NIGMRS R01-GM084947). Adult fly wings were dehydrated in isopropanol and mounted in a 3:1 dilution of CMCP-10 (Masters Company Inc, Wood Dale, IL) and lactic acid.

Bibliography

Adryan, B., and Teichmann, S.A. (2006). FlyTF: a systematic review of site-specific transcription factors in the fruit fly Drosophila melanogaster. Bioinformatics 22, 1532-1533.

Adryan, B., and Teichmann, S.A. (2007). Computational identification of site-specific transcription factors in Drosophila. Fly 1, 142-145.

Adryan, B., and Teichmann, S.A. (2010). The developmental expression dynamics of Drosophila melanogaster transcription factors. Genome biology 11, R40.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signalling: cell fate control and signal integration in development. Science 284, 770-776.

Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the human autophagy system. Nature 466, 68-76.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., et al. (2011). The transcriptional diversity of 25 Drosophila cell lines. Genome research 21, 301-314.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes & development 22, 746-755.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic acids research 11, 1475-1489.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature methods 4, 207-214.

Eng, J.K., Fischer, B., Grossmann, J., and Maccoss, M.J. (2008). A fast SEQUEST cross correlation algorithm. Journal of proteome research 7, 4598-4602.

Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., et al. (2005). Protein interaction mapping: a Drosophila case study. Genome research 15, 376-384.

Friedman, A.A., Tucker, G., Singh, R., Yan, D., Vinayagam, A., Hu, Y., Binari, R., Hong, P., Sun, X., Porto, M., et al. (2011). Proteomic and functional genomic landscape of receptor tyrosine kinase and ras to extracellular signal-regulated kinase signaling. Science signaling 4, rs10.

Frolov, M.V., Huen, D.S., Stevaux, O., Dimova, D., Balczarek-Strang, K., Elsdon, M., and Dyson, N.J. (2001). Functional antagonism between E2F family members. Genes & development 15, 2146-2160.

Guruharsha, K.G., Rual, J.F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O., et al. (2011). A protein complex network of Drosophila melanogaster. Cell 147, 690-703.

Helms, W., Lee, H., Ammerman, M., Parks, A.L., Muskavitch, M.A., and Yedvobnick, B. (1999). Engineered truncations in the Drosophila mastermind protein disrupt Notch pathway function. Developmental biology 215, 358-374.

Johnson, D.G., Schwarz, J.K., Cress, W.D., and Nevins, J.R. (1993). Expression of transcription factor E2F1 induces quiescent cells to enter S phase. Nature 365, 349-352.

Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell 131, 530-543.

Kankel, M.W., Hurlbut, G.D., Upadhyay, G., Yajnik, V., Yedvobnick, B., and Artavanis-Tsakonas, S. (2007). Investigating the genetic circuitry of mastermind in Drosophila, a notch signal effector. Genetics 177, 2493-2505.

Kitagawa, M., Oyama, T., Kawashima, T., Yedvobnick, B., Kumar, A., Matsuno, K., and Harigaya, K. (2001). A human protein with sequence similarity to Drosophila mastermind coordinates the nuclear form of notch and a CSL protein to build a transcriptional activator complex on target promoters. Molecular and cellular biology 21, 4337-4346.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125, 301-313.

Marygold, S.J., Leyland, P.C., Seal, R.L., Goodman, J.L., Thurmond, J., Strelets, V.B., and Wilson, R.J. (2013). FlyBase: improvements to the bibliography. Nucleic acids research 41, D751-757.

Mathur, D., Danford, T.W., Boyer, L.A., Young, R.A., Gifford, D.K., and Jaenisch, R. (2008). Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET. Genome biology 9, R126.

Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G., 3rd, and Finley, R.L., Jr. (2011). DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. Nucleic acids research 39, D736-743.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Muller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. The EMBO journal 26, 4078-4088.

Parks, A.L., Cook, K.R., Belvin, M., Dompe, N.A., Fawcett, R., Huppert, K., Tan, L.R., Winter, C.G., Bogart, K.P., Deal, J.E., et al. (2004). Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. Nature genetics 36, 288-292.

Pfreundt, U., James, D.P., Tweedie, S., Wilson, D., Teichmann, S.A., and Adryan, B. (2010). FlyTF: improved annotation and enhanced functionality of the Drosophila transcription factor database. Nucleic acids research 38, D443-447.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330, 1787-1797.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., et al. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell 147, 1270-1282.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell 138, 389-403.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics 13, 613-626.

Suzuki, H., Forrest, A.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J., et al. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nature genetics 41, 553-562.

Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H.L., et al. (2004). A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nature genetics 36, 283-287.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome research 13, 2129-2141.

van den Heuvel, S., and Dyson, N.J. (2008). Conserved functions of the pRB and E2F families. Nature reviews Molecular cell biology 9, 713-724.

Veraksa, A., Bauer, A., and Artavanis-Tsakonas, S. (2005). Analyzing protein complexes in Drosophila with tandem affinity purification-mass spectrometry. Developmental dynamics : an official publication of the American Association of Anatomists 232, 827-834.

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research 38, W214-220.

Wu, L., Aster, J.C., Blacklow, S.C., Lake, R., Artavanis-Tsakonas, S., and Griffin, J.D. (2000). MAML1, a human homologue of Drosophila mastermind, is a transcriptional co-activator for NOTCH receptors. Nature genetics 26, 484-489.

Yanagawa, S., Lee, J.S., and Ishimoto, A. (1998). Identification and characterization of a novel line of Drosophila Schneider S2 cells that respond to wingless signaling. The Journal of biological chemistry 273, 32353-32359.

Yu, C., Wan, K.H., Hammonds, A.S., Stapleton, M., Carlson, J.W., and Celniker, S.E. (2011). Development of expression-ready constructs for generation of proteomic libraries. Methods Mol Biol 723, 257-272.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. Science 322, 104-110.

Chapter 3

Integration of the Drosophila TF Interaction Network

Attributions

I carried out all analyses with the following exceptions: Dong-Yeon Cho[1] executed the TSPS algorithm, outlier tissue specificity scoring, and wrote and executed the script to identify TF regulatory motifs. Lijia Ma[2] and Matt Slattery[2] generated the chromatin immunoprecipitation datasets.

1. Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD USA 20894

2. Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL 60637, USA

Summary

As TF protein interactions define just one component of TF biology, we integrated our PPI data with a number of existing datasets to better define the contribution of TF protein complexes to the biology of the cell. These analyses build on the recent availability of genome wide datasets from the modENCODE project and others, and include gene expression studies, TF occupancy studies, as well as inferred regulatory network models. These datasets, respectively, allow us to address the importance of TFs in tissue specificity, to identify shared physical targets of interacting TFs, and to connect TF protein complexes to the gene regulatory networks in a cell. We classify proteins in our interaction network into bins based on tissue specificity and construct 24 tissue-specific interaction networks, outlining likely interactions within specific contexts. We then examined physical TF targets from the perspective of the protein complex, identifying likely targets of combinatorial regulation and lastly, constructed integrated networks, combining regulatory edges with our PPI data to examine regulatory connections from the viewpoint of protein complexes. We ultimately use each of these analyses to form testable hypotheses, which highlight the ultimate goal of this work, to provide a resource for the community as a whole to examine the biology of TFs and their role in regulating gene expression.

Introduction

In the first part of this work, we defined a protein interaction network of *Drosophila* transcription factors. While TF protein interactions constitute a central component to defining TF function, a number of other TF parameters must be incorporated in order to gain a more complete view of TF biology. These include examining the spatio-temporal expression of proteins in our interaction network, classifying TF targets as well as defining the regulatory relationships between TF protein complexes and their target genes.

TFs are often discussed in the context of conferring tissue specificity as they are frequently expressed within narrow domains and play a central role in developmental processes such as cell fate specification. In general terms, TFs fall into two broad groups, "general" factors that are broadly expressed and "specific" factors that exhibit restricted domains of expression (Ravasi et al., 2010). The underlying reasoning is that general factors enable transcription across many tissues, while specific factors are important in regulating tissue-specific gene programs, as is the case with "master regulator" genes in tissue specification. Proteins in these two categories frequently interact with one another in overlapping domains, altering TF function and resulting in even more specific activity. To address this fundamental component of TF biology, we utilized large-scale tissue expression datasets from the modENCODE project to score proteins in our PPI network, defining bins of TF expression specificity. We then examined the group of proteins exhibiting high tissue specificity, assigning these proteins to individual tissues, and ultimately used this analysis as the basis for constructing 24 tissue-specific interaction networks. These networks define likely interactions between proteins within specific contexts, providing a framework for understanding tissue-specific gene regulation and also provide insight into the interactions between general and specific factors.

A second crucial component to understanding TF function is defining the targets of TFs. TF target identification is frequently divided into two categories, physical and functional (Walhout 2006). Physical targets represent DNA sequences that are bound directly by TFs, often identified through chromatin-immunoprecipitation strategies or through yeast one-hybrid methods (Bulyk and Walhout 2012). Functional targets are identified through expression studies, where gene expression changes are characterized following perturbation of a particular TF (Capaldi et al., 2008). The disparity between these two categories is reflected in the fact that only 10-25% of defined physical targets, in higher eukaryotes, result in functional changes upon disruption (Spitz and Furlong 2012). An important component to these findings is that TF occupancy studies often examine each factor individually. As the bulk of TFs function through interactions with other proteins, it is expected that combinatorial interactions are playing an important role in regulating functional output. For instance, an interaction between a TF and a transcriptional suppressor or the lack of a co-factor could explain the discrepancy between a defined physical target and subsequent function. Thus examining TF targets from the perspective of the TF protein complex would provide significant insight into the activity of physical TF targets. Taking this into consideration, we combined multiple TF occupancy datasets from modENCODE and the Berkeley Drosophila Transcription Network Project (MacArthur et al., 2009, Roy et al., 2010) to define common physical targets between interacting TFs, defining potential targets for combinatorial regulation and to gain insight into potential mechanisms of regulation at these target genes.

Lastly, TF regulatory connections are often captured from the perspective of gene regulatory networks, where edges are represented not by physical interactions, but by regulatory relationships between proteins. The most extensive of these GRNs have been

constructed using learned regulatory network inference methods, incorporating many data types to construct large regulatory networks (Marbach et al., 2012). While these networks are broad in nature, they do not incorporate PPI data, thus limiting the scope to regulatory relationships from the perspective of the individual TF. Thus, we integrated our protein interaction network with such regulatory networks to examine TF regulatory relationships from the perspective of the protein complex. This integrated network analysis allows us to probe functional networks such as genetic screens, which we demonstrate by connecting genetic modifiers identified in the aforementioned, genome-wide screen for *mastermind* (Kankel et al., 2007). As regulatory programs are often conserved across species (Erwin and Davidson 2009), these analyses provide a universal framework from which to interrogate the biology of TFs and their targets.

Results

Tissue-Specific Interaction Networks

As a general rule, we expect that proteins that interact are expressed in the same place at the same time. To examine co-expression and tissue specificity of our interaction network, we utilized RNA-seq data from the modENCODE project spanning 29 tissues and developmental time points (Smibert et al., 2012). While TFs are often discussed in the context of conferring tissue specificity, a significant proportion of *Drosophila* TFs are expressed ubiquitously at some point during embryonic development and most exhibit a broad pattern of expression in the adult animal (Adryan and Teichmann 2010). TFs that show tissue specificity embryonically are usually not limited to a single tissue, but rather a narrow range of expression in several tissues. These findings suggest that it is not only the

presence of a specific TF that defines a particular tissue, but also the interactions of these TFs that establish tissue identity.

All proteins in our network were scored using tissue specificity score (TSPS, Ravasi et al., 2010). This particular method utilizes relative entropy to measure how the observed expression of a gene diverges from a distribution where a gene is uniformly expressed across all tissues. The distribution of TSPS scored proteins revealed three categories of expression, one representing broad or "general" expression across tissues, a group with high or "specific" tissue specificity, and a middle group exhibiting expression across several tissues (Figure 3.1, Supplemental Table 3.1).

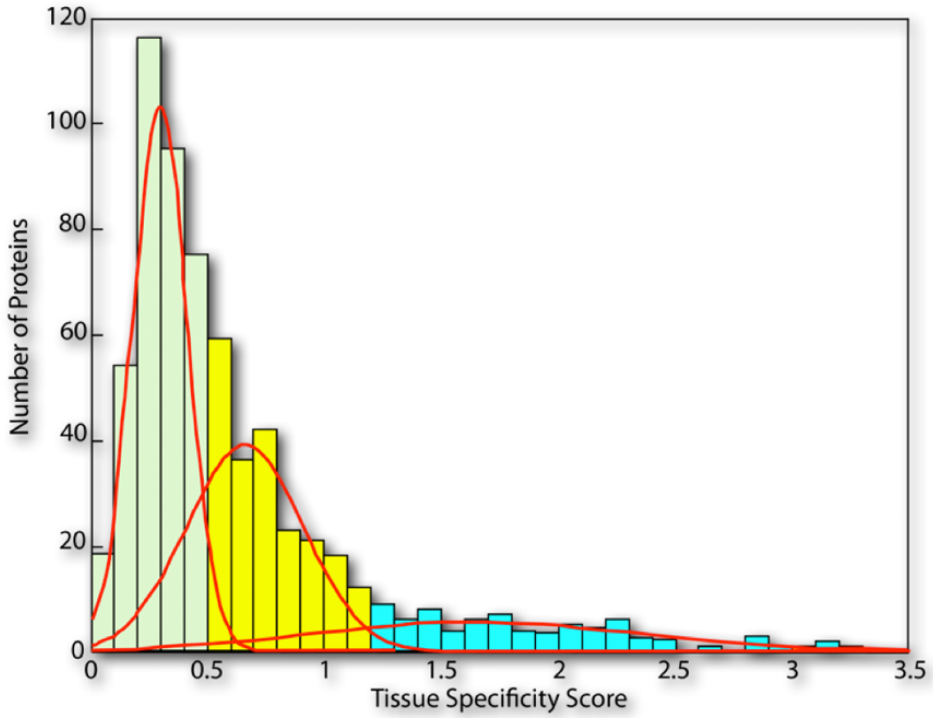

Figure 3.1 Distribution of TSPS Scored Proteins in TF Interaction Network
The distribution of TSPS for all proteins in our TF PPI network. Green bars represent the "specific" proteins, yellow represents moderate specificity and blue represents "general" or broad specificity. The distribution was fit to a trimodal Gaussian distribution to define the three separate groups.

Low TSPS proteins, representing broad expression, were assembled into a "core" network of 128 interactions which, based on their broad expression, are likely to be present across most tissues (Supplemental Figure 3.1). We then focused on the group of high scoring TSPS proteins, utilizing an outlier method (Kadota et al., 2003) to assign each protein to specific tissues (Supplemental Table 3.2). We combined these high-specificity proteins with our "core" network to build 24 different tissue-specific interaction networks (Figure 3.2, Supplemental Table 3.3, Supplemental Figure 3.1)

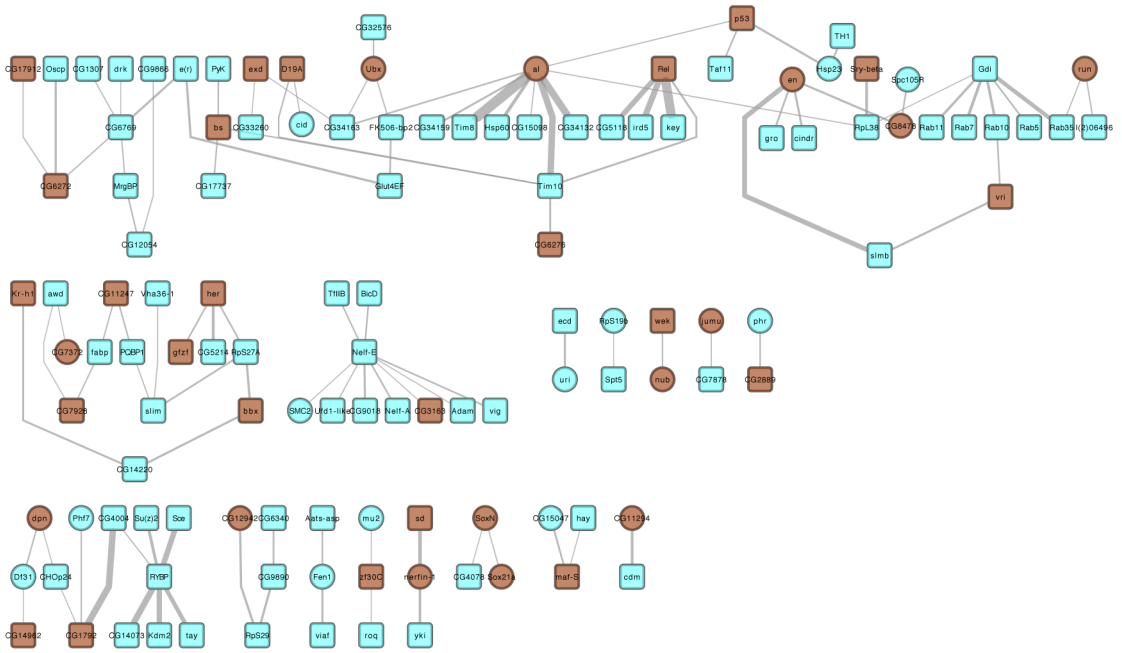

Figure 3.2: Third Instar Larval CNS-Specific TF Interaction Network
An example network from our tissue specific network analysis. "Specific" proteins are represented by circular nodes. "General" (low-specificity) proteins are represented by square nodes. Red nodes represent TFs. Protein-protein interactions are represented by gray edges, with the thickness relative to the HGScore.

Two very different protein complexes are illustrative of the value of this tissue-specificity analysis, one specific to the testis and another to the larval central nervous system (Figure 3.3). The first complex is centered on an unnamed protein CG8117, which according to our results is a part of the RNA polymerase II complex, connected to established RNA Polymerase II complex members through 8 physical edges (Figure 3.3A). CG8117 is electronically inferred to have transcription regulatory activity and to bind both zinc ions and nucleic acids. It is expressed at high levels in the adult testis, but is largely absent from other tissues. Outside of large-scale screens, CG8117 has not been independently studied in *Drosophila*. However, the human ortholog of this protein, TCEA2, has been characterized to be a testis-specific transcription factor (Weaver and Kane 1997), suggesting that this gene could play a similar tissue-specific role in *Drosophila*.

The second protein complex links two TFs, nervous fingers 1 (nerfin-1) and scalloped (sd) to the transcriptional co-activator yorkie (yki) (Figure 3.3B). sd is expressed in the developing nervous system, where it is essential for development of the sensory organs (Campbell et al., 1992). nerfin-1 has been shown to be important for axon guidance during early CNS development (Kuzin et al., 2005). yki is the *Drosophila* ortholog of the human protein YAP and is a transcriptional co-activator that functions in the hippo-yap pathway. yki and sd have been shown previously to interact (Goulev et al., 2008). It has also been suggested that Nerfin-1 is a binding partner of sd (Garg et al., 2007). Both Nerfin-1 and sd are expressed in a highly specific manner in the larval CNS and given their established importance in CNS development; their interaction suggests that they work together to regulate larval CNS development, possibly in tandem with the co-activator yki.

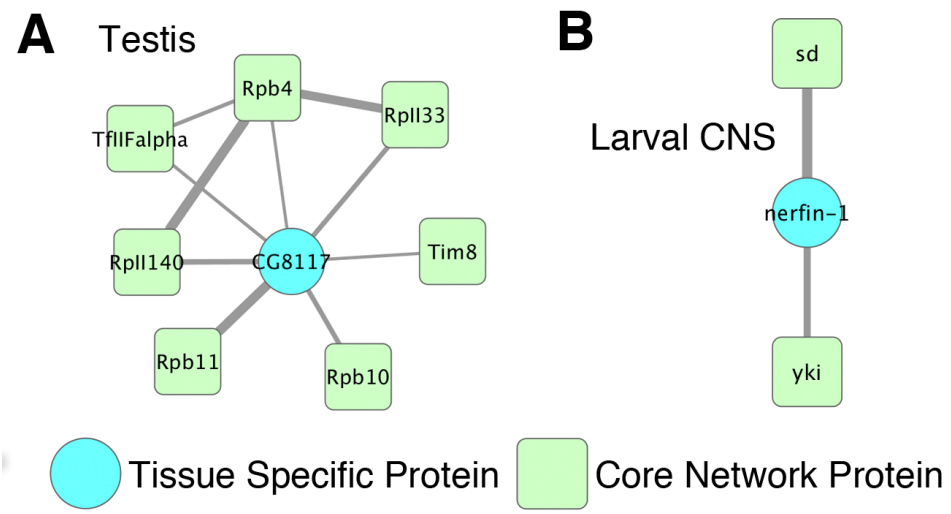

Figure 3.3 Tissue Specific Protein Complexes
Testis-specific and larval CNS-specific protein complexes. Circular nodes represent tissue specific proteins, while square nodes represent core network or "general" proteins. Gray edges represent PPI, with the width proportional to the HGScore. (A) A testis specific RNA Polymerase II protein complex. (B) A larval CNS-specific transcriptional complex containing the TFs scalloped (sd) and nerfin-1 (nervous fingers 1), along with the transcriptional co-activator yorkie (yki).

Combinatorial Targets of Interacting Transcription Factors

Given the importance of combinatorial TF interactions in gene regulation, we next compared our protein-protein interaction data with *in vivo* DNA binding data for all TF-TF pairs for which genome-wide ChIP data was available, defining shared targets for 10 TF pairs (Supplemental Table 3.4). For this analysis, we utilized TF occupancy datasets from the Berkeley *Drosophila* Transcription Network Project (MacArthur et al., 2009) and from modENCODE (Roy et al., 2010). We identified multiple pairs where the protein-protein interactions and DNA co-binding are consistent with the existing literature. For example, we observed an interaction between ecdysone receptor (EcR) and ultraspiracle (usp), which are the two proteins that comprise the complete ecdysone receptor; upon ligand binding, EcR-Usp are activated and coordinately regulate genes

including Eip75B and DHR3 (Yao et al., 1993; Figure 3.4A). We also recovered an interaction between polycomblike (pcl) and enhancer of zeste [E(z)], two proteins that are members of the polycomblike-polycomb repressive complex 2 (Pcl-PRC2; Figure 3.4B), as well an interaction between the segment polarity gene engrailed (en) and the co-repressor groucho (gro) (Figure 3.4C) (Hittinger and Carroll 2008).

Beyond these characterized interactions, we found several examples of less characterized protein-protein interactions that are supported by TF-TF co-localization on DNA. For instance, we observed an interaction between tramtrack (ttk) and Trithorax-like (Trl) (Figure 3.4D). Both are BTB/POZ (Br-C, ttk and bab/Pox virus and Zinc finger) domain containing proteins. This interaction has been described using the yeast two-hybrid method and in *Drosophila* S2 cells, providing additional evidence for this TF-TF interaction (Pagans et al., 2002). Ttk has been shown to function both as a transcriptional repressor and an activator, playing a variety of developmental roles including development of the nervous system, photoreceptor differentiation and in tracheal development (Arujo et al., 2007, Badenhorst 2001, Lai and Li 1999). Trl (also known as GAGA factor, or GAF) has been suggested to play a role in transcriptional activation through chromatin remodeling and in some cases, is necessary for full activation of transcription complexes (Bayarmagnai et al., 2012, Granok 1995). This would suggest that Ttk activity is modulated through interactions with Trl, likely playing a role in activation of expression of shared targets.

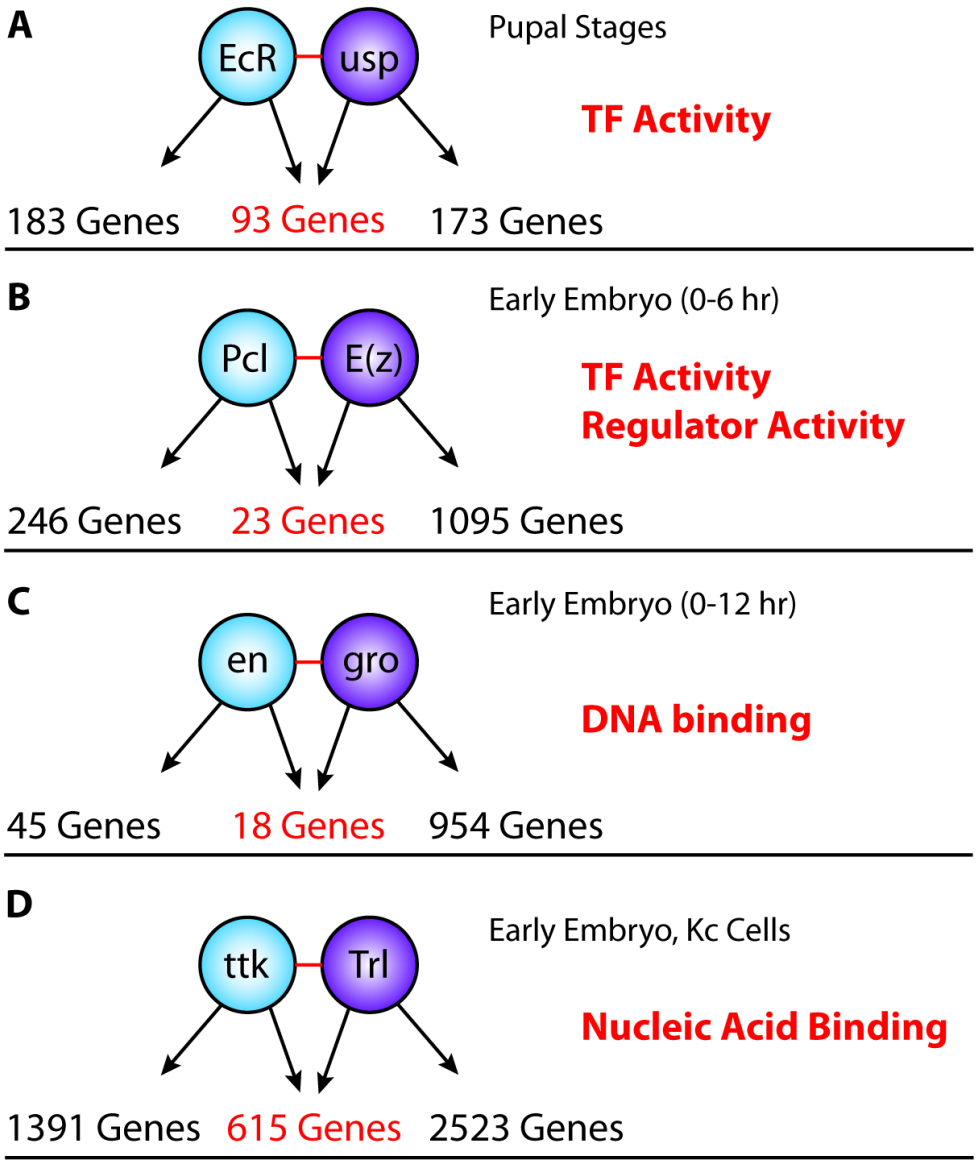

Figure 3.4 Common Physical Targets of Interacting TFs
Shared physical targets of interacting TF pairs. (A) ecdysone receptor (EcR) and ultraspiracle (usp) comprise the two parts of the complete Ecdysone receptor. They co-occupy 93 shared targets during pupal stages. (B) Polycomblike (Pcl) and Enhancer of zeste (E(z)), two members of the Pcl-PRC2 complex. (C) engrailed (en) and groucho (gro). (D) tramtrack (ttk) and Trithoraxlike (Trl), two BTB/POZ domain containing proteins. The red text indicates enrichment of shared TF targets for functional terms based on Panther analysis (Thomas et al., 2003).

Inferred Regulatory Motifs for TF complexes

To gain insight into the regulatory consequences of the PPI in our network, we have integrated our results with existing inferred regulatory network models (Marbach et al., 2012). These inferred networks integrate a wide range of data sets, including TF binding, gene expression and chromatin modifications, utilizing supervised and unsupervised machine-learning frameworks to predict regulatory edges. Supervised machine learning utilizes a training dataset (in this case, established regulatory relationships from the REDfly database (Gallo et al., 2010) to "teach" a network, providing either an error or reward based on this training set to achieve a certain range of outcomes for a set of inputs. Unsupervised machine learning lacks this training set, but rather looks for hidden organization within a dataset without the help of a "teacher."

These inferred networks have been shown to be a useful tool in predicting gene function, recovering previously identified regulatory edges at a higher rate when compared with other methods, such as TF binding data (Marbach et al., 2012). It is important to note, however, that protein-protein interaction data were not included in the assembly of these particular networks, nor do they contain PPI edges. By integrating our PPI data with such transcriptional regulatory networks, we provide a new dimension to this analysis, gaining insight into the combinatorial action of interacting TFs by linking regulatory edges directly to TF protein complexes.

To combine PPI with regulatory interactions and to probe these large integrated networks, we defined a set of TF regulatory motifs based on physical and regulatory interactions (Figure 3.5A, Supplemental Table 3.5). These three motifs represent instances where (1) An interacting protein is regulated by its binding partner; (2) Where two interacting proteins regulate the same target; and (3) a single factor regulates

interacting proteins. Each instance of these motifs essentially defines a biological hypothesis, representing an avenue for future inquiry.

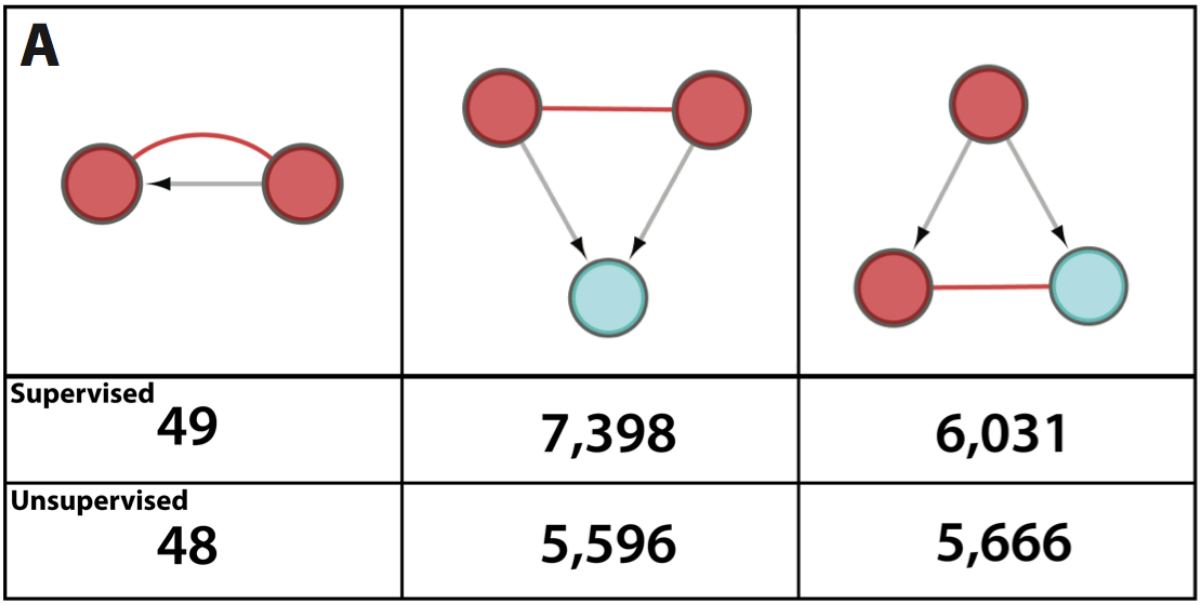

Figure 3.5 Transcriptional Regulatory Motifs
Transcriptional regulatory motifs, representing instances where (A) an interacting protein regulates its binding partner (1:1), (B) combinatorial regulation of a target by two interacting factors (2:1), and (C) regulation of interacting proteins by a single factor (1:2). Red edges indicate protein-protein interactions while grey edges with arrows indicate directional regulatory edges. The numbers indicate the total count uncovered for each motif within the supervised and unsupervised models.

By permuting the edges of both our high confidence PPI network and the inferred regulatory networks independently, we confirmed that these motifs are more frequent than expected by chance. Furthermore, as we have demonstrated the predictive power of the high-confidence interactions in our PPI network, focusing only on motifs containing one of our PPI edges, effectively filters the regulatory network based on experimental evidence. These motifs were then combined to build integrated PPI-regulatory networks containing 22,781 edges between 3,145 proteins and 19,062 edges between 2,331

proteins, corresponding to supervised and unsupervised models respectively (Figure 3.6, Supplemental Figure 3.2).

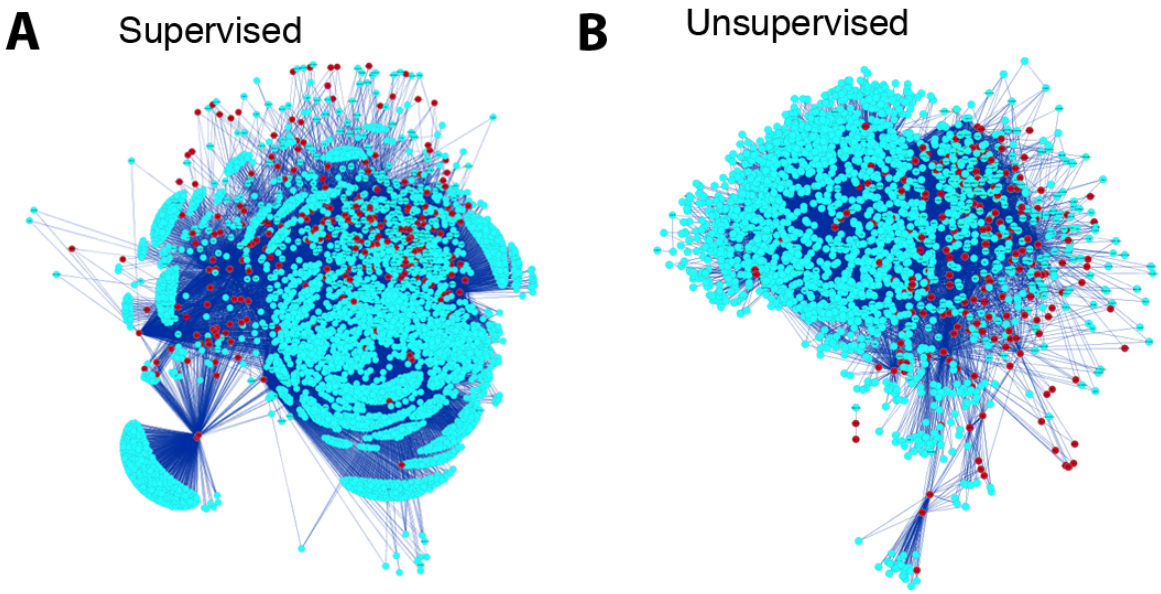

Figure 3.6 Integrated Networks
Integrated networks based on our transcriptional regulatory motifs. These networks contain both our high-confidence PPI interactions as well as regulatory edges that are directly linked to PPI edges. (A) Integrated network based on supervised regulatory network inference, containing 22,781 edges. (B) Integrated network based on unsupervised regulatory network inference, containing 19,062 edges.

Within the supervised integrated network, we have highlighted the network related to the Dp transcription factor and E2F, members of the dREAM (RBF, dE2F2, dMyb) complex (Figure 3.7). The dREAM complex is conserved in most eukaryotes and plays multiple roles including the regulation of development, cell division and apoptosis (van den Heuvel and Dyson 2008). Dp and E2f comprise a heterodimeric transcription factor that is important in the G1/S phase transition during the cell cycle, where E2f levels are rate-limiting for cell proliferation (Johnson et al., 1993).

Previous work has described interactions between E2F and Dp and Rbf and E2F2, corroborating the protein-protein interactions in our network (Figure 3.7A, Frolov et al., 2001). Another component of the dREAM complex, Myb, acts in a mutually exclusive manner with Dp/E2f to regulate target selection (Georlette et al., 2007). Though we did not recover Myb as a physical interactor, it is one of only three proteins that are inferred to both regulate Dp/E2f and are in turn targeted by the TF pair. The other two proteins are MTA1-like and CG17385, which have not been previously tied to dREAM functions, and thus define targets for functional analyses (Figure 3.7A). As expected, downstream targets of DP/E2f in our network include genes important for the cell cycle (Figure 3.7D) and DNA replication (Figure 3.7E).

The dREAM complex is thought to regulate transcription in three ways: the repressive binding of Rbf to E2f, inhibition of the basal transcription machinery and by recruiting chromatin-modifying proteins (Frolov et al., 2001). Our regulatory network reflects all three of these possibilities, showing a physical interaction between Rbf and E2f, the targeting of a number of basal transcriptional machinery components (Figure 3.7C), and the regulation of chromatin-modifying proteins such as brahma and MRG15 (Figure 3.7G). Other downstream targets of DP/E2f in our network include a group largely enriched for transcription-related proteins (Figure 3.7F) and 28 targets that are unannotated (Figure 3.7H). Dp and E2f are themselves targeted by a cohort of TFs and co-factors including DREF, Mad and trithorax-like (Figure 3.7B). Consequently, we have identified a well-characterized protein complex, a number of its known regulatory targets, and, most interestingly, targets that have not been previously linked to dREAM complex function.

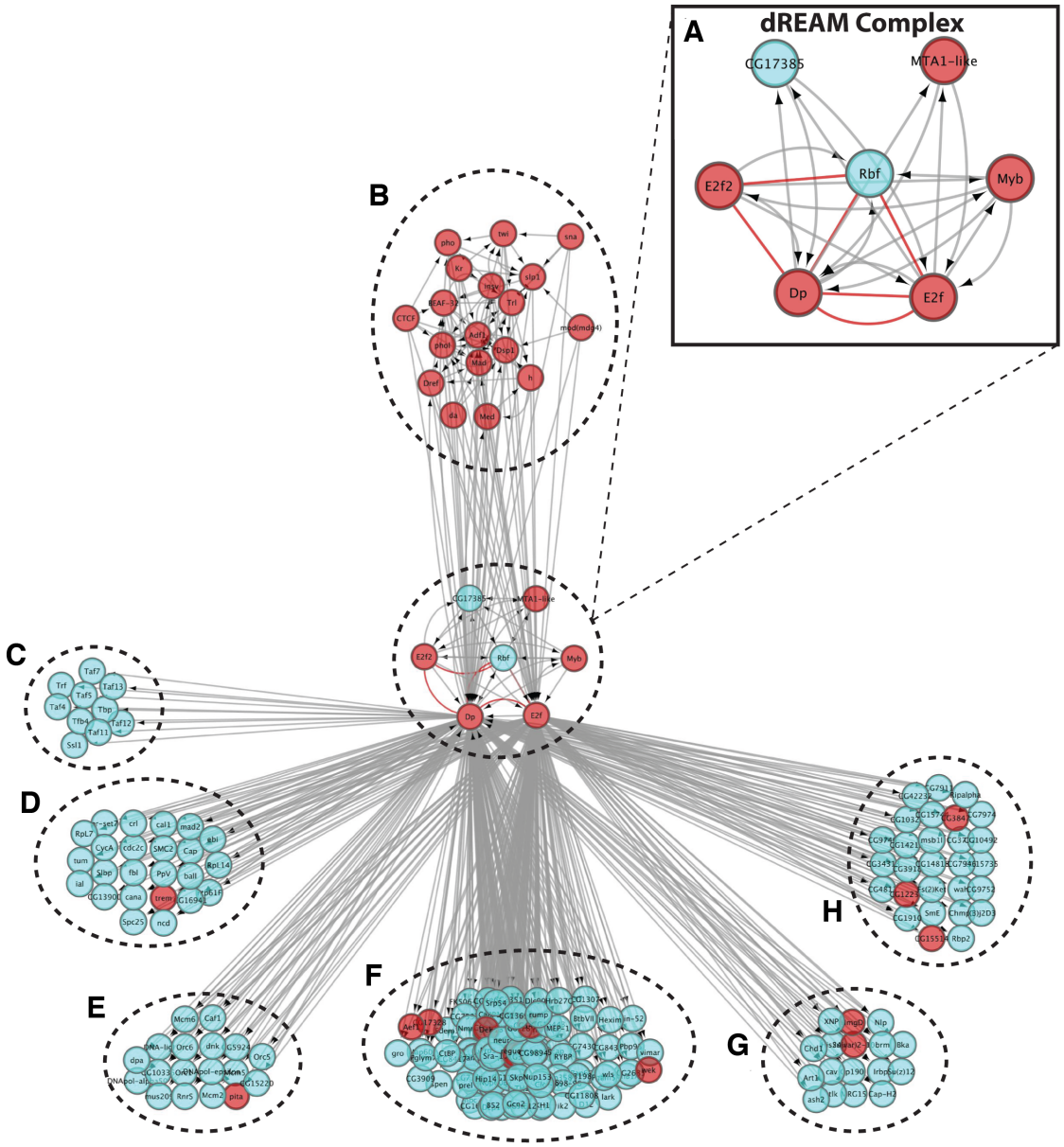

Figure 3.7 Integrated Network View of the dREAM Complex
The components of the *Drosophila* dREAM complex recovered in our integrated network. Red nodes represent TFs. Blue nodes represent non-TF proteins. Gray edges indicate regulatory interactions with the direction of regulation indicated by an arrow. Red edges indicate protein-protein interactions. (B) Transcriptional regulators of Dp-E2f, (C) Basal transcriptional machinery components, (D) Cell cycle proteins, (E) DNA Replication-related proteins, (F) Transcription-related proteins, (G) Chromatin-related proteins, (H) Unannotated targets of Dp-E2F.

Connecting Functional Networks

Genetic screens, especially in *Drosophila*, have been used as a powerful tool to define networks of proteins that share function (reviewed in St Johnston 2002). One of the resulting difficulties is in understanding, at a mechanistic level, how these proteins are connected to one another. On the other end of the spectrum, PPI networks describe the physical relationships between proteins, but do not capture functional relationships. As we have shown in the previous chapter, there is some overlap between these two network types, however, not every functional relationship is the result of a direct protein-protein interaction. As such, the majority of network edges between these two data types do not typically overlap. By combining GRNs with our PPI data, our integrated network allows us to bridge the gap between physical and functional relationships through defined regulatory edges.

As an example, we once again focused on the genetic interaction network of *mastermind*, defined in a genome-wide screen in *Drosophila* (Kankel et al., 2007). In this study, 408 genes were shown to genetically interact with *mastermind, in vivo*. Our supervised and unsupervised integrated networks contain 140 and 103 of these modifiers respectively. If we examine direct relationships between these in our networks, 88 and 34 proteins are directly linked to one another (Figure 3.8). If we expand this view to include first neighbor interactions, all *mam* modifiers in both instances are connected to one another.

The organization of these networks reveals several potential "hubs" of regulation, based on the total number of edges that connect to a particular node. For instance, the transcription factor *serpent* (srp) is connected by 12 separate network edges in our supervised network (Figure 3.8B). Though *srp* itself has not been demonstrated to be

directly regulated by the Notch pathway, it has been previously shown to function upstream of direct Notch targets during *Drosophila* larval hematopoiesis (Duvic et al., 2002). This would suggest a potential mechanism by which loss of *srp* would modulate Notch activity downstream, thus explaining the genetic interaction between *srp* and *mastermind*. Interestingly, our network identifies *mam* as a direct target of *srp*.

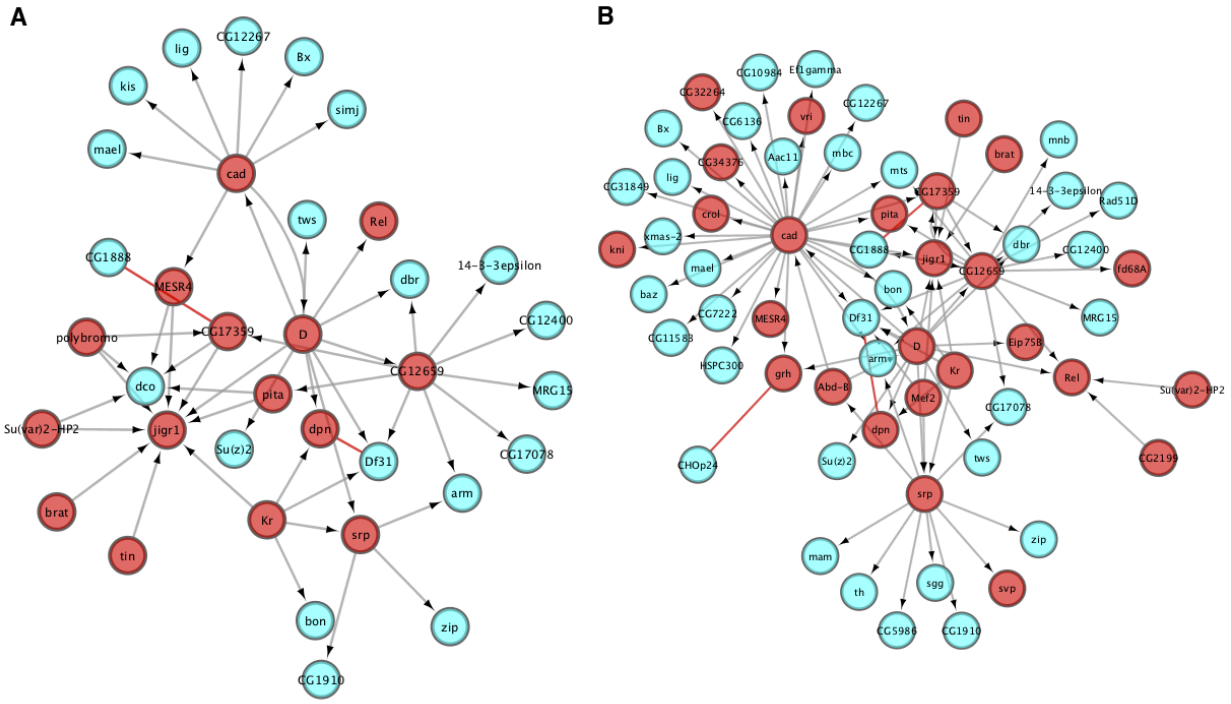

Figure 3.8 Unsupervised and Supervised Networks of the Functional Genetic Network of *mastermind* interactors.
(A) Unsupervised network view of 34 *mastermind* modifiers. (B) Supervised network view of 65 *mastermind* modifiers. Red nodes represent TFs. Blue nodes represent non-TF proteins. Red edges represent protein-protein interactions. Gray edges with arrows represent directional regulatory edges.

While these regulatory edges will certainly vary depending on context, this approach provides a network of hypotheses linking functional data points to be used as the basis for probing the mechanisms that link these proteins. We expect, as more data become available, that these networks will be further refined and expanded to provide higher resolution insight into the mechanisms driving biological function. As things stand, our integrated networks provide a substantial foundation from which to explore the mechanisms that connect functional datasets.

Discussion

We have taken our TF PPI interaction network and have integrated data derived from different experimental approaches, including distinct experimental parameters, to further explore the biology of TF protein complexes and their regulatory relationships. Our tissue-specific sub-networks emphasize the importance of context with regard to TF function. We have defined groups of proteins based on their broad or specific expression, and then connect these categories, providing insights into how general and specific TFs cooperate with one another to drive transcriptional programs. As has been suggested previously, it is likely that the presence of a particular TF protein interaction within a specific tissue, rather than the expression of a single tissue-specific TF, confers tissue identity (Ravasi et al., 2010). As such, we expect our tissue-specific interaction networks to be valuable tools for further probing the contributions of TFs to developmental processes.

We next examined the combinatorial targets of interacting TFs. As previous work has shown, TFs do not function in isolation, nor does physical binding of a single factor necessarily correlate to a change in gene expression. It is the combination of various TFs

and their interacting proteins that confers a specific activity. As such, TF targets should be viewed from the perspective of the TF protein complex and, indeed, we find multiple examples of PPI interactions that are supported by genome-wide DNA-binding data, as well as interactions that postulate novel functional hypotheses, warranting further exploration. It is important to note that despite the large number of TF occupancy datasets currently available, they are severely biased for proteins that have been previously studied, as ChIP methods depend on the availability of useful antibodies. Given that a substantial portion of our PPI network is composed of unstudied proteins, the overlap between these two datasets is still relatively small. Currently, the modENCODE project is systematically producing antibodies for TFs in *Drosophila* so it expected that more comprehensive target prediction in conjunction with our PPI network will be possible in the near future.

Finally, we connected TF protein complexes to gene regulatory networks using inferred regulatory edges, allowing us to expand target prediction beyond direct physical targets, and tying TFs directly to interacting groups of proteins. As we have demonstrated the predictive value of the physical edges in our network, this likely improves the quality of the inferred regulatory network, given that we examined only the edges that are directly linked to an experimentally observed physical interaction. We have demonstrated the utility of this integrated network in the characterization of a TF protein complex, including the identification of both characterized and novel targets, and used these integrated networks to interrogate large-scale functional data sets. While genetic screens have been used for decades, connecting the large number of functional modifiers identified in these screens to one another has been, and remains, a significant challenge. While Gene Ontology analysis certainly provides insights into the

categorization of genes within these datasets, the complex relationships between these components are only captured from a network perspective. Our integrated network provides a considerable foundation from which to build hypotheses as to how various functionally connected proteins are related to one another at a mechanistic level.

Ultimately, we view our data as a framework for developing specific hypotheses for future studies in both *Drosophila* and other metazoans. Given the conservation of regulatory programs, it is likely that many of the regulatory connections presented here will be preserved in other species, though possibly (and interestingly) used in different biological contexts. As transcription factors represent a fundamental point of regulation in the cell, we expect this present work to be relevant to the vast majority of biological processes.

Materials and Methods

Tissue Specificity Score (TSPS)

The tissue specificity score was executed as previously described in Ravasi et al., 2010, utilizing 24 mRNA-sequencing datasets from Smibert et al., 2012, encompassing 24 groups containing various tissues dissected from *Oregon R* wild type flies. The distribution for all proteins based on their TSPS was fit to a tri-modal Gaussian distribution, identifying cut-off values of 0.4781 for low (general) specificity proteins, while the cut-off for high specificity (specific) was 1.17406.

Specific Tissue Assignments and Network Construction

High specificity proteins, based on TSPS distribution were assigned to specific tissues using the method described in Kadota et al., 2002. This method searches for proteins that exhibit expression profiles that are very different in one tissue versus another, defining them as outliers and thus assigning them to a specific tissue. These tissue-specific proteins were then combined with the group of broadly expressed, low TSPS proteins, and assembled into respective tissue-specific networks using Cytoscape (Shannon et al., 2003).

Chromatin-Immunoprecipitation Data

ChIP data were used from both the modENCODE project (Roy et al., 2010) and the Berkeley Drosophila Transcription Network Project (MacArthurt et al., 2009). For published ChIP-chip and ChIP-seq datasets, filtered peaks were taken directly from the published analyses. New ChIP-seq datasets were generated as described in Roy et al., 2010, but analyzed through the Irreproducible Discovery Rate data analysis pipeline, described in detail here (https://sites.google.com/site/anshulkundaje/projects/idr).

Integrated Network Construction and Analysis

The supervised and unsupervised regulatory networks described in Marbach et al., 2012 were assembled together with our high-confidence TF PPI network. We then searched these combined networks for our three defined TF regulatory motifs. All edges that did not fall into one of these three motifs were filtered out and remaining edges were assembled to create our integrated networks. Network views and all subsequent analyses were performed in Cytoscape (Shannon et al., 2003).

Bibliography

Adryan, B., and Teichmann, S.A. (2010). The developmental expression dynamics of Drosophila melanogaster transcription factors. Genome biology 11, R40.

Araujo, S.J., Cela, C., and Llimargas, M. (2007). Tramtrack regulates different morphogenetic events during Drosophila tracheal development. Development 134, 3665-3676.

Badenhorst, P. (2001). Tramtrack controls glial number and identity in the Drosophila embryonic CNS. Development 128, 4093-4101.

Bayarmagnai, B., Nicolay, B.N., Islam, A.B., Lopez-Bigas, N., and Frolov, M.V. (2012). Drosophila GAGA factor is required for full activation of the dE2f1-Yki/Sd transcriptional program. Cell Cycle 11, 4191-4202.

Bulyk ML and Walhout A.J.M. Chapter 4, "Gene Regulatory Networks", in "Handbook of Systems Biology: Concepts and Insights", 2012, pages 65-88, Elsevier Inc. Edited by Albertha J.M. Walhout, Marc Vidal, and Job Dekker

Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes & development 6, 367-379.

Capaldi, A.P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nature genetics 40, 1300-1306.

Duvic, B., Hoffmann, J.A., Meister, M., and Royet, J. (2002). Notch signaling controls lineage specification during Drosophila larval hematopoiesis. Current biology : CB 12, 1923-1927.

Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nature reviews Genetics 10, 141-148.

Frolov, M.V., Huen, D.S., Stevaux, O., Dimova, D., Balczarek-Strang, K., Elsdon, M., and Dyson, N.J. (2001). Functional antagonism between E2F family members. Genes & development 15, 2146-2160.

Gallo, S.M., Gerrard, D.T., Miner, D., Simich, M., Des Soye, B., Bergman, C.M., and Halfon, M.S. (2011). REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic acids research 39, D118-123.

Garg, A., Kuzin, A., Brody, T., Odenwald, W., Bell, J. (2007). Nerfin-1: A novel binding partner of Scalloped. A. Dros. Res. Conf. 48:329B ABSTRACT

Georlette, D., Ahn, S., MacAlpine, D.M., Cheung, E., Lewis, P.W., Beall, E.L., Bell, S.P., Speed, T., Manak, J.R., and Botchan, M.R. (2007). Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. Genes & development 21, 2880-2896.

Goulev, Y., Fauny, J.D., Gonzalez-Marti, B., Flagiello, D., Silber, J., and Zider, A. (2008). SCALLOPED interacts with YORKIE, the nuclear effector of the hippo tumor-suppressor pathway in Drosophila. Current biology : CB 18, 435-441.

Granok, H., Leibovitch, B.A., Shaffer, C.D., and Elgin, S.C. (1995). Chromatin. Ga-ga over GAGA factor. Current biology : CB 5, 238-241.

Hittinger, C.T., and Carroll, S.B. (2008). Evolution of an insect-specific GROUCHO-interaction motif in the ENGRAILED selector protein. Evolution & development 10, 537-545.

Johnson, D.G., Schwarz, J.K., Cress, W.D., and Nevins, J.R. (1993). Expression of transcription factor E2F1 induces quiescent cells to enter S phase. Nature 365, 349-352.

Kadota, K., Nishimura, S., Bono, H., Nakamura, S., Hayashizaki, Y., Okazaki, Y., and Takahashi, K. (2003). Detection of genes with tissue-specific expression patterns using Akaike's information criterion procedure. Physiological genomics 12, 251-259.

Kankel, M.W., Hurlbut, G.D., Upadhyay, G., Yajnik, V., Yedvobnick, B., and Artavanis-Tsakonas, S. (2007). Investigating the genetic circuitry of mastermind in Drosophila, a notch signal effector. Genetics 177, 2493-2505.

Kuzin, A., Brody, T., Moore, A.W., and Odenwald, W.F. (2005). Nerfin-1 is required for early axon guidance decisions in the developing Drosophila CNS. Developmental biology 277, 347-365.

Lai, Z.C., and Li, Y. (1999). Tramtrack69 is positively and autonomously required for Drosophila photoreceptor development. Genetics 152, 299-305.

MacArthur, S., Li, X.Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keranen, S.V., et al. (2009). Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome biology 10, R80.

Marbach, D., Roy, S., Ay, F., Meyer, P.E., Candeias, R., Kahveci, T., Bristow, C.A., and Kellis, M. (2012). Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks. Genome research 22, 1334-1349.

Pagans, S., Ortiz-Lombardia, M., Espinas, M.L., Bernues, J., and Azorin, F. (2002). The Drosophila transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. Nucleic acids research 30, 4406-4413.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140, 744-752.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330, 1787-1797.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11), 2498-504.

Smibert, P., Miura, P., Westholm, J.O., Shenker, S., May, G., Duff, M.O., Zhang, D., Eads, B.D., Carlson, J., Brown, J.B., et al. (2012). Global patterns of tissue-specific alternative polyadenylation in Drosophila. Cell reports 1, 277-289.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics 13, 613-626.

St Johnston, D. (2002). The art and design of genetic screens: Drosophila melanogaster. Nature reviews Genetics 3, 176-188.

van den Heuvel, S., and Dyson, N.J. (2008). Conserved functions of the pRB and E2F families. Nature reviews Molecular cell biology 9, 713-724.

Walhout, A.J. (2006). Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome research 16, 1445-1454.

Weaver, Z.A., and Kane, C.M. (1997). Genomic characterization of a testis-specific TFIIS (TCEA2) gene. Genomics 46, 516-519.

Yao, T., Forman, B., Jiang, Z., Cherbas, L., Chen, J., McKeown, M., Cherbas, P., Evans, R. (1993) Functional ecdysone receptor is the product of *EcR* and *ultraspiracle* genes. Nature 366, 476-479.

Chapter 4

General Discussion

This body of work encompasses a study of transcription factors, proteins that play an essential role in the biology of the cell through the regulation of transcription. Though frequently discussed in the context of development, TFs play myriad roles in the embryo and the adult. The functions of these proteins are defined by several parameters, mainly their protein-protein interactions, protein-DNA interactions and the specific contexts in which they function. Although recent advances, in particular sequencing technologies, have allowed for genome-wide analyses of both gene expression and TF-DNA targets, the majority of TF protein interactions had not yet been defined. As the central component of the dissertation, I systematically probed these relationships using a co-AP/MS approach, describing connections for nearly half of the characterized TFs in *Drosophila*.

Chapter Two discusses the generation of this protein interaction dataset, the construction of a high-confidence protein interaction network and the use of these data to both predict and functionally validate relationships *in vivo*. Although the functional studies here are specifically focused on the Notch signaling pathway, these data represent a general resource that can be used to probe a multitude of biological questions. Of particular interest, ~40% of the proteins in the high-confidence TF interaction network are currently unstudied (many have annotations based only on electronic inference). As highlighted in the example of the extradenticle-homothorax transcription co-factor protein complex (Figure 2.4A), many of these "unknowns" are directly linked to proteins or protein complexes of known function, providing an entry point for further inquiry.

Although this study encompasses a significant fraction of *Drosophila* TFs, it represents a first pass analysis of these proteins. As new expression clones become available for the TFs not included in this work, these should be analyzed and incorporated to improve coverage of the *Drosophila* TF protein interaction network.

Additionally, these data were generated using *Drosophila* S2R+ cells, which express approximately half of the ~14,000 protein-coding genes in the *Drosophila* genome. It is certain that some interactions will be missed as some proteins are simply not present in our system. As such, the use of other cell lines that express these other proteins or experiments *in vivo* is necessary to capture the remaining portions of the TF protein interactome. It is also likely that the methods used here do not capture transient or weak interactions, which may be characterized using methods such as cross-linking prior to mass spectrometry analysis.

In Chapter Three, I integrated the TF protein interaction network with a number of data types to probe various aspects of TF function. First, expression data sets were used to define tissue specificity for proteins in the high-confidence network, which were then assigned to individual tissues and subsequently assembled into 24 different tissue-specific interaction networks. This analysis allows us to take data that were generated in a cell culture system and examines them within distinct settings in the animal, providing an atlas of relevant interactions for each tissue. It has been suggested that it is not only the expression of a particular TF, but also the presence of a specific TF interaction that is important in specifying a tissue or an activity. As such, each of these interactions represents a biological hypothesis for future study. It is important to note that while this analysis covers 24 tissues and time points, many of these expression datasets can and should be further refined. For example, the expression data for imaginal discs does not differentiate between different disc types (e.g., wing vs. eye), an important distinction as these represent precursors for completely different tissues. As more precise expression data become available, one would expect the resolution of these tissue-specific networks to increase.

Second, I combined multiple transcription factor occupancy datasets (ChIP-chip and ChIP-seq) to identify shared physical targets of TFs that interact with one another. Just as two proteins that interact should be expressed in the same tissue, two TFs that interact and function together should bind directly or indirectly at the same DNA targets. Indeed, I found that shared targets were identified for all interacting TFs where occupancy data were available. This analysis not only identifies potential targets of combinatorial regulation by these factors, but also provides additional evidence for the existence of each TF-TF interaction. The non-overlapping targets of interacting TFs are also of particular interest, as these may provide insights into the regulatory mechanisms affected by these genes. Although this comparison corroborates some of my interaction data, unfortunately, the overlap between existing TF occupancy datasets and our data was small. This is likely a reflection of the fact that ChIP based methods depend on the availability of suitable antibodies, thus biasing experiments towards previously characterized proteins and limiting the total number of available datasets. Currently, as a part of the modENCODE project, TF antibodies for ChIP are being systematically generated, so it is expected that more data will become available in the near future.

Finally, I utilized inferred regulatory networks to connect TF protein interactions to the regulatory network of the cell. These inferred networks were constructed using multiple large-scale datasets predicting regulatory connections, based on a machine-learning framework. While these methods are not a perfect substitute for experiments done at the bench, they provide a means to extract information from various data types and in principle, use a mathematical approach to define biological hypotheses. The inferred regulatory networks were combined with the TF interaction network and subsequently filtered by searching for three distinct transcriptional regulatory motifs, each

containing a protein-protein interaction. I highlighted the dREAM complex as an example of the utility of this integrated approach, identifying both well-characterized and novel targets of this TF protein complex. This case represents just one of many thousands of examples to be explored in these networks.

I subsequently used these integrated networks to probe the space among functionally related proteins identified in genetic modifier screens. It is often difficult to connect the genes identified in such studies, as most of these proteins do not directly interact with one another. For instance, for the genome-wide *mastermind* genetic screen that was used in this work, only about one-third of the modifiers are connected to one another in the current literature. This is likely an overestimate as the majority of the PPI edges that connect these proteins were derived from low-quality, unfiltered interaction data. The integrated networks in this study contain ~25% of the modifiers identified in the *mastermind* screen. Though this is a relatively small fraction of the entire functional dataset, the modifiers that are present are highly interconnected. In fact, for both the supervised and unsupervised integrated networks, all *mastermind* modifiers present are connected to one another when viewed as a first neighbor network. As each of these modifiers has already been demonstrated to interact at a functional level, the edges in this integrated analysis provide predictions regarding the mechanisms that lead to this shared function.

The data presented here have relied heavily on the use of currently available genomic datasets. Given the recent explosion of large-scale studies, it is expected that in the coming months and years that these networks can and should be expanded and further refined as more resources become available. I would expect better coverage of the transcription factor interactome as well as the whole protein interactome, in *Drosophila* as

well as in other species, including humans. Regulatory network models will also be more complete as both the modENCODE and ENCODE projects are systematically characterizing DNA elements that are relevant for the function of genes in both *Drosophila* and in humans. Combining these massive datasets using further improved computational models will provide a strong platform for identifying and exploring regulatory mechanisms. I also fully expect improved experimental methods to provide better data for such analyses. In particular, the recent development of small-scale chromatin immunoprecipitation (Adli and Bernstein 2011) is especially exciting as this opens the door to inquiry in a wide range of *in vivo* settings, especially during development, where TF expression frequently defines developmental domains (e.g., developing motor neuron pools). This raises the prospect of defining high-resolution regulatory networks within sub-compartments of developing tissues *in vivo*.

Taken together, the work in this dissertation explores multiple aspects of transcription factor biology to provide a set of tools from which to generate biological hypotheses. Though I worked exclusively in *Drosophila*, as TFs are well conserved, I fully anticipate these findings to be relevant for studies in other species including humans. Gene regulatory mechanisms are also frequently preserved from one species to the next, often used in different contexts. As such, the integrated regulatory networks described here can be used as the basis for studies in other species. Given the importance of transcriptional regulation in the biology of the cell, I expect these findings to be directly relevant to the majority of biological processes, from the earliest embryo all the way to the senescing adult.

Bibliography

Adli, M., and Bernstein, B.E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. Nature protocols 6, 1656-1668.

Appendix A

Supplemental Figures and Tables

Chapter 2

A Cytoscape file of the high-confidence transcription factor interaction network is attached electronically (Supplemental Figure 2.1). Supplemental Table 2.1 contains a list of all the unique proteins identified across all MS experiments in this study. Supplemental Table 2.2 is an electronic supplement containing an Excel file with the raw interaction data and a list of all binary TF-TF interactions identified in the raw, unscored, MS data from all experiments. Supplemental Table 2.3 (electronic supplement) is an Excel file containing the HGScore analysis. Supplemental Table 2.4 (electronic supplement) is an Excel file containing the high-confidence network edges. Supplemental Table 2.5 contains an Excel file with all DroID edges recovered by the TF interaction network analysis.

Supplemental Table 2.1 Unique Proteins Identified across all MS experiments A list of all unique proteins identified in the Co-AP/MS analysis. The Flybase Gene ID and Gene symbols are listed.

Supplemental Table 2.1			
Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0011710	Sep1	FBgn0027513	ana2
FBgn0010339	128up	FBgn0025111	Ant2
FBgn0020238	14-3-3epsilon	FBgn0000084	AnxB10
FBgn0004907	14-3-3zeta	FBgn0029512	Aos1
FBgn0053100	4EHP	FBgn0010380	AP-1-2beta
FBgn0027885	Aac11	FBgn0263350	AP-2alpha
FBgn0002069	Aats-asp	FBgn0263351	AP-2mu
FBgn0005674	Aats-glupro	FBgn0043012	AP-2sigma
FBgn0027084	Aats-lys	FBgn0032136	Apoltp
FBgn0000017	Abl	FBgn0033926	Arc1
FBgn0015331	abs	FBgn0013749	Arf102F
FBgn0033246	ACC	FBgn0010348	Arf79F
FBgn0027620	Acf1	FBgn0037182	ArfGAP3
FBgn0033749	achi	FBgn0037884	Arfip
FBgn0013955	Ack-like	FBgn0011745	Arp1
FBgn0263198	Acn	FBgn0031050	Arp10
FBgn0263120	Acsl	FBgn0011742	Arp2
FBgn0000042	Act5C	FBgn0262716	Arp3
FBgn0000045	Act79B	FBgn0038576	Arp5
FBgn0000047	Act88F	FBgn0011741	Arp6
FBgn0000667	Actn	FBgn0030877	Arp8
FBgn0053520	Ada2a	FBgn0038369	Arpc3A
FBgn0027619	Adam	FBgn0031781	Arpc4
FBgn0020513	ade5	FBgn0033062	Ars2
FBgn0000054	Adf1	FBgn0000139	ash2
FBgn0000055	Adh	FBgn0020407	asun
FBgn0022708	Adk2	FBgn0031876	Atac1
FBgn0262739	AGO1	FBgn0032691	Atac2
FBgn0087035	AGO2	FBgn0052343	Atac3
FBgn0027932	Akap200	FBgn0039946	ATbp
FBgn0000061	al	FBgn0029943	Atg5
FBgn0260972	alc	FBgn0010750	atms
FBgn0000064	Ald	FBgn0020236	ATPCL
FBgn0012036	Aldh	FBgn0010217	ATPsyn-beta
FBgn0013746	alien	FBgn0016119	ATPsyn-Cf6
FBgn0086346	ALiX	FBgn0020235	ATPsyn-gamma
FBgn0015575	alpha-Est7	FBgn0019637	Atu
FBgn0250789	alpha-Spec	FBgn0041188	Atx2
FBgn0025725	alphaCop	FBgn0000150	awd
FBgn0003884	alphaTub84B	FBgn0025185	az2

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0011758	B-H1	FBgn0004893	bowl
FBgn0004854	B-H2	FBgn0033155	Br140
FBgn0004587	B52	FBgn0039654	Brd8
FBgn0031977	baf	FBgn0086694	Bre1
FBgn0045866	bai	FBgn0038499	Brf
FBgn0027889	ball	FBgn0024250	brk
FBgn0004862	bap	FBgn0000212	brm
FBgn0042085	Bap170	FBgn0013755	Bro
FBgn0025716	Bap55	FBgn0004101	bs
FBgn0025463	Bap60	FBgn0000529	bsh
FBgn0024251	bbx	FBgn0263108	BtbVII
FBgn0000166	bcd	FBgn0030501	BthD
FBgn0026149	BCL7-like	FBgn0045862	btz
FBgn0015602	BEAF-32	FBgn0025457	Bub3
FBgn0037660	beag	FBgn0025458	BubR1
FBgn0263231	bel	FBgn0000242	Bx
FBgn0000173	ben	FBgn0004856	Bx42
FBgn0250788	beta-Spec	FBgn0010292	bys
FBgn0025724	beta'Cop	FBgn0040235	c12.1
FBgn0008635	betaCop	FBgn0004863	C15
FBgn0052598	betaNACtes6	FBgn0025678	CaBP1
FBgn0003887	betaTub56D	FBgn0000250	cact
FBgn0003888	betaTub60D	FBgn0031114	cactin
FBgn0003889	betaTub85D	FBgn0000251	cad
FBgn0003890	betaTub97EF	FBgn0263347	Caf1
FBgn0013753	Bgb	FBgn0033526	Caf1-105
FBgn0004581	bgcn	FBgn0030054	Caf1-180
FBgn0000181	bic	FBgn0000253	Cam
FBgn0000183	BicD	FBgn0015615	Cap
FBgn0039509	bigmax	FBgn0042134	Capr
FBgn0045759	bin	FBgn0261458	capt
FBgn0024491	Bin1	FBgn0004878	cas
FBgn0026262	bip2	FBgn0029093	cathD
FBgn0010520	Bka	FBgn0015919	caup
FBgn0015907	bl	FBgn0011571	caz
FBgn0035608	blanks	FBgn0033842	cbc
FBgn0035625	Blimp-1	FBgn0022943	Cbp20
FBgn0002906	Blm	FBgn0022942	Cbp80
FBgn0011211	blw	FBgn0043364	cbt
FBgn0023097	bon	FBgn0010621	Cct5
FBgn0032105	borr	FBgn0015019	Cctgamma

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0004106	cdc2	FBgn0037470	CG1091
FBgn0013435	cdc2rk	FBgn0031178	CG10918
FBgn0011573	Cdc37	FBgn0032858	CG10949
FBgn0010341	Cdc42	FBgn0030010	CG10959
FBgn0035136	Cdc5	FBgn0037379	CG10979
FBgn0037093	Cdk12	FBgn0036305	CG10984
FBgn0263237	Cdk7	FBgn0031736	CG11030
FBgn0015618	Cdk8	FBgn0039929	CG11076
FBgn0261532	cdm	FBgn0046222	CG1109
FBgn0029092	ced-6	FBgn0033160	CG11107
FBgn0030657	cerv	FBgn0039250	CG11120
FBgn0000286	Cf2	FBgn0039936	CG11148
FBgn0000289	cg	FBgn0037299	CG1115
FBgn0035720	CG10077	FBgn0034528	CG11180
FBgn0035715	CG10103	FBgn0031851	CG11188
FBgn0033971	CG10209	FBgn0037120	CG11247
FBgn0027514	CG1024	FBgn0031883	CG11266
FBgn0035690	CG10274	FBgn0030058	CG11294
FBgn0034654	CG10306	FBgn0039849	CG11334
FBgn0034643	CG10321	FBgn0037022	CG11396
FBgn0038454	CG10324	FBgn0040366	CG11398
FBgn0032690	CG10333	FBgn0024364	CG11417
FBgn0032698	CG10336	FBgn0037504	CG1142
FBgn0030342	CG10347	FBgn0024985	CG11448
FBgn0032707	CG10348	FBgn0039733	CG11504
FBgn0032814	CG10366	FBgn0035424	CG11505
FBgn0033019	CG10395	FBgn0039868	CG11563
FBgn0036277	CG10418	FBgn0035524	CG11583
FBgn0032730	CG10431	FBgn0030314	CG11696
FBgn0034636	CG10440	FBgn0031115	CG11710
FBgn0032822	CG10466	FBgn0031391	CG11723
FBgn0035630	CG10576	FBgn0033527	CG11777
FBgn0032717	CG10600	FBgn0034495	CG11788
FBgn0032731	CG10641	FBgn0034000	CG11808
FBgn0036292	CG10646	FBgn0036099	CG11811
FBgn0036294	CG10654	FBgn0039627	CG11837
FBgn0035590	CG10673	FBgn0039271	CG11839
FBgn0037881	CG10703	FBgn0039303	CG11857
FBgn0036314	CG10754	FBgn0039642	CG11882
FBgn0029979	CG10777	FBgn0034425	CG11906
FBgn0036207	CG10907	FBgn0039274	CG11920

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0037655	CG11984	FBgn0026566	CG1307
FBgn0040534	CG11985	FBgn0032050	CG13096
FBgn0037312	CG11999	FBgn0032051	CG13097
FBgn0027903	CG12018	FBgn0032150	CG13123
FBgn0032915	CG12050	FBgn0033750	CG13151
FBgn0039831	CG12054	FBgn0035526	CG1316
FBgn0030052	CG12065	FBgn0033667	CG13183
FBgn0030065	CG12075	FBgn0035692	CG13298
FBgn0043458	CG12084	FBgn0025634	CG13367
FBgn0035238	CG12104	FBgn0036450	CG13472
FBgn0030100	CG12106	FBgn0039209	CG13624
FBgn0030048	CG12112	FBgn0039210	CG13625
FBgn0033473	CG12128	FBgn0042092	CG13773
FBgn0033475	CG12129	FBgn0040954	CG13779
FBgn0037329	CG12162	FBgn0035165	CG13887
FBgn0037377	CG1218	FBgn0035162	CG13900
FBgn0043796	CG12219	FBgn0031677	CG14036
FBgn0029822	CG12236	FBgn0036814	CG14073
FBgn0039557	CG12259	FBgn0032961	CG1416
FBgn0038057	CG12267	FBgn0031037	CG14207
FBgn0032620	CG12288	FBgn0031036	CG14220
FBgn0032295	CG12299	FBgn0031062	CG14230
FBgn0036514	CG12301	FBgn0031061	CG14232
FBgn0036515	CG12304	FBgn0039504	CG14260
FBgn0038590	CG12320	FBgn0030554	CG1434
FBgn0033557	CG12325	FBgn0036771	CG14353
FBgn0035137	CG1233	FBgn0037183	CG14451
FBgn0037489	CG1234	FBgn0062442	CG1458
FBgn0033556	CG12343	FBgn0031186	CG14614
FBgn0033581	CG12391	FBgn0037220	CG14641
FBgn0035370	CG1240	FBgn0037275	CG14655
FBgn0032465	CG12404	FBgn0037282	CG14657
FBgn0035571	CG12493	FBgn0037317	CG14667
FBgn0038968	CG12499	FBgn0037920	CG14710
FBgn0037811	CG12592	FBgn0037924	CG14712
FBgn0035481	CG12605	FBgn0037930	CG14715
FBgn0030630	CG12608	FBgn0037943	CG14722
FBgn0040929	CG12659	FBgn0037997	CG14742
FBgn0033252	CG12769	FBgn0023514	CG14805
FBgn0033507	CG12909	FBgn0031174	CG1486
FBgn0033569	CG12942	FBgn0038428	CG14894

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0035407	CG14962	FBgn0040754	CG17059
FBgn0035532	CG15014	FBgn0086855	CG17078
FBgn0035541	CG15019	FBgn0032291	CG17118
FBgn0030938	CG15047	FBgn0032299	CG17127
FBgn0034398	CG15098	FBgn0039943	CG17168
FBgn0034400	CG15099	FBgn0038741	CG17186
FBgn0041702	CG15107	FBgn0037882	CG17187
FBgn0033463	CG1513	FBgn0038043	CG17202
FBgn0030322	CG15220	FBgn0036958	CG17233
FBgn0028531	CG15286	FBgn0035249	CG17249
FBgn0031144	CG1529	FBgn0031497	CG17259
FBgn0031377	CG15356	FBgn0033089	CG17266
FBgn0039828	CG1542	FBgn0038829	CG17271
FBgn0031608	CG15435	FBgn0038830	CG17272
FBgn0031161	CG15445	FBgn0036396	CG17359
FBgn0029955	CG15478	FBgn0030293	CG1737
FBgn0034120	CG15710	FBgn0030291	CG1738
FBgn0030374	CG15734	FBgn0033934	CG17385
FBgn0030364	CG15735	FBgn0039977	CG17454
FBgn0030474	CG15747	FBgn0039924	CG17471
FBgn0029999	CG1575	FBgn0030305	CG1749
FBgn0029766	CG15784	FBgn0040010	CG17493
FBgn0035337	CG15877	FBgn0032763	CG17568
FBgn0029860	CG15891	FBgn0035423	CG17737
FBgn0033185	CG1603	FBgn0033529	CG17765
FBgn0033183	CG1620	FBgn0032240	CG17768
FBgn0039600	CG1646	FBgn0038549	CG17802
FBgn0033454	CG1671	FBgn0038548	CG17806
FBgn0034529	CG16742	FBgn0025635	CG17829
FBgn0035393	CG16753	FBgn0030061	CG1785
FBgn0029941	CG1677	FBgn0030063	CG1789
FBgn0003715	CG16778	FBgn0032597	CG17904
FBgn0032488	CG16812	FBgn0032600	CG17912
FBgn0037728	CG16817	FBgn0039860	CG1792
FBgn0036574	CG16838	FBgn0033566	CG18004
FBgn0028931	CG16863	FBgn0033491	CG18011
FBgn0028919	CG16865	FBgn0036035	CG18178
FBgn0040394	CG16903	FBgn0034403	CG18190
FBgn0038464	CG16941	FBgn0029525	CG18273
FBgn0032481	CG16972	FBgn0030269	CG18292
FBgn0033122	CG17002	FBgn0032979	CG1832

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0033458	CG18446	FBgn0031643	CG3008
FBgn0037931	CG18476	FBgn0050122	CG30122
FBgn0038601	CG18600	FBgn0050185	CG30185
FBgn0260444	CG18616	FBgn0050349	CG30349
FBgn0032202	CG18619	FBgn0050382	CG30382
FBgn0042111	CG18766	FBgn0029925	CG3040
FBgn0042125	CG18787	FBgn0050491	CG30491
FBgn0042138	CG18815	FBgn0023527	CG3071
FBgn0033421	CG1888	FBgn0051211	CG31211
FBgn0033434	CG1902	FBgn0051249	CG31249
FBgn0037468	CG1943	FBgn0051301	CG31301
FBgn0037466	CG1965	FBgn0051320	CG31320
FBgn0035271	CG2021	FBgn0051368	CG31368
FBgn0015359	CG2034	FBgn0051388	CG31388
FBgn0037376	CG2051	FBgn0051441	CG31441
FBgn0033400	CG2063	FBgn0051457	CG31457
FBgn0033205	CG2064	FBgn0051510	CG31510
FBgn0030008	CG2129	FBgn0051549	CG31549
FBgn0030243	CG2186	FBgn0051550	CG31550
FBgn0035213	CG2199	FBgn0051612	CG31612
FBgn0030320	CG2247	FBgn0034961	CG3163
FBgn0030000	CG2260	FBgn0051712	CG31712
FBgn0035205	CG2469	FBgn0051739	CG31739
FBgn0032871	CG2611	FBgn0031678	CG31918
FBgn0025838	CG2652	FBgn0051935	CG31935
FBgn0024993	CG2662	FBgn0051955	CG31955
FBgn0014931	CG2678	FBgn0029887	CG3198
FBgn0030504	CG2691	FBgn0052075	CG32075
FBgn0024975	CG2712	FBgn0052176	CG32176
FBgn0027599	CG2790	FBgn0029885	CG3224
FBgn0031266	CG2807	FBgn0029882	CG3226
FBgn0034753	CG2852	FBgn0052264	CG32264
FBgn0031459	CG2862	FBgn0052344	CG32344
FBgn0029672	CG2875	FBgn0052409	CG32409
FBgn0030206	CG2889	FBgn0052425	CG32425
FBgn0023529	CG2918	FBgn0052479	CG32479
FBgn0037342	CG2931	FBgn0052486	CG32486
FBgn0030177	CG2972	FBgn0052500	CG32500
FBgn0029704	CG2982	FBgn0052576	CG32576
FBgn0050007	CG30007	FBgn0052700	CG32700
FBgn0050069	CG30069	FBgn0052708	CG32708

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0260741	CG3281	FBgn0038275	CG3817
FBgn0053107	CG33107	FBgn0032130	CG3838
FBgn0037980	CG3313	FBgn0023520	CG3857
FBgn0053260	CG33260	FBgn0027524	CG3909
FBgn0036018	CG3335	FBgn0038472	CG3995
FBgn0029874	CG3342	FBgn0030418	CG4004
FBgn0034987	CG3363	FBgn0058045	CG40045
FBgn0052831	CG33695	FBgn0034659	CG4021
FBgn0064126	CG33722	FBgn0011824	CG4038
FBgn0250851	CG33981	FBgn0029798	CG4078
FBgn0054008	CG34008	FBgn0028474	CG4119
FBgn0054039	CG34039	FBgn0031256	CG4164
FBgn0031573	CG3407	FBgn0250814	CG4169
FBgn0083968	CG34132	FBgn0038300	CG4203
FBgn0083985	CG34149	FBgn0038302	CG4210
FBgn0085188	CG34159	FBgn0250754	CG42232
FBgn0085192	CG34163	FBgn0259704	CG42358
FBgn0085208	CG34179	FBgn0259720	CG42374
FBgn0085220	CG34191	FBgn0259990	CG42487
FBgn0031229	CG3436	FBgn0260390	CG42516
FBgn0085405	CG34376	FBgn0260953	CG42585
FBgn0085446	CG34417	FBgn0034598	CG4266
FBgn0085451	CG34422	FBgn0261551	CG42669
FBgn0034854	CG3493	FBgn0261562	CG42676
FBgn0034791	CG3501	FBgn0261641	CG42724
FBgn0027571	CG3523	FBgn0034114	CG4282
FBgn0029714	CG3527	FBgn0262104	CG42857
FBgn0035995	CG3529	FBgn0031287	CG4291
FBgn0031492	CG3542	FBgn0034742	CG4294
FBgn0035033	CG3548	FBgn0030455	CG4318
FBgn0031493	CG3605	FBgn0036274	CG4328
FBgn0038461	CG3678	FBgn0263047	CG43342
FBgn0037027	CG3680	FBgn0263110	CG43367
FBgn0035987	CG3689	FBgn0038787	CG4360
FBgn0029824	CG3726	FBgn0032138	CG4364
FBgn0038271	CG3731	FBgn0064122	CG43736
FBgn0034933	CG3735	FBgn0030434	CG4400
FBgn0031657	CG3756	FBgn0031296	CG4415
FBgn0038692	CG3773	FBgn0038765	CG4424
FBgn0034802	CG3800	FBgn0034089	CG44242
FBgn0029861	CG3815	FBgn0050420	CG44247

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0034734	CG4554	FBgn0036754	CG5589
FBgn0037841	CG4565	FBgn0039537	CG5590
FBgn0035016	CG4612	FBgn0038046	CG5641
FBgn0029936	CG4617	FBgn0036258	CG5642
FBgn0035021	CG4622	FBgn0036254	CG5645
FBgn0035036	CG4707	FBgn0037082	CG5664
FBgn0043456	CG4747	FBgn0034313	CG5726
FBgn0032348	CG4751	FBgn0032193	CG5727
FBgn0032354	CG4788	FBgn0039182	CG5728
FBgn0260456	CG4806	FBgn0039186	CG5746
FBgn0039566	CG4849	FBgn0032454	CG5787
FBgn0038766	CG4854	FBgn0032455	CG5792
FBgn0037011	CG4858	FBgn0030855	CG5800
FBgn0034232	CG4866	FBgn0027574	CG5815
FBgn0031318	CG4887	FBgn0032171	CG5846
FBgn0032194	CG4901	FBgn0039385	CG5913
FBgn0028897	CG4935	FBgn0032587	CG5953
FBgn0038768	CG4936	FBgn0038056	CG5961
FBgn0039563	CG4951	FBgn0038927	CG6015
FBgn0039558	CG4980	FBgn0039488	CG6066
FBgn0039554	CG5003	FBgn0038339	CG6118
FBgn0028744	CG5033	FBgn0033859	CG6197
FBgn0038331	CG5073	FBgn0037794	CG6254
FBgn0031317	CG5118	FBgn0036126	CG6272
FBgn0034345	CG5174	FBgn0038316	CG6276
FBgn0031909	CG5181	FBgn0030648	CG6340
FBgn0036994	CG5199	FBgn0263398	CG6364
FBgn0032473	CG5204	FBgn0034269	CG6406
FBgn0037891	CG5214	FBgn0039261	CG6422
FBgn0031912	CG5261	FBgn0032643	CG6453
FBgn0036987	CG5274	FBgn0030933	CG6470
FBgn0032248	CG5343	FBgn0032361	CG6488
FBgn0038950	CG5382	FBgn0030874	CG6506
FBgn0032216	CG5384	FBgn0032363	CG6509
FBgn0036568	CG5389	FBgn0032509	CG6523
FBgn0032476	CG5439	FBgn0030943	CG6540
FBgn0032429	CG5446	FBgn0034210	CG6568
FBgn0039450	CG5484	FBgn0030944	CG6617
FBgn0039160	CG5510	FBgn0037855	CG6621
FBgn0039560	CG5514	FBgn0038301	CG6654
FBgn0032444	CG5525	FBgn0036685	CG6664

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0035902	CG6683	FBgn0034432	CG7461
FBgn0032388	CG6686	FBgn0038108	CG7518
FBgn0037877	CG6689	FBgn0037087	CG7519
FBgn0037878	CG6693	FBgn0035793	CG7546
FBgn0033889	CG6701	FBgn0030990	CG7556
FBgn0032408	CG6712	FBgn0036734	CG7564
FBgn0032298	CG6724	FBgn0037094	CG7611
FBgn0030878	CG6769	FBgn0033548	CG7637
FBgn0037921	CG6808	FBgn0036686	CG7728
FBgn0036405	CG6833	FBgn0032016	CG7818
FBgn0036828	CG6841	FBgn0036124	CG7839
FBgn0036810	CG6885	FBgn0033059	CG7845
FBgn0036490	CG6888	FBgn0026738	CG7857
FBgn0038293	CG6904	FBgn0037549	CG7878
FBgn0031711	CG6907	FBgn0035235	CG7879
FBgn0038989	CG6937	FBgn0039730	CG7903
FBgn0030959	CG6961	FBgn0039735	CG7911
FBgn0027587	CG7028	FBgn0039740	CG7928
FBgn0030086	CG7033	FBgn0036505	CG7945
FBgn0030088	CG7039	FBgn0039743	CG7946
FBgn0030091	CG7065	FBgn0035253	CG7971
FBgn0030963	CG7101	FBgn0038585	CG7993
FBgn0034422	CG7137	FBgn0036096	CG8003
FBgn0038593	CG7146	FBgn0027554	CG8042
FBgn0031947	CG7154	FBgn0038597	CG8064
FBgn0038586	CG7168	FBgn0027567	CG8108
FBgn0035872	CG7185	FBgn0030663	CG8117
FBgn0030894	CG7192	FBgn0030871	CG8142
FBgn0035868	CG7194	FBgn0037702	CG8176
FBgn0038571	CG7215	FBgn0030863	CG8188
FBgn0030081	CG7246	FBgn0037624	CG8223
FBgn0036791	CG7271	FBgn0027607	CG8230
FBgn0036500	CG7275	FBgn0033351	CG8235
FBgn0030969	CG7288	FBgn0033349	CG8243
FBgn0038551	CG7357	FBgn0033342	CG8258
FBgn0036522	CG7372	FBgn0035824	CG8281
FBgn0038546	CG7379	FBgn0030854	CG8289
FBgn0035691	CG7386	FBgn0026573	CG8290
FBgn0037135	CG7414	FBgn0030851	CG8326
FBgn0031979	CG7429	FBgn0037634	CG8359
FBgn0036762	CG7430	FBgn0035707	CG8368

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0034062	CG8388	FBgn0033101	CG9436
FBgn0034084	CG8435	FBgn0031092	CG9577
FBgn0037670	CG8436	FBgn0038166	CG9588
FBgn0038235	CG8461	FBgn0038360	CG9590
FBgn0037746	CG8478	FBgn0037578	CG9601
FBgn0037756	CG8507	FBgn0030787	CG9609
FBgn0033741	CG8545	FBgn0031483	CG9641
FBgn0035714	CG8549	FBgn0037550	CG9667
FBgn0030699	CG8578	FBgn0037583	CG9684
FBgn0027602	CG8611	FBgn0034617	CG9754
FBgn0033317	CG8635	FBgn0037270	CG9769
FBgn0029629	CG8636	FBgn0037609	CG9773
FBgn0026577	CG8677	FBgn0037261	CG9775
FBgn0036900	CG8765	FBgn0027866	CG9776
FBgn0036397	CG8783	FBgn0037621	CG9797
FBgn0028473	CG8801	FBgn0038146	CG9799
FBgn0031476	CG8813	FBgn0031420	CG9866
FBgn0036386	CG8833	FBgn0034814	CG9890
FBgn0031664	CG8892	FBgn0031453	CG9894
FBgn0030710	CG8924	FBgn0030734	CG9911
FBgn0034504	CG8929	FBgn0030738	CG9915
FBgn0030720	CG8939	FBgn0038196	CG9922
FBgn0030680	CG8944	FBgn0036671	CG9951
FBgn0034186	CG8950	FBgn0035371	CG9977
FBgn0034181	CG8963	FBgn0032781	CG9987
FBgn0035318	CG9018	FBgn0021760	chb
FBgn0031764	CG9107	FBgn0000319	Chc
FBgn0030629	CG9123	FBgn0035499	Chd64
FBgn0030791	CG9132	FBgn0014141	cher
FBgn0034496	CG9143	FBgn0029504	CHES-1-like
FBgn0035181	CG9205	FBgn0013764	Chi
FBgn0030659	CG9215	FBgn0000308	chic
FBgn0032925	CG9246	FBgn0086758	chinmo
FBgn0032919	CG9253	FBgn0015371	chn
FBgn0030672	CG9281	FBgn0029709	CHOp24
FBgn0036886	CG9300	FBgn0029503	CHORD
FBgn0034564	CG9344	FBgn0043002	Chrac-14
FBgn0034572	CG9346	FBgn0044324	Chro
FBgn0026582	CG9418	FBgn0040477	cid
FBgn0033092	CG9422	FBgn0027598	cindr
FBgn0032485	CG9426	FBgn0015024	CkIalpha

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0000258	CkIIalpha	FBgn0033890	Ctf4
FBgn0015025	CkIIalpha-i1	FBgn0011760	ctp
FBgn0000259	CkIIbeta	FBgn0262707	CTPsyn
FBgn0037613	Cks85A	FBgn0261268	Cul-3
FBgn0024814	Clc	FBgn0031452	Cwc25
FBgn0020503	CLIP-190	FBgn0004597	CycC
FBgn0015621	Clp	FBgn0022936	CycH
FBgn0034087	clu	FBgn0025674	CycK
FBgn0262975	cnc	FBgn0032210	CYLD
FBgn0000339	cni	FBgn0004432	Cyp1
FBgn0013765	cnn	FBgn0038681	Cyp12a4
FBgn0033265	coil	FBgn0028382	cyp33
FBgn0010434	cora	FBgn0035141	Cypl
FBgn0261573	CoRest	FBgn0000411	D
FBgn0033109	coro	FBgn0000412	D1
FBgn0019624	CoVa	FBgn0027490	D12
FBgn0031830	CoVb	FBgn0022935	D19A
FBgn0013770	Cp1	FBgn0022699	D19B
FBgn0000283	Cp190	FBgn0033015	d4
FBgn0034577	cpa	FBgn0005677	dac
FBgn0011570	cpb	FBgn0020493	Dad
FBgn0027873	Cpsf100	FBgn0028833	Dak1
FBgn0024698	Cpsf160	FBgn0030093	dalao
FBgn0261065	Cpsf73	FBgn0023388	Dap160
FBgn0005585	Crc	FBgn0020305	dbe
FBgn0004396	CrebA	FBgn0004556	Dbp73D
FBgn0014467	CrebB-17A	FBgn0067779	dbr
FBgn0000377	crn	FBgn0002413	dco
FBgn0020309	crol	FBgn0034921	Dcp1
FBgn0001994	crp	FBgn0015075	Ddx1
FBgn0032346	Csl4	FBgn0013799	Deaf1
FBgn0027055	CSN3	FBgn0036038	defl
FBgn0027054	CSN4	FBgn0026533	Dek
FBgn0027053	CSN5	FBgn0028969	deltaCOP
FBgn0028837	CSN6	FBgn0259784	Det
FBgn0028836	CSN7	FBgn0022893	Df31
FBgn0039867	CstF-50	FBgn0027836	Dgp-1
FBgn0027841	CstF-64	FBgn0039710	dgt1
FBgn0004198	ct	FBgn0032390	dgt2
FBgn0020496	CtBP	FBgn0034569	dgt3
FBgn0035769	CTCF	FBgn0026085	dgt4

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0033740	dgt5	FBgn0000520	dwg
FBgn0039638	dgt6	FBgn0000541	E(bx)
FBgn0261797	Dhc64C	FBgn0000581	E(Pc)
FBgn0011274	Dif	FBgn0011586	e(r)
FBgn0030891	dik	FBgn0000591	E(spl)m8-HLH
FBgn0031601	Dim1	FBgn0260243	E(var)3-9
FBgn0023091	dimm	FBgn0000617	e(y)1
FBgn0040467	Dip1	FBgn0000618	e(y)2
FBgn0040466	Dip2	FBgn0087008	e(y)3
FBgn0040465	Dip3	FBgn0000629	E(z)
FBgn0039183	Dis3	FBgn0011766	E2f
FBgn0024432	Dlc90F	FBgn0024371	E2f2
FBgn0030276	Dlic	FBgn0008646	E5
FBgn0034537	DMAP1	FBgn0035624	Eaf6
FBgn0021825	Dmn	FBgn0035063	Eap
FBgn0263106	DnaJ-1	FBgn0026441	ear
FBgn0259676	DNApol-alpha60	FBgn0261954	east
FBgn0005696	DNApol-alpha73	FBgn0027066	Eb1
FBgn0259220	Doa	FBgn0023444	ebi
FBgn0028789	Doc1	FBgn0000543	ecd
FBgn0035956	Doc2	FBgn0000546	EcR
FBgn0035954	Doc3	FBgn0032198	eEF1delta
FBgn0010583	dock	FBgn0000556	Ef1alpha48D
FBgn0015379	dod	FBgn0028737	Ef1beta
FBgn0020306	dom	FBgn0029176	Ef1gamma
FBgn0000482	dor	FBgn0000559	Ef2b
FBgn0011763	Dp	FBgn0086908	egg
FBgn0027835	Dp1	FBgn0261609	eIF-2alpha
FBgn0010109	dpn	FBgn0004926	eIF-2beta
FBgn0032293	Dpy-30L1	FBgn0003600	eIF-2gamma
FBgn0000492	Dr	FBgn0022023	eIF-3p40
FBgn0002183	dre4	FBgn0040227	eIF-3p66
FBgn0033051	dream	FBgn0001942	eIF-4a
FBgn0015664	Dref	FBgn0020660	eIF-4B
FBgn0004638	drk	FBgn0015218	eIF-4E
FBgn0038145	Droj2	FBgn0034967	eIF-5A
FBgn0020304	drongo	FBgn0034858	eIF2B-delta
FBgn0026479	Drp1	FBgn0023512	eIF2B-epsilon
FBgn0010269	Dsor1	FBgn0037249	eIF3-S10
FBgn0011764	Dsp1	FBgn0034258	eIF3-S8
FBgn0000504	dsx	FBgn0034237	eIF3-S9

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0037573	eIF4AIII	FBgn0003062	Fib
FBgn0023213	eIF4G	FBgn0025519	fidipidine
FBgn0030719	eIF5	FBgn0024238	Fim
FBgn0034915	eIF6	FBgn0037255	Fip1
FBgn0000568	Eip75B	FBgn0013269	FK506-bp1
FBgn0004858	elB	FBgn0013954	FK506-bp2
FBgn0020443	Elf	FBgn0010470	Fkbp13
FBgn0039066	EloA	FBgn0029174	FKBP59
FBgn0023212	Elongin-B	FBgn0000662	fl(2)d
FBgn0023211	Elongin-C	FBgn0033806	FLASH
FBgn0037926	Elp1	FBgn0000709	fliI
FBgn0031604	Elp3	FBgn0000711	flw
FBgn0000576	ems	FBgn0028734	Fmr1
FBgn0000577	en	FBgn0262477	FoxP
FBgn0000578	ena	FBgn0004652	fru
FBgn0004875	enc	FBgn0004656	fs(1)h
FBgn0028515	EndoGI	FBgn0000810	fs(1)K10
FBgn0035500	ens	FBgn0262743	Fs(2)Ket
FBgn0035060	Eps-15	FBgn0001078	ftz-f1
FBgn0027496	epsilonCOP	FBgn0029173	fu2
FBgn0036974	eRF1	FBgn0001086	fzy
FBgn0035909	ergic53	FBgn0040372	G9a
FBgn0033663	ERp60	FBgn0031213	galectin
FBgn0000588	esc	FBgn0028968	gammaCop
FBgn0005660	Ets21C	FBgn0028552	gammaSnap
FBgn0039225	Ets96B	FBgn0260639	gammaTub23C
FBgn0004510	Ets97D	FBgn0001092	Gapdh2
FBgn0005659	Ets98B	FBgn0038391	GATAe
FBgn0250753	exba	FBgn0086736	GckIII
FBgn0000611	exd	FBgn0020388	Gcn5
FBgn0260946	exo84	FBgn0004868	Gdi
FBgn0000615	exu	FBgn0032340	Ge-1
FBgn0005558	ey	FBgn0050011	gem
FBgn0037913	fabp	FBgn0033081	geminin
FBgn0014163	fax	FBgn0250732	gfzf
FBgn0039937	fd102C	FBgn0030141	Gga
FBgn0004896	fd59A	FBgn0262126	gho
FBgn0036134	fd68A	FBgn0037551	Gie
FBgn0025832	Fen1	FBgn0033539	Git
FBgn0037937	Fer3	FBgn0259139	glo
FBgn0024891	ferrochelatase	FBgn0015391	glu

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0263097	Glut4EF	FBgn0022740	HLH54F
FBgn0034697	GM130	FBgn0004362	HmgD
FBgn0004913	Gnf1	FBgn0010228	HmgZ
FBgn0001124	Got1	FBgn0085448	Hmx
FBgn0039562	Gp93	FBgn0004914	Hnf4
FBgn0036919	Grasp65	FBgn0015393	hoip
FBgn0001133	grau	FBgn0032250	holn1
FBgn0259211	grh	FBgn0025777	homer
FBgn0026430	Grip84	FBgn0024352	Hop
FBgn0001139	gro	FBgn0017397	how
FBgn0261278	grp	FBgn0030082	HP1b
FBgn0001147	gsb-n	FBgn0039019	HP1c
FBgn0051992	gw	FBgn0261456	hpo
FBgn0001168	h	FBgn0037382	Hpr1
FBgn0001169	H	FBgn0261239	Hr39
FBgn0016660	H15	FBgn0000448	Hr46
FBgn0001170	H2.0	FBgn0004838	Hrb27C
FBgn0032812	Hakai	FBgn0004237	Hrb87F
FBgn0026575	hang	FBgn0001215	Hrb98DE
FBgn0001179	hay	FBgn0015949	hrg
FBgn0001180	hb	FBgn0031450	Hrs
FBgn0008636	hbn	FBgn0001216	Hsc70-1
FBgn0039904	Hcf	FBgn0001217	Hsc70-2
FBgn0025825	Hdac3	FBgn0001218	Hsc70-3
FBgn0026428	HDAC6	FBgn0001219	Hsc70-4
FBgn0014189	Hel25E	FBgn0001220	Hsc70-5
FBgn0022787	Hel89B	FBgn0001222	Hsf
FBgn0011771	Hem	FBgn0001223	Hsp22
FBgn0001185	her	FBgn0001224	Hsp23
FBgn0031107	HERC2	FBgn0001225	Hsp26
FBgn0040318	HGTX	FBgn0001226	Hsp27
FBgn0035142	hipk	FBgn0015245	Hsp60
FBgn0053864	His1:CG33864	FBgn0001230	Hsp68
FBgn0053826	His2A:CG33826	FBgn0013275	Hsp70Aa
FBgn0001197	His2Av	FBgn0001233	Hsp83
FBgn0053882	His2B:CG33882	FBgn0001235	hth
FBgn0053833	His3:CG33833	FBgn0263391	hts
FBgn0053909	His4:CG33909	FBgn0002431	hyd
FBgn0001202	hk	FBgn0037657	hyx
FBgn0001565	Hlc	FBgn0024227	ial
FBgn0011276	HLH3B	FBgn0028546	ics

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0028427	Ilk	FBgn0015721	king-tubby
FBgn0013983	imd	FBgn0086902	kis
FBgn0039139	Ime4	FBgn0010235	Klc
FBgn0262735	Imp	FBgn0030268	Klp10A
FBgn0001258	ImpL3	FBgn0004378	Klp61F
FBgn0260991	Incenp	FBgn0027259	Kmn1
FBgn0025582	Int6	FBgn0001320	kni
FBgn0035462	IntS10	FBgn0001323	knrl
FBgn0039691	IntS11	FBgn0051232	koko
FBgn0039459	IntS12	FBgn0028420	Kr-h1
FBgn0262117	IntS3	FBgn0004167	kst
FBgn0026679	IntS4	FBgn0001324	kto
FBgn0261383	IntS6	FBgn0041627	Ku80
FBgn0025830	IntS8	FBgn0038476	kuk
FBgn0036570	IntS9	FBgn0001330	kz
FBgn0001269	inv	FBgn0001491	l(1)10Bb
FBgn0025366	Ip259	FBgn0027334	l(1)G0004
FBgn0036053	iPLA2-VIA	FBgn0027330	l(1)G0020
FBgn0011774	Irbp	FBgn0027291	l(1)G0156
FBgn0024222	ird5	FBgn0028342	l(1)G0230
FBgn0036999	isoQC	FBgn0010551	l(2)03709
FBgn0011604	Iswi	FBgn0010622	l(2)06496
FBgn0001276	ix	FBgn0022288	l(2)09851
FBgn0040309	Jafrac1	FBgn0261535	l(2)34Fd
FBgn0040308	Jafrac2	FBgn0001986	l(2)35Df
FBgn0039350	jigr1	FBgn0086447	l(2)37Cg
FBgn0001291	Jra	FBgn0086451	l(2)k09022
FBgn0015396	jumu	FBgn0033029	l(2)NC136
FBgn0051363	Jupiter	FBgn0010704	l(2)s5379
FBgn0032704	Jwa	FBgn0010741	l(3)01239
FBgn0024889	Kap-alpha1	FBgn0010926	l(3)07882
FBgn0027338	Kap-alpha3	FBgn0263599	l(3)72Ab
FBgn0087013	Karybeta3	FBgn0263605	l(3)72Dn
FBgn0001297	kay	FBgn0002283	l(3)73Ah
FBgn0022268	KdelR	FBgn0011335	l(3)j2D3
FBgn0037659	Kdm2	FBgn0002441	l(3)mbt
FBgn0033233	Kdm4A	FBgn0086910	l(3)neo38
FBgn0011236	ken	FBgn0011638	La
FBgn0041205	key	FBgn0002525	Lam
FBgn0001308	Khc	FBgn0086372	lap
FBgn0024887	kin17	FBgn0011640	lark

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0261618	larp	FBgn0017578	Max
FBgn0260771	Larp7	FBgn0027950	MBD-like
FBgn0063485	Lasp	FBgn0038016	MBD-R2
FBgn0005654	lat	FBgn0262732	mbf1
FBgn0026634	ldlCp	FBgn0086912	mbm
FBgn0002542	lds	FBgn0026207	mbo
FBgn0031759	lid	FBgn0005536	Mbs
FBgn0020279	lig	FBgn0262559	Mdh2
FBgn0038035	lig3	FBgn0004419	me31B
FBgn0041588	ligatin	FBgn0011655	Med
FBgn0026411	Lim1	FBgn0037109	MED1
FBgn0002023	Lim3	FBgn0036581	MED10
FBgn0035626	lin-28	FBgn0036811	MED11
FBgn0029800	lin-52	FBgn0035145	MED14
FBgn0030274	Lint-1	FBgn0027592	MED15
FBgn0025687	LKR	FBgn0034707	MED16
FBgn0039039	lmd	FBgn0038578	MED17
FBgn0261565	Lmpt	FBgn0026873	MED18
FBgn0020278	loco	FBgn0036761	MED19
FBgn0005630	lola	FBgn0013531	MED20
FBgn0022238	lolal	FBgn0040020	MED21
FBgn0263594	lost	FBgn0040339	MED22
FBgn0067622	LSm-4	FBgn0034795	MED23
FBgn0261067	LSm1	FBgn0035851	MED24
FBgn0051184	LSm3	FBgn0038760	MED25
FBgn0261068	LSm7	FBgn0037359	MED27
FBgn0030142	Lst8	FBgn0039337	MED28
FBgn0029688	lva	FBgn0035149	MED30
FBgn0011648	Mad	FBgn0037262	MED31
FBgn0026326	Mad1	FBgn0035754	MED4
FBgn0035640	mad2	FBgn0024330	MED6
FBgn0034534	maf-S	FBgn0051390	MED7
FBgn0002736	mago	FBgn0034503	MED8
FBgn0034641	mahj	FBgn0011656	Mef2
FBgn0002645	Map205	FBgn0260986	mei-38
FBgn0010342	Map60	FBgn0260856	membrin
FBgn0034282	Mapmodulin	FBgn0035357	MEP-1
FBgn0033845	mars	FBgn0037207	Mes2
FBgn0043884	mask	FBgn0034240	MESR4
FBgn0024956	Mat1	FBgn0002723	Met
FBgn0261286	Mat89Ba	FBgn0035294	Mfap1

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0030731	Mfe2	FBgn0262737	mub
FBgn0086783	Mhc	FBgn0002891	mus205
FBgn0262519	Mi-2	FBgn0005655	mus209
FBgn0033846	mip120	FBgn0004698	mus210
FBgn0023509	mip130	FBgn0002914	Myb
FBgn0034430	mip40	FBgn0086347	Myo31DF
FBgn0035725	Mis12	FBgn0010246	Myo61F
FBgn0004687	Mlc-c	FBgn0028471	Nab2
FBgn0002774	mle	FBgn0086904	Nacalpha
FBgn0014863	Mlp84B	FBgn0015268	Nap1
FBgn0023215	Mnt	FBgn0010488	NAT1
FBgn0017572	Mo25	FBgn0028926	NC2beta
FBgn0039581	Moca-cyp	FBgn0010352	Nc73EF
FBgn0039280	Mocs2	FBgn0002924	ncd
FBgn0002780	mod	FBgn0263510	nclb
FBgn0002781	mod(mdg4)	FBgn0086707	ncm
FBgn0011661	Moe	FBgn0030500	Ndc80
FBgn0014340	mof	FBgn0261617	nej
FBgn0002783	mor	FBgn0038872	Nelf-A
FBgn0020270	mre11	FBgn0027553	NELF-B
FBgn0027378	MRG15	FBgn0017430	Nelf-E
FBgn0033341	MrgBP	FBgn0024542	Neos
FBgn0035107	mri	FBgn0028999	nerfin-1
FBgn0261109	mrn	FBgn0032848	nesd
FBgn0039507	mrt	FBgn0002931	net
FBgn0035209	msd1	FBgn0035993	Nf-YA
FBgn0035210	msd5	FBgn0032816	Nf-YB
FBgn0036486	Msh6	FBgn0029905	Nf-YC
FBgn0011666	msi	FBgn0029148	NHP2
FBgn0026252	msk	FBgn0053554	Nipped-A
FBgn0002775	msl-3	FBgn0026401	Nipped-B
FBgn0010909	msn	FBgn0027548	nito
FBgn0261836	Msp-300	FBgn0021874	Nle
FBgn0027948	msps	FBgn0016685	Nlp
FBgn0027951	MTA1-like	FBgn0039254	Nmnat
FBgn0013756	Mtor	FBgn0022069	Nnp-1
FBgn0028479	Mtpalpha	FBgn0261710	nocte
FBgn0036916	Mtr3	FBgn0014366	noi
FBgn0004177	mts	FBgn0004227	nonA
FBgn0010438	mtSSB	FBgn0015520	nonA-l
FBgn0002872	mu2	FBgn0026196	nop5

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0038964	Nop56	FBgn0261885	osa
FBgn0259937	Nop60B	FBgn0016691	Oscp
FBgn0037137	Nopp140	FBgn0015524	otp
FBgn0085436	Not1	FBgn0003028	ovo
FBgn0039348	Npl4	FBgn0053105	p24-2
FBgn0038473	ns1	FBgn0034259	P32
FBgn0034243	ns2	FBgn0024846	p38b
FBgn0013998	Nsf2	FBgn0033179	p47
FBgn0031145	Ntf-2	FBgn0039044	p53
FBgn0033457	Ntmt	FBgn0037718	P58IPK
FBgn0085424	nub	FBgn0030294	Pa1
FBgn0052190	NUCB1	FBgn0261619	pAbp
FBgn0021768	nudC	FBgn0005648	Pabp2
FBgn0027868	Nup107	FBgn0038418	pad
FBgn0061200	Nup153	FBgn0025809	Paf-AHalpha
FBgn0262647	Nup160	FBgn0260934	par-1
FBgn0010660	Nup214	FBgn0052528	parvin
FBgn0039302	Nup358	FBgn0016693	Past1
FBgn0039301	Nup37	FBgn0028470	Patr-1
FBgn0038609	Nup43	FBgn0263197	Patronin
FBgn0033247	Nup44A	FBgn0011692	pav
FBgn0033264	Nup50	FBgn0041789	Pax
FBgn0034118	Nup62	FBgn0038371	Pbp45
FBgn0034310	Nup75	FBgn0037540	Pbp95
FBgn0039120	Nup98-96	FBgn0003042	Pc
FBgn0016687	Nurf-38	FBgn0051453	pch2
FBgn0005636	nvy	FBgn0036184	PCID2
FBgn0036640	nxf2	FBgn0003044	Pcl
FBgn0028411	Nxt1	FBgn0030520	Pdcd4
FBgn0033901	O-fut1	FBgn0014002	Pdi
FBgn0002985	odd	FBgn0004394	pdm2
FBgn0026058	OdsH	FBgn0261588	pdm3
FBgn0038168	omd	FBgn0260012	pds5
FBgn0003002	opa	FBgn0086895	pea
FBgn0050443	Opbp	FBgn0004181	Peb
FBgn0030606	opm	FBgn0011823	Pen
FBgn0022772	Orc1	FBgn0015527	pen
FBgn0015270	Orc2	FBgn0004401	Pep
FBgn0023181	Orc4	FBgn0032407	Pex19
FBgn0015271	Orc5	FBgn0035405	pfk
FBgn0023180	Orc6	FBgn0003071	Pfk

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0023517	Pgam5	FBgn0030057	Ppt1
FBgn0004654	Pgd	FBgn0003139	PpV
FBgn0003074	Pgi	FBgn0039270	PQBP1
FBgn0250906	Pgk	FBgn0011474	pr-set7
FBgn0014869	Pglym78	FBgn0032059	PrBP
FBgn0011270	Pglym87	FBgn0261837	pre-mod(mdg4)-T
FBgn0004860	ph-d	FBgn0024734	PRL-1
FBgn0025334	PHDP	FBgn0003149	Prm
FBgn0031091	Phf7	FBgn0014269	prod
FBgn0002521	pho	FBgn0004595	pros
FBgn0035997	phol	FBgn0086134	Prosalpha2
FBgn0003082	phr	FBgn0004066	Prosalpha4
FBgn0033669	PI31	FBgn0010590	Prosbeta1
FBgn0020622	Pi3K21B	FBgn0026380	Prosbeta3
FBgn0260962	pic	FBgn0261119	Prp19
FBgn0038966	pinta	FBgn0036915	Prp3
FBgn0025140	pit	FBgn0036487	Prp31
FBgn0016696	Pitslre	FBgn0050342	Prp38
FBgn0037737	Pnn	FBgn0033688	Prp8
FBgn0003117	pnr	FBgn0030329	prtp
FBgn0003118	pnt	FBgn0038570	Prx5
FBgn0013726	pnut	FBgn0261552	ps
FBgn0053526	PNUTS	FBgn0005624	Psc
FBgn0036354	Poc1	FBgn0014870	Psi
FBgn0011230	poe	FBgn0263102	psq
FBgn0027559	Pol32	FBgn0052133	ptip
FBgn0003124	polo	FBgn0028577	pUf68
FBgn0039227	polybromo	FBgn0003165	pum
FBgn0040078	pont	FBgn0022361	Pur-alpha
FBgn0036239	Pop2	FBgn0003178	PyK
FBgn0004363	porin	FBgn0259785	pzg
FBgn0003130	Poxn	FBgn0022987	qkr54B
FBgn0003132	Pp1-13C	FBgn0022986	qkr58E-1
FBgn0004103	Pp1-87B	FBgn0022985	qkr58E-2
FBgn0003134	Pp1alpha-96A	FBgn0022984	qkr58E-3
FBgn0260439	Pp2A-29B	FBgn0004636	R
FBgn0042693	PP2A-B'	FBgn0003189	r
FBgn0023177	Pp4-19C	FBgn0016700	Rab1
FBgn0010770	ppan	FBgn0015789	Rab10
FBgn0027945	ppl	FBgn0015790	Rab11
FBgn0030208	PPP4R2r	FBgn0014009	Rab2

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0031090	Rab35	FBgn0028700	RfC38
FBgn0029959	Rab39	FBgn0260985	RfC4
FBgn0014010	Rab5	FBgn0017550	Rga
FBgn0015795	Rab7	FBgn0260442	rhea
FBgn0262518	Rab8	FBgn0038747	RhoGAP92B
FBgn0030221	Rab9Db	FBgn0023172	RhoGEF2
FBgn0020618	Rack1	FBgn0003254	rib
FBgn0026777	Rad23	FBgn0050085	Rif1
FBgn0034728	rad50	FBgn0015778	rin
FBgn0034646	Rae1	FBgn0027335	Rip11
FBgn0036624	RAF2	FBgn0014022	Rlb1
FBgn0020255	Ran	FBgn0003261	Rm62
FBgn0053180	Ranbp16	FBgn0030753	rngo
FBgn0262114	RanBPM	FBgn0023171	rnh1
FBgn0003346	RanGAP	FBgn0037707	RnpS1
FBgn0040080	raps	FBgn0011703	RnrL
FBgn0003205	Ras85D	FBgn0011704	RnrS
FBgn0031868	Rat1	FBgn0024196	robl
FBgn0004903	Rb97D	FBgn0036697	rogdi
FBgn0036973	Rbbp5	FBgn0039152	Rootletin
FBgn0023458	Rbcn-3A	FBgn0004574	Rop
FBgn0023510	Rbcn-3B	FBgn0036621	roq
FBgn0015799	Rbf	FBgn0033998	row
FBgn0030067	Rbm13	FBgn0010173	RpA-70
FBgn0260944	Rbp1	FBgn0032906	RPA2
FBgn0030479	Rbp1-like	FBgn0039218	Rpb10
FBgn0262734	Rbp2	FBgn0032634	Rpb11
FBgn0261064	Rbsn-5	FBgn0262954	Rpb12
FBgn0002638	Rcc1	FBgn0033571	Rpb5
FBgn0031047	Rcd-1	FBgn0051155	Rpb7
FBgn0033897	Rcd1	FBgn0037121	Rpb8
FBgn0035489	Rcd5	FBgn0015805	Rpd3
FBgn0262907	rdx	FBgn0019938	RpI1
FBgn0003231	ref(2)P	FBgn0038903	RpI12
FBgn0010774	Ref1	FBgn0003278	RpI135
FBgn0029133	REG	FBgn0262955	RpII140
FBgn0014018	Rel	FBgn0003275	RpII18
FBgn0011701	repo	FBgn0003277	RpII215
FBgn0032341	Reps	FBgn0026373	RpII33
FBgn0040075	rept	FBgn0004463	RpIII128
FBgn0032244	RfC3	FBgn0024733	RpL10

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0036213	RpL10Ab	FBgn0015756	RpL9
FBgn0013325	RpL11	FBgn0000100	RpLP0
FBgn0034968	RpL12	FBgn0033485	RpLP0-like
FBgn0011272	RpL13	FBgn0002593	RpLP1
FBgn0037351	RpL13A	FBgn0003274	RpLP2
FBgn0017579	RpL14	FBgn0028695	Rpn1
FBgn0028697	RpL15	FBgn0015283	Rpn10
FBgn0029897	RpL17	FBgn0028694	Rpn11
FBgn0035753	RpL18	FBgn0028693	Rpn12
FBgn0010409	RpL18A	FBgn0033886	Rpn13
FBgn0002607	RpL19	FBgn0028692	Rpn2
FBgn0032987	RpL21	FBgn0261396	Rpn3
FBgn0015288	RpL22	FBgn0028690	Rpn5
FBgn0010078	RpL23	FBgn0028689	Rpn6
FBgn0026372	RpL23A	FBgn0028688	Rpn7
FBgn0032518	RpL24	FBgn0002787	Rpn8
FBgn0037899	RpL24-like	FBgn0028691	Rpn9
FBgn0036825	RpL26	FBgn0022246	Rpp30
FBgn0039359	RpL27	FBgn0027494	RpS10a
FBgn0261606	RpL27A	FBgn0261593	RpS10b
FBgn0035422	RpL28	FBgn0033699	RpS11
FBgn0016726	RpL29	FBgn0260441	RpS12
FBgn0020910	RpL3	FBgn0010265	RpS13
FBgn0086710	RpL30	FBgn0004404	RpS14b
FBgn0025286	RpL31	FBgn0034138	RpS15
FBgn0002626	RpL32	FBgn0010198	RpS15Aa
FBgn0037686	RpL34b	FBgn0033555	RpS15Ab
FBgn0029785	RpL35	FBgn0034743	RpS16
FBgn0037328	RpL35A	FBgn0005533	RpS17
FBgn0002579	RpL36	FBgn0010411	RpS18
FBgn0031980	RpL36A	FBgn0010412	RpS19a
FBgn0030616	RpL37a	FBgn0039129	RpS19b
FBgn0261608	RpL37A	FBgn0004867	RpS2
FBgn0040007	RpL38	FBgn0019936	RpS20
FBgn0003279	RpL4	FBgn0015521	RpS21
FBgn0064225	RpL5	FBgn0033912	RpS23
FBgn0039857	RpL6	FBgn0261596	RpS24
FBgn0005593	RpL7	FBgn0086472	RpS25
FBgn0032404	RpL7-like	FBgn0261597	RpS26
FBgn0014026	RpL7A	FBgn0039300	RpS27
FBgn0261602	RpL8	FBgn0003942	RpS27A

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0030136	RpS28b	FBgn0030788	Sap30
FBgn0261599	RpS29	FBgn0038947	Sar1
FBgn0002622	RpS3	FBgn0029755	Sas10
FBgn0038834	RpS30	FBgn0010575	sbb
FBgn0017545	RpS3A	FBgn0003321	sbr
FBgn0011284	RpS4	FBgn0040286	SC35
FBgn0261592	RpS6	FBgn0261872	scaf6
FBgn0039757	RpS7	FBgn0040285	Scamp
FBgn0039713	RpS8	FBgn0041781	SCAR
FBgn0010408	RpS9	FBgn0003330	Sce
FBgn0028687	Rpt1	FBgn0025682	scf
FBgn0015282	Rpt2	FBgn0003334	Scm
FBgn0028686	Rpt3	FBgn0260936	scny
FBgn0028685	Rpt4	FBgn0261385	scra
FBgn0028684	Rpt5	FBgn0003345	sd
FBgn0020369	Rpt6	FBgn0010415	Sdc
FBgn0039788	Rpt6R	FBgn0053497	Sdic2
FBgn0004584	Rrp1	FBgn0024509	Sec13
FBgn0034879	Rrp4	FBgn0052654	Sec16
FBgn0260648	Rrp40	FBgn0260855	Sec22
FBgn0034065	Rrp42	FBgn0262125	Sec23
FBgn0030789	Rrp45	FBgn0033460	sec24
FBgn0037815	Rrp46	FBgn0033339	sec31
FBgn0030711	Rrp47	FBgn0031537	sec5
FBgn0038269	Rrp6	FBgn0263260	sel
FBgn0021995	Rs1	FBgn0261270	SelD
FBgn0011305	Rsf1	FBgn0002573	sens
FBgn0020909	Rtc1	FBgn0003360	sesB
FBgn0034722	Rtf1	FBgn0014879	Set
FBgn0015803	RtGEF	FBgn0040022	Set1
FBgn0260010	rump	FBgn0025571	SF1
FBgn0003300	run	FBgn0040284	SF2
FBgn0025381	rush	FBgn0032475	Sfmbt
FBgn0020617	Rx	FBgn0036804	Sgf11
FBgn0034763	RYBP	FBgn0050390	Sgf29
FBgn0020616	SA	FBgn0032640	Sgt
FBgn0002842	sa	FBgn0260939	Sgt1
FBgn0039229	Saf-B	FBgn0035772	Sh3beta
FBgn0005278	Sam-S	FBgn0015296	Shc
FBgn0024188	san	FBgn0052423	shep
FBgn0262714	Sap130	FBgn0003392	shi

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0003396	shn	FBgn0032005	Snx6
FBgn0013733	shot	FBgn0004892	sob
FBgn0086656	shrb	FBgn0003462	Sod
FBgn0032741	Side	FBgn0036411	Sox21a
FBgn0004666	sim	FBgn0029123	SoxN
FBgn0010762	simj	FBgn0015544	spag
FBgn0022764	Sin3A	FBgn0037025	Spc105R
FBgn0024191	sip1	FBgn0015546	spel1
FBgn0031878	sip2	FBgn0086683	Spf45
FBgn0003411	sisA	FBgn0029764	spoon
FBgn0003415	skd	FBgn0037981	Spt3
FBgn0032487	Ski6	FBgn0028683	spt4
FBgn0025637	skpA	FBgn0040273	Spt5
FBgn0041186	Slbp	FBgn0028982	Spt6
FBgn0037810	sle	FBgn0015818	Spx
FBgn0015816	Slh	FBgn0263396	sqd
FBgn0261477	slim	FBgn0003514	sqh
FBgn0023423	slmb	FBgn0038320	Sra-1
FBgn0003430	slp1	FBgn0036340	SRm160
FBgn0039626	Slu7	FBgn0003507	srp
FBgn0003435	sm	FBgn0015298	Srp19
FBgn0262601	SmB	FBgn0024285	Srp54
FBgn0040283	SMC1	FBgn0010747	Srp54k
FBgn0027783	SMC2	FBgn0035947	Srp68
FBgn0261933	SmD1	FBgn0038810	Srp72
FBgn0261789	SmD2	FBgn0035827	Srp9
FBgn0023167	SmD3	FBgn0003511	Sry-beta
FBgn0261790	SmE	FBgn0003512	Sry-delta
FBgn0000426	SmF	FBgn0011481	Ssdp
FBgn0261791	SmG	FBgn0037202	Ssl1
FBgn0016983	smid	FBgn0036248	ssp
FBgn0025800	Smox	FBgn0036389	ssp2
FBgn0024308	Smr	FBgn0011016	SsRbeta
FBgn0026170	smt3	FBgn0010278	Ssrp
FBgn0086129	snama	FBgn0024987	ssx
FBgn0250791	Snap	FBgn0003517	sta
FBgn0003449	snf	FBgn0027363	Stam
FBgn0011715	Snr1	FBgn0016917	Stat92E
FBgn0016978	snRNP-U1-70K	FBgn0020249	stck
FBgn0031534	Snx1	FBgn0002466	sti
FBgn0038065	Snx3	FBgn0003459	stwl

Supplemental Table 2.1 Continued

Flybase ID	Gene Symbol	Flybase ID	Gene Symbol
FBgn0003559	su(f)	FBgn0034451	TBCB
FBgn0004837	Su(H)	FBgn0003687	Tbp
FBgn0003567	su(Hw)	FBgn0025790	TBPH
FBgn0003612	Su(var)2-10	FBgn0037632	Tcp-1eta
FBgn0026427	Su(var)2-HP2	FBgn0027329	Tcp-1zeta
FBgn0003607	Su(var)205	FBgn0037874	Tctp
FBgn0260397	Su(var)3-3	FBgn0041180	Tep4
FBgn0003638	su(w[a])	FBgn0261014	TER94
FBgn0020887	Su(z)12	FBgn0037569	tex
FBgn0008654	Su(z)2	FBgn0033929	Tfb1
FBgn0037462	sunz	FBgn0031309	Tfb4
FBgn0019925	Surf4	FBgn0011289	TfIIA-L
FBgn0038746	Surf6	FBgn0013347	TfIIA-S
FBgn0003651	svp	FBgn0004915	TfIIB
FBgn0003654	sw	FBgn0015828	TfIIEalpha
FBgn0002044	swm	FBgn0015829	TfIIEbeta
FBgn0261403	sxc	FBgn0010282	TfIIFalpha
FBgn0003660	Syb	FBgn0010421	TfIIFbeta
FBgn0037371	Sym	FBgn0010422	TfIIS
FBgn0038826	Syp	FBgn0015014	tgo
FBgn0003676	T-cp1	FBgn0010416	TH1
FBgn0086358	Tab2	FBgn0031390	tho2
FBgn0026620	tacc	FBgn0034939	thoc5
FBgn0010355	Taf1	FBgn0036263	thoc6
FBgn0028398	Taf10	FBgn0035110	thoc7
FBgn0026324	Taf10b	FBgn0027360	Tim10
FBgn0011291	Taf11	FBgn0027359	Tim8
FBgn0011290	Taf12	FBgn0030480	Tim9a
FBgn0032847	Taf13	FBgn0026080	Tip60
FBgn0011836	Taf2	FBgn0086899	tlk
FBgn0010280	Taf4	FBgn0003721	Tm1
FBgn0010356	Taf5	FBgn0082582	tmod
FBgn0010417	Taf6	FBgn0036285	toe
FBgn0024909	Taf7	FBgn0004924	Top1
FBgn0022724	Taf8	FBgn0003732	Top2
FBgn0030365	Tango4	FBgn0021796	Tor
FBgn0033902	Tango7	FBgn0033636	tou
FBgn0051852	Tap42	FBgn0032586	Tpr2
FBgn0021795	Tapdelta	FBgn0003742	tra2
FBgn0040071	tara	FBgn0041775	tral
FBgn0260938	tay	FBgn0026761	Trap1

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**	**Flybase ID**	**Gene Symbol**
FBgn0010287	Trf	FBgn0022097	Vha36-1
FBgn0261793	Trf2	FBgn0262511	Vha44
FBgn0015834	Trip1	FBgn0005671	Vha55
FBgn0013263	Trl	FBgn0263598	Vha68-2
FBgn0260861	Trs23	FBgn0036237	viaf
FBgn0036666	TSG101	FBgn0262468	vib
FBgn0003866	tsh	FBgn0024183	vig
FBgn0011726	tsr	FBgn0046214	vig2
FBgn0033378	tsu	FBgn0027936	vih
FBgn0030502	tth	FBgn0022960	vimar
FBgn0003870	ttk	FBgn0004397	Vinc
FBgn0035121	Tudor-SN	FBgn0003977	vir
FBgn0003896	tup	FBgn0033748	vis
FBgn0003900	twi	FBgn0052418	vito
FBgn0011725	twin	FBgn0263251	vnc
FBgn0004889	tws	FBgn0033194	Vps13
FBgn0035631	Txl	FBgn0014411	Vps26
FBgn0026083	tyf	FBgn0021814	Vps28
FBgn0033210	U2A	FBgn0027605	Vps4
FBgn0017457	U2af38	FBgn0261049	Vps45
FBgn0005411	U2af50	FBgn0016076	vri
FBgn0053505	U3-55K	FBgn0263512	Vsx2
FBgn0036733	U4-U6-60K	FBgn0086680	vvl
FBgn0023143	Uba1	FBgn0035120	wac
FBgn0003943	Ubi-p63E	FBgn0262527	wah
FBgn0031057	Ubqn	FBgn0004655	wapl
FBgn0003944	Ubx	FBgn0033692	wash
FBgn0010288	Uch	FBgn0262560	wcd
FBgn0011327	Uch-L5	FBgn0039067	wda
FBgn0030370	Uch-L5R	FBgn0005642	wdn
FBgn0035601	Uev1A	FBgn0032030	Wdr82
FBgn0036136	Ufd1-like	FBgn0040066	wds
FBgn0004395	unk	FBgn0001990	wek
FBgn0263352	Unr	FBgn0034876	wmd
FBgn0035025	uri	FBgn0010328	woc
FBgn0003963	ush	FBgn0028554	x16
FBgn0003964	usp	FBgn0039338	XNP
FBgn0260749	Utx	FBgn0261850	Xpd
FBgn0039269	veli	FBgn0026751	XRCC1
FBgn0262524	ver	FBgn0043842	Yeti
FBgn0015324	Vha26	FBgn0026749	Yippee

Supplemental Table 2.1 Continued

Flybase ID	**Gene Symbol**
FBgn0034970	yki
FBgn0032321	YL-1
FBgn0022959	yps
FBgn0027616	YT521-B
FBgn0021895	ytr
FBgn0004050	z
FBgn0052685	ZAP3
FBgn0083919	Zasp52
FBgn0004053	zen
FBgn0040512	zetaCOP
FBgn0022720	zf30C
FBgn0004606	zfh1
FBgn0037446	Zif
FBgn0005634	zip
FBgn0263603	Zn72D
FBgn0030096	Zpr1
FBgn0061476	zwilch

Chapter 3

Supplemental Figure 3.1 (electronic) contains a Cytoscape file containing the tissue specific network analysis. Supplemental Figure 3.2 (electronic) contains a Cytoscape file with the supervised and unsupervised integrated network analysis. Supplemental Table 3.1 is a list of all proteins in the high confidence interaction network, scored using tissue specificity score (TSPS). Supplemental Table 3.2 (electronic) is an Excel file containing the tissue specificity assignments. Supplemental Table 3.3 (electronic) is an Excel file containing all the nodes for each tissue specific network. Supplemental Table 3.4 (electronic) is an Excel file containing all shared targets between interacting TFs. Supplemental Table 3.5 (electronic) is an Excel file containing all instances of the transcriptional regulatory motifs from the integrated network analysis.

Supplemental Table 3.1 TSPS Scored Proteins

A table containing all proteins from the high confidence interaction network, the corresponding tissue specificity score and the specificity group that each proteins falls into.

Gene Symbol	Flybase ID	TSPS	Specificity Bin
CG14260	FBgn0039504	3.402211	High
betaTub85D	FBgn0003889	3.235619	High
CG17127	FBgn0032299	3.19419	High
CG17118	FBgn0032291	3.162656	High
nerfin-1	FBgn0028999	3.087146	High
bcd	FBgn0000166	2.92827	High
CG2652	FBgn0025838	2.869528	High
CG15286	FBgn0028531	2.849702	High
RpS19b	FBgn0039129	2.80711	High
Act79B	FBgn0000045	2.610427	High
al	FBgn0000061	2.456623	High
Ant2	FBgn0025111	2.452437	High
CG15047	FBgn0030938	2.387919	High
dpn	FBgn0010109	2.381318	High
vis	FBgn0033748	2.316395	High
rib	FBgn0003254	2.267388	High
cid	FBgn0040477	2.2666	High
CG42857	FBgn0262104	2.236477	High
CG14451	FBgn0037183	2.234482	High
hb	FBgn0001180	2.2088	High
CG5204	FBgn0032473	2.204518	High
sunz	FBgn0037462	2.190314	High
SoxN	FBgn0029123	2.146592	High
CG4415	FBgn0031296	2.145169	High
mei-38	FBgn0260986	2.118121	High
GATAe	FBgn0038391	2.087002	High
sa	FBgn0002842	2.070977	High
CG9576	FBgn0031091	2.064074	High
CG10918	FBgn0031178	2.06104	High
CG16838	FBgn0036574	2.025879	High
CG15734	FBgn0030374	1.958918	High
Doc1	FBgn0028789	1.955515	High
Hsp23	FBgn0001224	1.918802	High
fzy	FBgn0001086	1.914295	High
lin-28	FBgn0035626	1.859709	High
toe	FBgn0036285	1.842211	High
CG8478	FBgn0037746	1.832647	High
msd1	FBgn0035209	1.808326	High
CG8117	FBgn0030663	1.794007	High
Peb	FBgn0004181	1.782664	High
sisA	FBgn0003411	1.770145	High

Supplemental Table 3.1 Continued

E(spl)	FBgn0000591	1.758087	High
Sox21a	FBgn0036411	1.749953	High
cad	FBgn0000251	1.73576	High
Spc105R	FBgn0037025	1.702963	High
OdsH	FBgn0026058	1.669813	High
Hr46	FBgn0000448	1.651436	High
en	FBgn0000577	1.636496	High
nub	FBgn0085424	1.629565	High
CG16972	FBgn0032481	1.61095	High
ovo	FBgn0003028	1.604128	High
Ubx	FBgn0003944	1.516771	High
CG11294	FBgn0030058	1.514692	High
SMC2	FBgn0027783	1.512524	High
Rab9Db	FBgn0030221	1.502993	High
jumu	FBgn0015396	1.472634	High
twi	FBgn0003900	1.45904	High
CG7372	FBgn0036522	1.451602	High
sens	FBgn0002573	1.450008	High
mu2	FBgn0002872	1.449963	High
Doc3	FBgn0035954	1.433114	High
Irbp	FBgn0011774	1.41088	High
Prm	FBgn0003149	1.402176	High
odd	FBgn0002985	1.386128	High
mus309	FBgn0002906	1.361422	High
Mhc	FBgn0086783	1.354061	High
run	FBgn0003300	1.34278	High
phr	FBgn0003082	1.329612	High
CG42374	FBgn0259720	1.317671	High
rdx	FBgn0262907	1.280413	High
CG17802	FBgn0038549	1.255355	High
CG12942	FBgn0033569	1.25515	High
bin	FBgn0045759	1.241559	High
Df31	FBgn0022893	1.239521	High
esc	FBgn0000588	1.236276	High
dimm	FBgn0023091	1.233353	High
inv	FBgn0001269	1.21797	High
HLH54F	FBgn0022740	1.215623	High
CG5199	FBgn0036994	1.197013	High
sob	FBgn0004892	1.191442	High
Fen1	FBgn0025832	1.185704	High
uri	FBgn0035025	1.175876	High
HLH3B	FBgn0011276	1.170332	Mid

Supplemental Table 3.1 Continued

tra2	FBgn0003742	1.154453	Mid
D12	FBgn0027490	1.152831	Mid
H15	FBgn0016660	1.144378	Mid
CG10440	FBgn0034636	1.144047	Mid
Hsp27	FBgn0001226	1.138696	Mid
Map60	FBgn0010342	1.13333	Mid
Hsp26	FBgn0001225	1.101763	Mid
fd59A	FBgn0004896	1.099617	Mid
Mlp84B	FBgn0014863	1.091103	Mid
dac	FBgn0005677	1.086116	Mid
jigr1	FBgn0039350	1.07733	Mid
Jafrac2	FBgn0040308	1.070012	Mid
CG4707	FBgn0035036	1.068773	Mid
CG10959	FBgn0030010	1.05755	Mid
Kmn1	FBgn0027259	1.05457	Mid
Pcl	FBgn0003044	1.053532	Mid
CG42487	FBgn0259990	1.052137	Mid
tsh	FBgn0003866	1.04931	Mid
Zasp52	FBgn0083919	1.046752	Mid
CG7271	FBgn0036791	1.037779	Mid
CG17359	FBgn0036396	1.018198	Mid
CG15478	FBgn0029955	1.01739	Mid
CG14742	FBgn0037997	1.01671	Mid
achi	FBgn0033749	1.005527	Mid
CG14712	FBgn0037924	1.004333	Mid
CG7386	FBgn0035691	0.990494	Mid
Side	FBgn0032741	0.984178	Mid
pdm3	FBgn0261588	0.977746	Mid
CG3838	FBgn0032130	0.975759	Mid
caup	FBgn0015919	0.972204	Mid
CG15710	FBgn0034120	0.958581	Mid
CG2129	FBgn0030008	0.953378	Mid
ftz-f1	FBgn0001078	0.953053	Mid
Vsx2	FBgn0263512	0.951946	Mid
dco	FBgn0002413	0.951744	Mid
Kdm4A	FBgn0033233	0.951369	Mid
Br140	FBgn0033155	0.946818	Mid
spel1	FBgn0015546	0.946201	Mid
CG7101	FBgn0030963	0.94614	Mid
grh	FBgn0259211	0.934301	Mid
C15	FBgn0004863	0.926386	Mid
phol	FBgn0035997	0.925904	Mid

Supplemental Table 3.1 Continued

CG9123	FBgn0030629	0.919134	Mid
CG12769	FBgn0033252	0.91695	Mid
ey	FBgn0005558	0.908134	Mid
Psc	FBgn0005624	0.903983	Mid
E2f2	FBgn0024371	0.895311	Mid
CG9609	FBgn0030787	0.885389	Mid
Adh	FBgn0000055	0.882474	Mid
Poxn	FBgn0003130	0.873721	Mid
CG12659	FBgn0040929	0.870024	Mid
CycH	FBgn0022936	0.867439	Mid
CG7246	FBgn0030081	0.856109	Mid
CG3975	FBgn0027559	0.848754	Mid
BCL7-like	FBgn0026149	0.837404	Mid
CG8944	FBgn0030680	0.836704	Mid
Atf-2	FBgn0050420	0.836134	Mid
CG4854	FBgn0038766	0.827171	Mid
Ku80	FBgn0041627	0.824816	Mid
CG15107	FBgn0041702	0.815451	Mid
Xpd	FBgn0261850	0.810569	Mid
ct	FBgn0004198	0.810157	Mid
CG6540	FBgn0030943	0.8099	Mid
D	FBgn0000411	0.809752	Mid
sim	FBgn0004666	0.806786	Mid
FoxP	FBgn0262477	0.806633	Mid
CG12112	FBgn0030048	0.806595	Mid
mrn	FBgn0261109	0.8003	Mid
Dip1	FBgn0040467	0.800039	Mid
grp	FBgn0261278	0.799319	Mid
Rx	FBgn0020617	0.794784	Mid
Trf	FBgn0010287	0.79374	Mid
chn	FBgn0015371	0.787073	Mid
ade5	FBgn0020513	0.786135	Mid
CG6808	FBgn0037921	0.784478	Mid
pnr	FBgn0003117	0.784385	Mid
pr-set7	FBgn0011474	0.784234	Mid
CG2712	FBgn0024975	0.782554	Mid
CG12299	FBgn0032295	0.78253	Mid
tup	FBgn0003896	0.782384	Mid
PHDP	FBgn0025334	0.78162	Mid
dom	FBgn0020306	0.78128	Mid
Mta70	FBgn0039139	0.781007	Mid
Set	FBgn0014879	0.780209	Mid

Supplemental Table 3.1 Continued

Dip3	FBgn0040465	0.773208	Mid
wdn	FBgn0005642	0.765253	Mid
pinta	FBgn0038966	0.763101	Mid
CG6693	FBgn0037878	0.762802	Mid
CG31457	FBgn0051457	0.762631	Mid
pad	FBgn0038418	0.760528	Mid
Hsp22	FBgn0001223	0.755672	Mid
H	FBgn0001169	0.75388	Mid
CG10395	FBgn0033019	0.752967	Mid
pdm2	FBgn0004394	0.752386	Mid
CG2611	FBgn0032871	0.749811	Mid
srp	FBgn0003507	0.739729	Mid
hang	FBgn0026575	0.736038	Mid
slp1	FBgn0003430	0.728714	Mid
Nf-YC	FBgn0029905	0.724797	Mid
CG14655	FBgn0037275	0.722328	Mid
msl-3	FBgn0002775	0.719092	Mid
gsb-n	FBgn0001147	0.718412	Mid
CG10979	FBgn0037379	0.716482	Mid
woc	FBgn0010328	0.714973	Mid
Lmpt	FBgn0261565	0.713384	Mid
nvy	FBgn0005636	0.712479	Mid
Tbp	FBgn0003687	0.710463	Mid
B-H2	FBgn0004854	0.710173	Mid
CG12325	FBgn0033557	0.708435	Mid
CG43342	FBgn0263047	0.708413	Mid
Actn	FBgn0000667	0.70247	Mid
Cdk7	FBgn0263237	0.697637	Mid
Rbf	FBgn0015799	0.696272	Mid
E(bx)	FBgn0000541	0.691309	Mid
Bgb	FBgn0013753	0.691027	Mid
r	FBgn0003189	0.690022	Mid
CG4565	FBgn0037841	0.688012	Mid
vvl	FBgn0086680	0.685101	Mid
CG31612	FBgn0051612	0.684524	Mid
Nipped-A	FBgn0053554	0.672053	Mid
Su(H)	FBgn0004837	0.670603	Mid
mahj	FBgn0034641	0.662642	Mid
bsh	FBgn0000529	0.661242	Mid
CG1738	FBgn0030291	0.654851	Mid
spag	FBgn0015544	0.652776	Mid
MED21	FBgn0040020	0.652422	Mid

Supplemental Table 3.1 Continued

Sym	FBgn0037371	0.64662	Mid
Lim3	FBgn0002023	0.646076	Mid
CG3726	FBgn0029824	0.642033	Mid
CG8950	FBgn0034186	0.641307	Mid
H2.0	FBgn0001170	0.639246	Mid
Atg5	FBgn0029943	0.638347	Mid
CG5953	FBgn0032587	0.635807	Mid
Ets98B	FBgn0005659	0.63519	Mid
CG4936	FBgn0038768	0.634092	Mid
koko	FBgn0051232	0.633264	Mid
msk	FBgn0026252	0.624556	Mid
otp	FBgn0015524	0.624285	Mid
CG3407	FBgn0031573	0.622833	Mid
Cbp20	FBgn0022943	0.621894	Mid
mof	FBgn0014340	0.617538	Mid
CG32425	FBgn0052425	0.616217	Mid
svp	FBgn0003651	0.610021	Mid
pfk	FBgn0035405	0.608681	Mid
Cp190	FBgn0000283	0.608538	Mid
king-tubby	FBgn0015721	0.607468	Mid
bon	FBgn0023097	0.604322	Mid
su(Hw)	FBgn0003567	0.598735	Mid
cathD	FBgn0029093	0.598467	Mid
LSm7	FBgn0261068	0.594851	Mid
Sdc	FBgn0010415	0.589504	Mid
CG10654	FBgn0036294	0.586572	Mid
Mad1	FBgn0026326	0.586115	Mid
sesB	FBgn0003360	0.582176	Mid
CG9754	FBgn0034617	0.580734	Mid
ppl	FBgn0027945	0.580659	Mid
Chro	FBgn0044324	0.58062	Mid
CG3363	FBgn0034987	0.57996	Mid
NELF-B	FBgn0027553	0.579174	Mid
B-H1	FBgn0011758	0.577723	Mid
Dr	FBgn0000492	0.575395	Mid
E(Pc)	FBgn0000581	0.57165	Mid
CG7357	FBgn0038551	0.570993	Mid
Bx	FBgn0000242	0.569231	Mid
Cul-3	FBgn0261268	0.564957	Mid
CG8833	FBgn0036386	0.564142	Mid
tws	FBgn0004889	0.563023	Mid
Max	FBgn0017578	0.562241	Mid

Supplemental Table 3.1 Continued

vig2	FBgn0046214	0.560138	Mid
wda	FBgn0039067	0.558622	Mid
CG7818	FBgn0032016	0.558583	Mid
kni	FBgn0001320	0.557289	Mid
CTPsyn	FBgn0262707	0.55688	Mid
CG7429	FBgn0031979	0.556824	Mid
EcR	FBgn0000546	0.554341	Mid
hth	FBgn0001235	0.551041	Mid
CG16753	FBgn0035393	0.544225	Mid
Cdk8	FBgn0015618	0.543988	Mid
Ge-1	FBgn0032340	0.542917	Mid
CG15445	FBgn0031161	0.542791	Mid
Nc73EF	FBgn0010352	0.54019	Mid
RpII18	FBgn0003275	0.539112	Mid
CG8783	FBgn0036397	0.538002	Mid
Arp5	FBgn0038576	0.536999	Mid
CG9922	FBgn0038196	0.536206	Mid
brk	FBgn0024250	0.535507	Mid
CG9894	FBgn0031453	0.534899	Mid
mus210	FBgn0004698	0.523058	Mid
Blimp-1	FBgn0035625	0.521105	Mid
CG12106	FBgn0030100	0.519254	Mid
Mi-2	FBgn0262519	0.517987	Mid
TfIIFbeta	FBgn0010421	0.517783	Mid
l(3)73Ah	FBgn0002283	0.516973	Mid
CG1234	FBgn0037489	0.516699	Mid
Ald	FBgn0000064	0.515218	Mid
Hr39	FBgn0261239	0.513933	Mid
RpII215	FBgn0003277	0.513008	Mid
CG3773	FBgn0038692	0.512541	Mid
Nf-YA	FBgn0035993	0.51187	Mid
spt4	FBgn0028683	0.510007	Mid
CG6907	FBgn0031711	0.509872	Mid
east	FBgn0261954	0.505533	Mid
Rpb7	FBgn0051155	0.505183	Mid
CG18011	FBgn0033491	0.505129	Mid
Eip75B	FBgn0000568	0.50393	Mid
CG16863	FBgn0028931	0.502179	Mid
CG14710	FBgn0037920	0.499803	Mid
CG34149	FBgn0083985	0.49964	Mid
net	FBgn0002931	0.498672	Mid
CG1529	FBgn0031144	0.498357	Mid

Supplemental Table 3.1 Continued

knrl	FBgn0001323	0.495568	Mid
lid	FBgn0031759	0.495135	Mid
CG6664	FBgn0036685	0.491176	Mid
Hop	FBgn0024352	0.490691	Mid
tlk	FBgn0086899	0.488329	Mid
chb	FBgn0021760	0.485645	Mid
Pur-alpha	FBgn0022361	0.485411	Mid
CG1888	FBgn0033421	0.48537	Mid
CG34376	FBgn0085405	0.484684	Mid
dsx	FBgn0000504	0.484439	Mid
SA	FBgn0020616	0.483984	Mid
TepIV	FBgn0041180	0.481794	Mid
Nelf-A	FBgn0038872	0.480294	Mid
CG3680	FBgn0037027	0.479921	Mid
zf30C	FBgn0022720	0.479099	Mid
Ssdp	FBgn0011481	0.478568	Mid
vri	FBgn0016076	0.476095	Low
MED15	FBgn0027592	0.475307	Low
Taf2	FBgn0011836	0.47417	Low
wac	FBgn0035120	0.473684	Low
RYBP	FBgn0034763	0.471832	Low
Nelf-E	FBgn0017430	0.471722	Low
CG14962	FBgn0035407	0.468327	Low
fs(1)K10	FBgn0000810	0.468167	Low
CG10347	FBgn0030342	0.467356	Low
BicD	FBgn0000183	0.465887	Low
fs(1)h	FBgn0004656	0.462738	Low
Dp	FBgn0011763	0.462084	Low
CG1792	FBgn0039860	0.46009	Low
psq	FBgn0263102	0.456004	Low
wek	FBgn0001990	0.4549	Low
RpII140	FBgn0262955	0.453397	Low
CG32700	FBgn0052700	0.451505	Low
BEAF-32	FBgn0015602	0.448884	Low
Dsp1	FBgn0011764	0.447449	Low
CG17272	FBgn0038830	0.446712	Low
sel	FBgn0263260	0.44568	Low
CG1943	FBgn0037468	0.445326	Low
Su(z)12	FBgn0020887	0.445303	Low
Pbp45	FBgn0038371	0.442042	Low
simj	FBgn0010762	0.440324	Low
Brd8	FBgn0039654	0.439309	Low

Supplemental Table 3.1 Continued

l(2)03709	FBgn0010551	0.438898	Low
LSm3	FBgn0051184	0.437949	Low
Atac3	FBgn0052343	0.435531	Low
qkr54B	FBgn0022987	0.435392	Low
Vha36-1	FBgn0022097	0.433929	Low
Trf2	FBgn0261793	0.433364	Low
4EHP	FBgn0053100	0.432368	Low
gro	FBgn0001139	0.432286	Low
Pp4-19C	FBgn0023177	0.429878	Low
CG9799	FBgn0038146	0.428439	Low
CG7928	FBgn0039740	0.427458	Low
CG12219	FBgn0043796	0.426621	Low
l(3)mbt	FBgn0002441	0.425817	Low
CG9422	FBgn0033092	0.425523	Low
Glut4EF	FBgn0263097	0.424327	Low
MEP-1	FBgn0035357	0.421523	Low
dwg	FBgn0000520	0.419455	Low
E(z)	FBgn0000629	0.41313	Low
CG17002	FBgn0033122	0.411863	Low
gfzf	FBgn0250732	0.411119	Low
CG3281	FBgn0260741	0.410949	Low
TfIIEalpha	FBgn0015828	0.410176	Low
Taf12	FBgn0011290	0.405633	Low
chic	FBgn0000308	0.404501	Low
atms	FBgn0010750	0.404352	Low
CG6254	FBgn0037794	0.403712	Low
ref(2)P	FBgn0003231	0.403711	Low
pho	FBgn0002521	0.403597	Low
CG16865	FBgn0028919	0.402575	Low
Sce	FBgn0003330	0.398496	Low
Syb	FBgn0003660	0.398261	Low
TfIIB	FBgn0004915	0.397624	Low
slmb	FBgn0023423	0.396568	Low
CG1307	FBgn0026566	0.395264	Low
CG34132	FBgn0083968	0.393863	Low
TfIIFalpha	FBgn0010282	0.392988	Low
CG4424	FBgn0038765	0.392801	Low
MED4	FBgn0035754	0.392317	Low
nej	FBgn0261617	0.391996	Low
Sry-beta	FBgn0003511	0.39074	Low
CG11015	FBgn0031830	0.387396	Low
CG6568	FBgn0034210	0.386976	Low

Supplemental Table 3.1 Continued

tgo	FBgn0015014	0.386667	Low
CG14073	FBgn0036814	0.385812	Low
D19A	FBgn0022935	0.384138	Low
E2f	FBgn0011766	0.383128	Low
P58IPK	FBgn0037718	0.382854	Low
Rpb11	FBgn0032634	0.38178	Low
ttk	FBgn0003870	0.381353	Low
e(r)	FBgn0011586	0.378531	Low
snama	FBgn0086129	0.377347	Low
bs	FBgn0004101	0.37634	Low
CG18292	FBgn0030269	0.374879	Low
CG14667	FBgn0037317	0.374672	Low
RpII33	FBgn0026373	0.373491	Low
CoVa	FBgn0019624	0.372032	Low
Rtc1	FBgn0020909	0.371331	Low
CG42724	FBgn0261641	0.371172	Low
CG34163	FBgn0085192	0.37024	Low
CG11505	FBgn0035424	0.370113	Low
san	FBgn0024188	0.368514	Low
z	FBgn0004050	0.367042	Low
Kr-h1	FBgn0028420	0.361114	Low
CG5118	FBgn0031317	0.360925	Low
CG10466	FBgn0032822	0.360027	Low
hay	FBgn0001179	0.357623	Low
Arp14D	FBgn0011742	0.357426	Low
lolal	FBgn0022238	0.354972	Low
Sgf11	FBgn0036804	0.353738	Low
CG3731	FBgn0038271	0.352626	Low
Sam-S	FBgn0005278	0.351047	Low
CG11723	FBgn0031391	0.350804	Low
homer	FBgn0025777	0.349734	Low
Mfe2	FBgn0030731	0.349453	Low
p53	FBgn0039044	0.347986	Low
MED25	FBgn0038760	0.345537	Low
l(3)j2D3	FBgn0011335	0.344397	Low
PyK	FBgn0003178	0.343304	Low
Mnt	FBgn0023215	0.341005	Low
CG14657	FBgn0037282	0.33909	Low
MBD-like	FBgn0027950	0.338795	Low
CG8928	FBgn0030711	0.337478	Low
Oscp	FBgn0016691	0.336612	Low
roq	FBgn0036621	0.336415	Low

Supplemental Table 3.1 Continued

CG5003	FBgn0039554	0.335992	Low
CG33981	FBgn0250851	0.335454	Low
TfIIS	FBgn0010422	0.334647	Low
CG9436	FBgn0033101	0.334644	Low
Hakai	FBgn0032812	0.334501	Low
fliI	FBgn0000709	0.332981	Low
lark	FBgn0011640	0.331012	Low
CG14894	FBgn0038428	0.330542	Low
Shc	FBgn0015296	0.327778	Low
vig	FBgn0024183	0.327651	Low
ird5	FBgn0024222	0.324416	Low
CG9426	FBgn0032485	0.324021	Low
cerv	FBgn0030657	0.324019	Low
crol	FBgn0020309	0.322587	Low
CG8064	FBgn0038597	0.322534	Low
CG7878	FBgn0037549	0.321968	Low
CG10366	FBgn0032814	0.320781	Low
usp	FBgn0003964	0.32052	Low
Kdm2	FBgn0037659	0.318603	Low
cdm	FBgn0261532	0.318442	Low
CG17806	FBgn0038548	0.317862	Low
Zif	FBgn0037446	0.314764	Low
CG10321	FBgn0034643	0.314372	Low
CG7192	FBgn0030894	0.313345	Low
Chi	FBgn0013764	0.313185	Low
CG9018	FBgn0035318	0.313135	Low
Hsp60	FBgn0015245	0.312521	Low
TH1	FBgn0010416	0.309946	Low
CG5727	FBgn0032193	0.309909	Low
Su(z)2	FBgn0008654	0.309149	Low
Rab35	FBgn0031090	0.308242	Low
CG2021	FBgn0035271	0.308166	Low
Atac1	FBgn0031876	0.305604	Low
CG8461	FBgn0038235	0.304278	Low
sd	FBgn0003345	0.30296	Low
CG10641	FBgn0032731	0.302651	Low
CG13151	FBgn0033750	0.302607	Low
cni	FBgn0000339	0.30193	Low
Taf1	FBgn0010355	0.301645	Low
ssp	FBgn0036248	0.300105	Low
Myb	FBgn0002914	0.299496	Low
CG4004	FBgn0030418	0.29942	Low

Supplemental Table 3.1 Continued

Adf1	FBgn0000054	0.299145	Low
CG1908	FBgn0030274	0.298746	Low
CG32264	FBgn0052264	0.298551	Low
dbr	FBgn0067779	0.297874	Low
D19B	FBgn0022699	0.297571	Low
CG4360	FBgn0038787	0.296335	Low
CG1416	FBgn0032961	0.29478	Low
PQBP1	FBgn0039270	0.294505	Low
elB	FBgn0004858	0.292806	Low
CG3163	FBgn0034961	0.292732	Low
Pgi	FBgn0003074	0.291552	Low
fabp	FBgn0037913	0.291318	Low
Nmnat	FBgn0039254	0.290624	Low
CG4078	FBgn0029798	0.290435	Low
l(3)01239	FBgn0010741	0.289183	Low
Cdc42	FBgn0010341	0.288633	Low
mbf1	FBgn0262732	0.288047	Low
h	FBgn0001168	0.287797	Low
Trl	FBgn0013263	0.286529	Low
Actr13E	FBgn0011741	0.285852	Low
CG9890	FBgn0034814	0.285653	Low
cact	FBgn0000250	0.285545	Low
loco	FBgn0020278	0.285423	Low
Bx42	FBgn0004856	0.28455	Low
CG10600	FBgn0032717	0.282454	Low
CG2790	FBgn0027599	0.279642	Low
Mef2	FBgn0011656	0.278488	Low
CG8436	FBgn0037670	0.277194	Low
CG9866	FBgn0031420	0.27617	Low
pre-mod(mdg4)-T	FBgn0261837	0.276105	Low
CG5446	FBgn0032429	0.275169	Low
Rab5	FBgn0014010	0.27398	Low
HmgZ	FBgn0010228	0.273395	Low
cbt	FBgn0043364	0.272715	Low
drongo	FBgn0020304	0.272223	Low
CG5382	FBgn0038950	0.272011	Low
CG18476	FBgn0037931	0.270935	Low
Mad	FBgn0011648	0.268705	Low
viaf	FBgn0036237	0.26817	Low
tay	FBgn0260938	0.267783	Low
Pgd	FBgn0004654	0.266525	Low
CG12054	FBgn0039831	0.26645	Low

Supplemental Table 3.1 Continued

Tip60	FBgn0026080	0.266397	Low
Sry-delta	FBgn0003512	0.265997	Low
CG5214	FBgn0037891	0.265405	Low
CG9977	FBgn0035371	0.265275	Low
Ufd1-like	FBgn0036136	0.2643	Low
CG11985	FBgn0040534	0.264168	Low
CG10274	FBgn0035690	0.257464	Low
l(1)10Bb	FBgn0001491	0.25695	Low
Txl	FBgn0035631	0.256299	Low
Rpb4	FBgn0053520	0.255598	Low
CG5590	FBgn0039537	0.255171	Low
slim	FBgn0261477	0.254467	Low
CG32576	FBgn0052576	0.253781	Low
CG6340	FBgn0030648	0.252244	Low
CG15735	FBgn0030364	0.251718	Low
yki	FBgn0034970	0.250937	Low
Taf10b	FBgn0026324	0.250756	Low
CG4210	FBgn0038302	0.249457	Low
Stam	FBgn0027363	0.249205	Low
mri	FBgn0035107	0.247392	Low
Tim8	FBgn0027359	0.246695	Low
Eps-15	FBgn0035060	0.24631	Low
CG18005	FBgn0037660	0.246262	Low
Vha44	FBgn0262511	0.245343	Low
CG9776	FBgn0027866	0.244827	Low
CG14220	FBgn0031036	0.243237	Low
Tim10	FBgn0027360	0.24095	Low
Aats-asp	FBgn0002069	0.240307	Low
bbx	FBgn0024251	0.240107	Low
Ppt1	FBgn0030057	0.239541	Low
RhoGEF2	FBgn0023172	0.239391	Low
CG1434	FBgn0030554	0.238797	Low
Rel	FBgn0014018	0.238441	Low
CG18619	FBgn0032202	0.238183	Low
CG11247	FBgn0037120	0.23674	Low
Rga	FBgn0017550	0.236008	Low
CG13624	FBgn0039209	0.23589	Low
CG34179	FBgn0085208	0.235359	Low
CG15098	FBgn0034398	0.232826	Low
CG4294	FBgn0034742	0.229988	Low
membrin	FBgn0260856	0.229773	Low
Pgk	FBgn0250906	0.229743	Low

Supplemental Table 3.1 Continued

HDAC6	FBgn0026428	0.229741	Low
Scamp	FBgn0040285	0.228594	Low
CG6276	FBgn0038316	0.228428	Low
CG17912	FBgn0032600	0.227655	Low
Adk2	FBgn0022708	0.223901	Low
Nf-YB	FBgn0032816	0.223043	Low
exd	FBgn0000611	0.222744	Low
CG33260	FBgn0053260	0.219124	Low
Spt5	FBgn0040273	0.218414	Low
Tpr2	FBgn0032586	0.216557	Low
eIF2B-epsilon	FBgn0023512	0.2164	Low
l(2)NC136	FBgn0033029	0.214049	Low
hrg	FBgn0015949	0.212588	Low
Brf	FBgn0038499	0.212461	Low
CG40351	FBgn0040022	0.211426	Low
Hel89B	FBgn0022787	0.211108	Low
DnaJ-1	FBgn0263106	0.210195	Low
CG2889	FBgn0030206	0.209517	Low
Dim1	FBgn0031601	0.208125	Low
Taf11	FBgn0011291	0.207993	Low
ecd	FBgn0000543	0.207652	Low
CG7461	FBgn0034432	0.207369	Low
KdelR	FBgn0022268	0.207334	Low
CG3342	FBgn0029874	0.205117	Low
CG18815	FBgn0042138	0.204514	Low
Jwa	FBgn0032704	0.203686	Low
CG13887	FBgn0035165	0.202958	Low
coro	FBgn0033109	0.202843	Low
CG8446	FBgn0034089	0.202776	Low
su(f)	FBgn0003559	0.201479	Low
CG12084	FBgn0043458	0.199951	Low
robl	FBgn0024196	0.198932	Low
drk	FBgn0004638	0.19615	Low
her	FBgn0001185	0.19007	Low
eIF-5A	FBgn0034967	0.189577	Low
CG11710	FBgn0031115	0.188264	Low
TfIIEbeta	FBgn0015829	0.187644	Low
skpA	FBgn0025637	0.187366	Low
key	FBgn0041205	0.185159	Low
CG8578	FBgn0030699	0.184583	Low
Karybeta3	FBgn0087013	0.183853	Low
awd	FBgn0000150	0.183497	Low

Supplemental Table 3.1 Continued

Rab7	FBgn0015795	0.182445	Low
CG42669	FBgn0261551	0.181072	Low
Rpb12	FBgn0262954	0.180494	Low
CG17471	FBgn0039924	0.180464	Low
Srp9	FBgn0035827	0.180087	Low
AnnX	FBgn0000084	0.178949	Low
CG3040	FBgn0029925	0.175999	Low
CG2862	FBgn0031459	0.175121	Low
Met	FBgn0002723	0.173104	Low
Spt3	FBgn0037981	0.172554	Low
CG34159	FBgn0085188	0.171461	Low
Rab10	FBgn0015789	0.169876	Low
CkIalpha	FBgn0015024	0.168651	Low
Moe	FBgn0011661	0.168385	Low
CG17259	FBgn0031497	0.167822	Low
CG17765	FBgn0033529	0.166858	Low
poe	FBgn0011230	0.164698	Low
FK506-bp2	FBgn0013954	0.162094	Low
CG4389	FBgn0028479	0.160219	Low
CG6769	FBgn0030878	0.157596	Low
CHOp24	FBgn0029709	0.156324	Low
Gdi	FBgn0004868	0.153874	Low
l(2)06496	FBgn0010622	0.15226	Low
MrgBP	FBgn0033341	0.151951	Low
PP2A-B'	FBgn0042693	0.15158	Low
ens	FBgn0035500	0.151522	Low
Uev1A	FBgn0035601	0.150626	Low
shrb	FBgn0086656	0.14664	Low
CrebB-17A	FBgn0014467	0.146564	Low
Rab8	FBgn0262518	0.146142	Low
Cf2	FBgn0000286	0.139956	Low
SsRbeta	FBgn0011016	0.137574	Low
CG1233	FBgn0035137	0.135	Low
CG17385	FBgn0033934	0.132604	Low
Rpb10	FBgn0039218	0.129598	Low
RpL18A	FBgn0010409	0.128129	Low
TfIIA-S	FBgn0013347	0.127964	Low
Tapdelta	FBgn0021795	0.125018	Low
cindr	FBgn0027598	0.119681	Low
maf-S	FBgn0034534	0.119035	Low
RpL38	FBgn0040007	0.105881	Low
CG17059	FBgn0040754	0.101469	Low

Supplemental Table 3.1 Continued

RpL29	FBgn0016726	0.099474	Low
Rab11	FBgn0015790	0.098307	Low
CG17737	FBgn0035423	0.097278	Low
CG34191	FBgn0085220	0.096843	Low
CG6272	FBgn0036126	0.088832	Low
CG6523	FBgn0032509	0.085011	Low
RpL22	FBgn0015288	0.084221	Low
CG1458	FBgn0062442	0.080682	Low
RpL37a	FBgn0030616	0.080663	Low
Ef2b	FBgn0000559	0.076297	Low
CG14782	FBgn0025381	0.071587	Low
RpL13A	FBgn0037351	0.068282	Low
Surf4	FBgn0019925	0.064215	Low
Adam	FBgn0027619	0.060017	Low
RpS29	FBgn0261599	0.058276	Low
RpS27A	FBgn0003942	0.049811	Low

Appendix B

A Protein Complex Network of *Drosophila melanogaster*

Attributions:

This appendix contains work published as:

Guruharsha KG*, Rual JF*, Zhai B*, Mintseris J*, Vaidya P, Vaidya N, Beekman C, Wong C, **Rhee DY**, Cenaj O, McKillip E, Stapleton M, Wan KH, Yu C, Parsa B, Carlson JW, Chen X, Kapadia B, VijayRaghavan K, Gygi SP, Celniker SE, Obar RA, Artavanis-Tsakonas S. A Protein Complex Network of *Drosophila melanogaster*. Cell. 2011, 147:690-703

DY Rhee performed immunoprecipitation experiments, which contributed to the interaction network, contributed to the text and participated in editing of the manuscript.

K VijayRaghavan, SP Gygi, SE Celniker, RA Obar and S Artavanis-Tsakonas advised the project.

All other authors performed the remaining experiments and analyses.

*Authors contributed equally to this work

Cell

Resource

A Protein Complex Network of *Drosophila melanogaster*

K.G. Guruharsha,[1,4] Jean-François Rual,[1,4] Bo Zhai,[1,4] Julian Mintseris,[1,4] Pujita Vaidya,[1] Namita Vaidya,[1] Chapman Beekman,[1] Christina Wong,[1] David Y. Rhee,[1] Odise Cenaj,[1] Emily McKillip,[1] Saumini Shah,[1] Mark Stapleton,[2] Kenneth H. Wan,[2] Charles Yu,[2] Bayan Parsa,[2] Joseph W. Carlson,[2] Xiao Chen,[2] Bhaveen Kapadia,[2] K. VijayRaghavan,[3] Steven P. Gygi,[1] Susan E. Celniker,[2] Robert A. Obar,[1,*] and Spyros Artavanis-Tsakonas[1,*]
[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[2]Berkeley *Drosophila* Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India
[4]These authors contributed equally to this work
*Correspondence: robert_obar@hms.harvard.edu (R.A.O.), artavanis@hms.harvard.edu (S.A.-T.)
DOI 10.1016/j.cell.2011.08.047

SUMMARY

Determining the composition of protein complexes is an essential step toward understanding the cell as an integrated system. Using coaffinity purification coupled to mass spectrometry analysis, we examined protein associations involving nearly 5,000 individual, FLAG-HA epitope-tagged *Drosophila* proteins. Stringent analysis of these data, based on a statistical framework designed to define individual protein-protein interactions, led to the generation of a *Drosophila* *p*rotein *i*nteraction *m*ap (DPiM) encompassing 556 protein complexes. The high quality of the DPiM and its usefulness as a paradigm for metazoan proteomes are apparent from the recovery of many known complexes, significant enrichment for shared functional attributes, and validation in human cells. The DPiM defines potential novel members for several important protein complexes and assigns functional links to 586 protein-coding genes lacking previous experimental annotation. The DPiM represents, to our knowledge, the largest metazoan protein complex map and provides a valuable resource for analysis of protein complex evolution.

INTRODUCTION

The vast majority of proteins work as parts of assemblies composed of several elements, thereby defining protein complexes as essential cellular functional units. The functionality of proteins relies on their ability to interact with one another, whereas pathogenic conditions can reflect the loss of such function. Given the fundamental importance of protein interactions, proteome-wide "interactome" maps based on pairwise protein interactions using the yeast two-hybrid (Y2H) system have been determined for several organisms (Giot et al., 2003; Ito et al., 2001; Li et al., 2004; Rual et al., 2005; Stanyon et al., 2004; Stelzl et al., 2005; Uetz et al., 2000). Alternatively, protein complex isolation based on coaffinity purification combined with tandem mass spectrometry (coAP-MS) has been used to generate protein complex maps at proteome scale for *Saccharomyces cerevisiae* (Gavin et al., 2006; Ho et al., 2002; Krogan et al., 2006), *Escherichia coli* (Hu et al., 2009), and *Mycoplasma pneumoniae* (Kühner et al., 2009). This approach has been proven successful in the study of defined metazoan proteomic subspaces (Behrends et al., 2010; Bouwmeester et al., 2004; Ewing et al., 2007; Guerrero et al., 2008; Sowa et al., 2009), but there are no large-scale protein complex maps available for metazoans (reviewed in Gavin et al., 2011). Here, we present a substantial resource of affinity-tagged proteins, as well as the generation of a protein complex map of *Drosophila* that serves as a blueprint of interactions in a metazoan proteome.

Extensive genetic analyses in *Drosophila* have contributed fundamentally to our understanding of metazoan morphogenesis. However, many functional associations defined genetically in the animal lack mechanistic explanations. A comprehensive protein complex map would serve as a powerful resource to uncover the molecular basis of these genetic interactions and provide necessary mechanistic insights. Moreover, despite the success of the extensive molecular genetic studies in *Drosophila*, one-third (~14,000) of predicted *Drosophila* proteins (Adams et al., 2000) remains without functional annotation (Tweedie et al., 2009). The genetic tools available in *Drosophila* enable testing of predicted physical interactions in vivo, making it an ideal model organism for the generation of a comprehensive protein complex map. Such a map is a compelling tool for gene annotation, which is also incomplete in mammals, so a *Drosophila* map will be of considerable value for annotating mammalian proteomes.

Here, we describe the generation of a large-scale *Drosophila* *P*rotein *i*nteraction *M*ap (DPiM) by coAP-MS analysis based on ~3,500 affinity purifications. We developed a semiquantitative statistical approach to score protein interactions and defined a high-quality map. The map recovers many known and hundreds of previously uncharacterized protein complexes, thus providing functional associations and biological context for 586 proteins that previously lacked annotation. To our knowledge, the DPiM is the first large-scale metazoan protein complex analysis that is not focused on a specific subproteomic space,

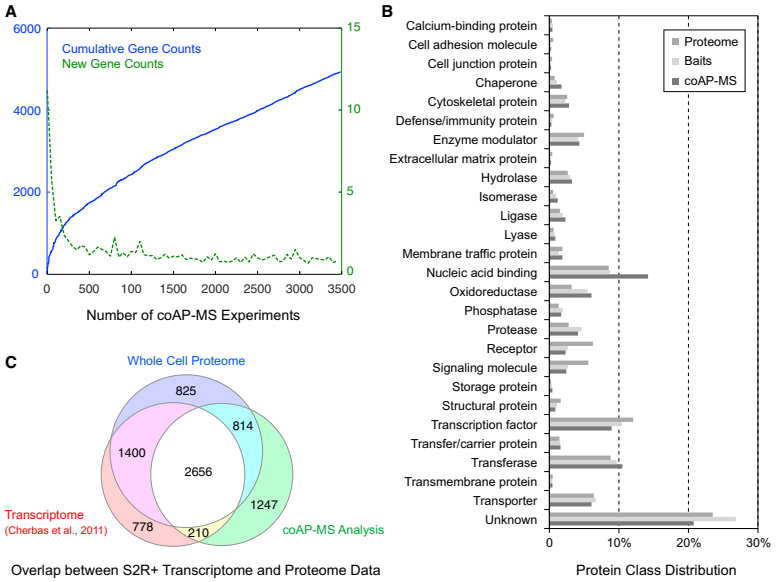

Figure 1. Analysis of Proteins Identified in the coAP-MS Pipeline

(A) Cumulative number of gene counts (blue) and unique gene counts (green) detected as a function of the number of high-quality affinity purification experiments.

(B) Comparison of protein class distribution between the *Drosophila* proteome, baits used and proteins identified in DPiM analysis (coAP-MS) using PANTHER (Thomas et al., 2003).

(C) A conservative estimate of overlap between the S2R+ cell transcriptome (5,044 protein-coding genes with gene score ≥300; Cherbas et al. [2011]), S2R+ proteome whole-cell lysate MS analysis (5,695 proteins), and the proteins identified in coAP-MS analysis (4,927 proteins). The intersections of the data sets are as follows: 4,056 (Transcriptome and Whole Cell Proteome), 3,470 (coAP-MS and Whole Cell Proteome), and 2,866 (Transcriptome and coAP-MS).

See also Figure S1 and Tables S1 and S2.

thereby providing a systems biology view of a metazoan proteome. The map defines a primary protein interaction landscape for *Drosophila* cells that allows study of the developmental dynamics and tissue level variation of any protein complex in the map. Finally, the DPiM offers a new reference point in the analysis of protein complex evolution.

RESULTS

High-Throughput *Drosophila* Proteomics Platform

To systematically isolate *Drosophila* protein complexes and determine their composition, we developed a large collection of affinity-tagged clones called the Universal Proteomics Resource (Yu et al., 2011; http://www.fruitfly.org/EST/proteomics.shtml) as part of the Berkeley *Drosophila* Genome Project (BDGP; see Experimental Procedures). From this collection, 4,273 individual clones were transiently transfected into S2R+ cells. Approximately 80% of the clones successfully expressed "bait" protein at detectable levels, and associated protein complexes were affinity purified. Purifications that resulted in detection of one or more unique, bait-derived peptides by mass spectrometry were considered for subsequent analysis, with few exceptions (see Experimental Procedures). This resulted in identification of a total of 4,927 *Drosophila* proteins (at 0.8% false discovery rate [FDR]) from 3,488 individual affinity purifications (Figure 1A). In general, mass spectrometric analysis of tryptic peptides cannot distinguish a specific protein isoform with confidence. So, for this analysis all the identified isoforms were traced back to the genes encoding them. From here on, all gene products are referred to as proteins without specifying isoforms. The raw mass spectrometry data are available in Table S1 (available online) and are accessible through FlyBase Linkouts and the DPiM website (https://interfly.med.harvard.edu/).

Comparison of protein functional class distribution using the PANTHER classification system (Thomas et al., 2003) indicates that the distribution of protein categories of baits used and proteins identified in coAP-MS is very similar to the overall distribution of the *Drosophila* proteome, much of which remains

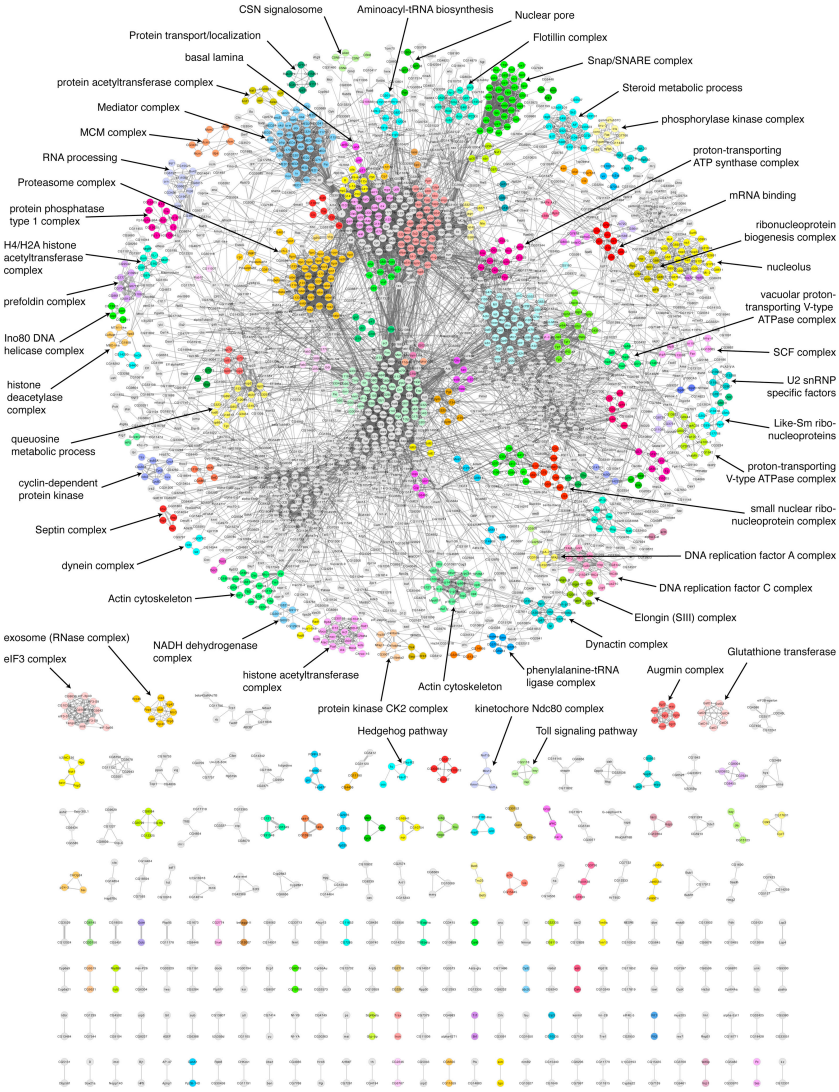

Figure 2. DPiM

Graphical representation of the DPiM comprising 10,969 high-confidence co-complex membership interactions (at 0.05% FDR) involving 2,297 proteins. Protein interactions are shown as gray lines with thickness proportional to the HGSCore for the interaction in the DPiM. The map defines 556 clusters, 377 of which are

unannotated (Figure 1B). A few minor differences are noted: nucleic acid-binding proteins and oxidoreductases are overrepresented, whereas receptor and signaling molecules are underrepresented in the coAP-MS data set (Figure 1B).

We determined the proteome composition of the S2R+ cell by high-resolution mass spectrometry, resulting in the identification of 6,081 proteins corresponding to 5,695 genes (1% FDR) in S2R+ cells (Figure 1C) (see Experimental Procedures; Figure S1; Table S2). The transcriptome data (Cherbas et al., 2011) and whole-cell proteome analyses indicate that more than one-third of the predicted *Drosophila* proteome is expressed in these cells. A large fraction of baits used for generating this map is expressed in S2R+ cells (61%), and 75% of proteins identified by coAP-MS were found in either transcriptome or whole-cell proteome analysis. Our analysis has interrogated a large portion of the S2R+ cell proteome but not saturated it. These are conservative estimates because strict comparisons with the transcriptome data are not possible given the methodological differences and absence of a rigorously defined FDR for the transcriptome data.

A *Drosophila* Protein Interaction Map

Proteins identified by coAP-MS represent a mixture of genuine direct or indirect interactors and nonspecific interactors (Ewing et al., 2007; Rees et al., 2011). The nonspecific interactors are present in a large number of data sets independent of the bait used, whereas genuine interactors tend to co-occur across relevant experiments. We developed a scoring system based on the hypergeometric probability distribution (Hart et al., 2007) to calculate the significance of co-occurrence of protein pairs by incorporating the total spectral counts (TSCs) for each protein. The number of TSCs correlates roughly with protein abundance in a sample (Liu et al., 2004) and, thus, increases the sensitivity of our approach by providing a semiquantitative dimension to the score. We refer to this scoring system as the HGSCore (HyperGeometric Spectral Counts score; see Experimental Procedures). A matrix model was used for both bait-prey and prey-prey interactions, and a total of 209,912 potential protein-protein interactions were scored among 4,927 *Drosophila* proteins (Table S3).

This statistical analysis led to the prediction of 10,969 high-confidence co-complex membership interactions (0.05% FDR) involving 2,297 *Drosophila* proteins, which are visualized as a network (Figure 2; Data S1). Further analyses of these high-confidence co-complex membership interactions based on the Markov clustering algorithm (MCL) (Enright et al., 2002) defined 556 putative complexes encompassing 2,240 proteins (Table S4). We use the term DPiM to refer to the composite data set and the resulting network. The map shows a distinct grouping of 1,817 (80% of total) proteins as the giant component of the network encompassing 377 (68%) putative complexes with a high degree of interconnectedness (Figure 2). A second group of 179 (32%) independent complexes defined by the map are not connected to other complexes. Among the baits that are expressed in S2R+ cells and part of the same MCL cluster, 36% (159 of 442) are found in direct reciprocal pull-downs. Some of the well-known complexes recovered in the DPiM are indicated in Figure 2.

DPiM Quality Assessment

The quality of the DPiM was evaluated using four approaches. First, we examined whether the coAP-MS approach was capable of identifying known interactions. Second, we asked if the members of complexes tend to share Gene Ontology (GO) annotation. Third, we examined whether the genes encoding proteins of the same complex tend to be coexpressed. Finally, we tested the ability of DPiM interactions to be validated across species using human proteins as baits in human embryonic kidney (HEK) 293 cells.

Defining a positive *Drosophila* reference set in order to assess the sensitivity and specificity of different scoring methods is difficult because existing data sets show little overlap (Yu et al., 2008), and there are no hand-curated databases similar to those available for the yeast and human proteomes. Hence, we used the extent of overlap from multiple diverse sources as an estimate of reliability of a given pairwise interaction. The DroID database (Murali et al., 2011) consolidates protein interaction data from seven discrete sources. Four bins of interactions were defined with increasing levels of confidence, i.e., those supported by at least one, two, three, or four independent DroID sources, and the overlaps with the DPiM were computed (Figure 3A). The coAP-MS data set was also analyzed using published scoring methods (Breitkreutz et al., 2010; Choi et al., 2011; Gavin et al., 2006; Hart et al., 2007; Sowa et al., 2009). Because these methods produce different numbers of interactions, we compared the top 25,000 interactions reported from each method with those listed in DroID. The HGSCore method recovered more interactions than other published scoring methods across all confidence levels, reflecting a 15% increase on average that is significant even when compared to the next best method (p value 6.9×10^{-12}) (Figure 3A). We find that the top 25,000 HGSCore interactions recover between 68% and 84% of the highest confidence interactions, i.e., physical interactions supported by either three or four independent DroID data sets (n = 247 and 61, respectively). When considering only those interactions above the 0.05 FDR threshold of HGSCore, the DPiM recovers between 56% and 71% of the highest confidence interactions. The overall increase in recall at increasing reference set confidence levels across multiple analysis methods suggests that the underlying data in the DPiM are of high quality, whereas the robust improvement HGSCore makes over established methods validates our approach. Nearly 86% of the interactions in the DPiM are novel when considering all the interactions reported in DroID, which includes interolog data from three species (yeast, worm, and human).

Proteins belonging to the same protein complex can be expected to be enriched for GO annotations, share the same KEGG pathways, and contain similar protein domains. The DAVID Functional Annotation Tools (Huang da et al., 2009)

interconnected, representing nearly 80% of the proteins in the network. The remaining 179 clusters are not connected to members of other complexes. Depicted with different colors are 153 clusters enriched for GO terms, KEGG pathways, or Pfam/InterPro domains. Proteins in other clusters that are not enriched are shown as gray circles. Selected complexes with known molecular function/biological role are indicated. See also Tables S3 and S4.

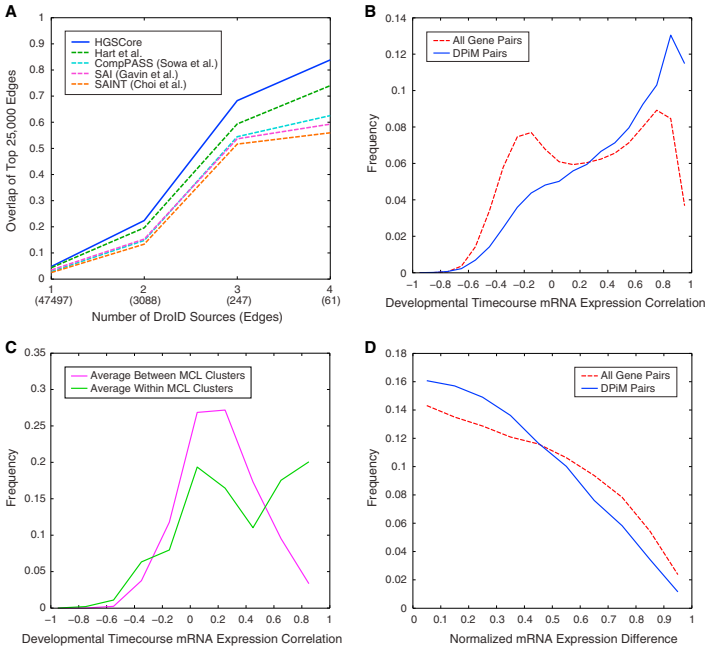

Figure 3. Evaluation of Quality of DPiM Protein Interactions

(A) Comparison of interactions in the DPiM data set and DroID. Four bins with increasing levels of confidence supported by at least one, two, three, or four DroID sources were defined. The overlap between the top 25,000 interactions defined by each of the co-occurrence analysis methods and DroID is shown. The number of interactions supported by given number of sources is indicated in parentheses along the x axis.

(B) Distribution of correlation coefficients between mRNAs corresponding to interacting proteins in the DPiM compared to all gene pairs, based on the RNA-Seq data (Graveley et al., 2011).

(C) Distribution of correlation coefficients of mRNAs corresponding to proteins within MCL clusters compared to those between MCL clusters, analysis similar to (B).

(D) Normalized absolute mRNA expression differences between DPiM interactors and all gene pairs (Cherbas et al., 2011).

See also Figure S2 and Table S5.

were used to calculate enrichment for annotations, pathways, and domains within each protein cluster generated by the DPiM. About 28% of the MCL-derived protein clusters (153 of 556) are enriched for one or more of these features (multiple hypothesis testing-adjusted $p < 0.01$) (Figure S4). In total, almost half of the proteins in the DPiM network fall into a GO term-enriched cluster (Table S4). Due to the nature of MCL clustering, some components of larger complexes tend to separate into smaller independent clusters, making it statistically less likely to find significant enrichment due to the small sample size.

Genes expressing subunits of protein complexes often tend to be coexpressed (Jansen et al., 2002; Krogan et al., 2006). Therefore, we used the developmental time course transcription profiling data sets from the modEncode project (Graveley et al., 2011) to examine the mRNA expression profile correlation between genes encoding interacting proteins. The frequency distribution of the correlation coefficients calculated between genes connected by DPiM edges is clearly skewed toward coregulated expression when compared with all-to-all gene correlations (Figure 3B). Similarly, transcripts corresponding to the same MCL clusters tend to be coexpressed more frequently than those belonging to different clusters (Figure 3C). Aside from correlated profiles, it has been suggested that both the expression profiles and the absolute level of expression of interacting partners may be maintained at similar levels in the cell as a consequence of coregulation of complex subunit stoichiometry (Jansen et al., 2002). Following Jansen et al. (2002), we calculated the normalized differences between absolute mRNA

expression levels from the modEncode RNA-Seq data (Cherbas et al., 2011) and confirmed this trend in flies (Figure 3D). Similar results involving both expression profiling and absolute levels were obtained from analogous analysis of gene expression data from 26 *Drosophila* tissues in FlyAtlas (Chintapalli et al., 2007) (Figure S2).

Cross-Species Validation of DPiM Interactions

Using orthologous HA-tagged human proteins as coAP-MS baits in HEK293 cell line (Graham et al., 1977), we examined whether DPiM-defined interactions can be validated across species. A set of 118 human bait proteins was selected based on whether an ORF clone was available in the CCSB human ORFeome collection (Lamesch et al., 2007; Rual et al., 2004), and if the corresponding *Drosophila* ortholog involved high HGSCore interactions in the DPiM.

After Gateway cloning of the corresponding ORF inserts into the pHAGE-N-FLAG-HA vector (Behrends et al., 2010), we successfully cloned and affinity purified 80% (94 of 118) of the baits, but the data set was too small to be analyzed by the HGSCore method. In the DPiM, a total of 2,641 interactions involves *Drosophila* orthologs of 1 of these 94 human proteins. Transcriptome data of HEK293 cells (Shaw et al., 2002; Williams et al., 2004) suggested that several human orthologs of interactors predicted by DPiM are not expressed in this cell type. Therefore, the analysis was restricted to 114 DPiM interactions that are found as "bait-prey" interactors in the raw *Drosophila* data set for which both human orthologs are expressed in 293 cells; the success rate was 51% (58 of 114) (Table S5). This validation rate illustrates the high specificity of our coAP-MS approach and the value of the DPiM as a reliable resource for biological hypothesis in human cells. A total of 268 human-validated DPiM interactions were novel (Table S5). Examples of these cross-species validated interactions are considered further below.

Proteasome and SNARE Complexes

To further assess the quality of the DPiM at protein complex level, we performed an in-depth analysis of two previously well-characterized complexes: the proteasome and the SNARE (SNAP [soluble NSF attachment protein] receptor) complex. The proteasome is a large multiprotein complex involved in protein degradation and has been extensively characterized in a variety of organisms but little studied in *Drosophila* (Hölzl et al., 2000). We used the KEGG database (Kanehisa et al., 2010), FlyBase (Tweedie et al., 2009), and original literature to generate a list of 51 putative *Drosophila* proteasome subunits (described in Table S6).

Affinity purification was performed for 32 individual proteasome subunits, and 42 of the 51 classified proteasome subunits were detected as copurifying proteins in at least 2 bait purifications. On average, 70% of the copurifying proteins are common between replicate proteasome bait purifications, and 84% of the high-confidence (DPiM) interactors were detected in both replicates (Table S6). It is noteworthy that proteins predicted to be from the same proteasome substructure, i.e., core, base, or lid, consistently copurified (Figure 4A). Consistent with yeast and human proteasome studies (Leggett et al., 2002; Wang et al., 2007), Rpn11—a proteasomal lid subunit—pulled down the majority of the proteasome components. Consistent with its predicted role in maturation of the proteasome core (Fricke et al., 2007), the proteasome maturation protein (Pomp) copurified with only a few core members (Figure 4A).

Of the 51 annotated proteasome subunits, 6 were detected only when they were used as bait. Interestingly, these were all recently described as testis-specific proteasome proteins (Belote and Zhong, 2009), and indeed, expression profiling analysis confirmed that they are not expressed in the *Drosophila* embryo-derived S2R+ cells (Cherbas et al., 2011). Nevertheless, when used as baits, the testis-specific proteins interacted with other proteasome components with profiles similar to those of their respective ubiquitous paralogs (Figure 4A). The fact that paralogous proteins produce similar interaction profiles illustrates the reproducibility of our coAP-MS approach and also suggests that the DPiM provides valuable information that can reach beyond the S2R+ proteome.

Importantly, this study also uncovered a set of seven additional subunits not originally predicted to be part of the proteasome complex: CG12321, CG11885, CG2046, CG13319, GNBP2, CG3812, and RPR (Figure 4B). Sequence similarity analysis revealed that CG12321 and CG11885 are the *Drosophila* homologs of proteasome assembly chaperone 2 and 3, respectively (KEGG). Nothing is known about the functions of CG2046 or CG13319, and the sequences or domain structures of GNBP2, CG3812, and RPR do not suggest a plausible relationship to the proteasome. Direct experimentation will be essential to explore their functionality and potential role in the proteasome complex.

We next examined the SNARE complex. SNARE proteins are a large protein superfamily implicated in mediating membrane fusion events during protein trafficking (Südhof and Rothman, 2009). In *Drosophila*, 23 SNARE proteins have been described (KEGG pathway: dme04130), and all of them are well connected in the DPiM. All SNARE proteins with the exception of Syntaxin 6 fall into two clusters (clusters #7 and #162; Figure 4C). Among nine proteins in cluster #7 (Table S4) that are not classified in KEGG as SNARE proteins, seven (Syb, Snap, Slh, gammaSnap, Syx13, CG6208, and Nsf2) have "SNAP receptor activity" or "SNAP activity" GO annotations and, thus, represent potential genuine interactors of the SNARE proteins. The remaining two proteins in cluster #7 (AttD and Rme-8) do not have prior annotation related to SNAP receptor activity. We also found that Syb is linked to several proteins in the map, which suggests that it is a shared component of multiple complexes. Connections of particular interest are the ones that link Syb with members of cluster #22 (the Flotillin complex), which is involved in protein transport and control of subcellular localization (Figure 4C). In total, 57 interactions (31 novel) from the SNAP/SNARE complex and 10 interactions (9 novel) from the Flotillin complex were independently validated in human 293F cells (Table S5).

The analyses of the proteasome and SNARE complexes confirm previously reported interactions, further validating the quality of the DPiM. Consequently, this also strengthens the potential of the DPiM to formulate functional hypotheses at the levels of both pairwise interactions and protein complex definition.

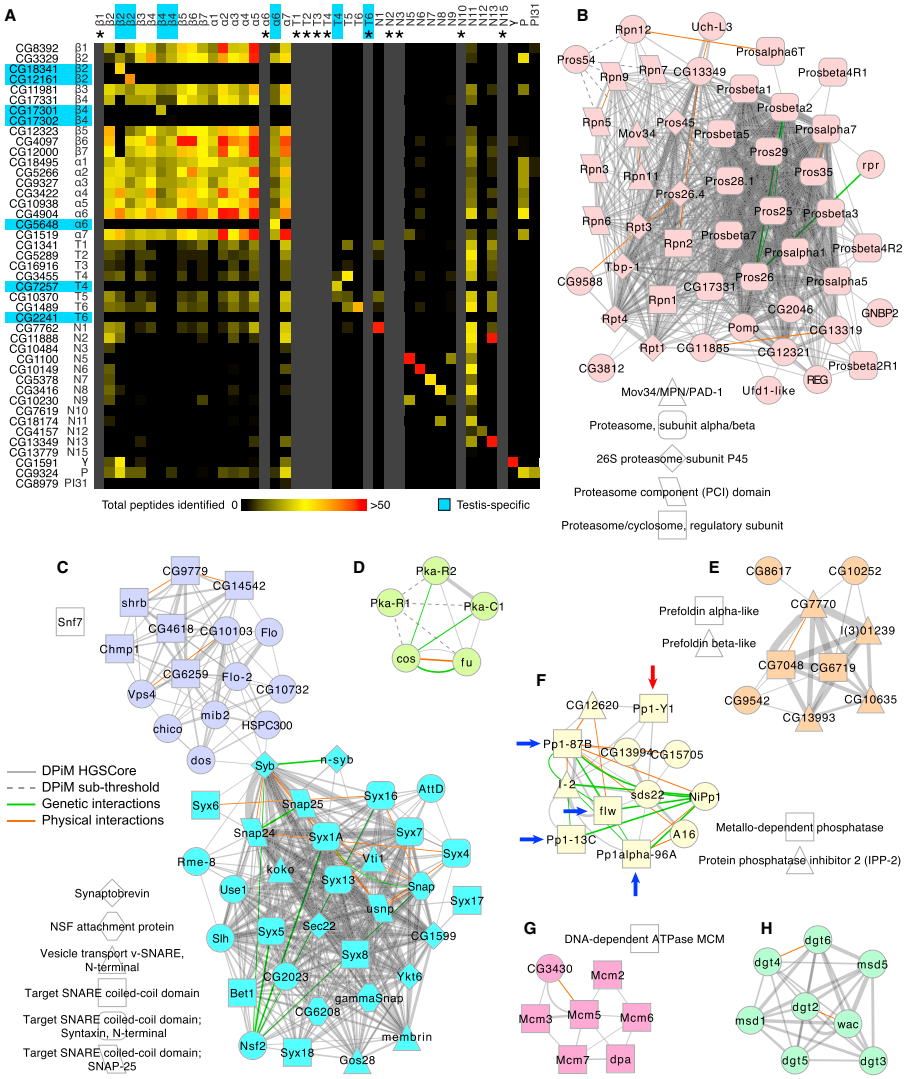

Figure 4. Biological Implications of Protein Complexes in the DPiM

(A) Two-dimensional heat map showing the number of peptides identified for each proteasome subunit. Each column corresponds to proteins copurified in a particular proteasome bait experiment. Gray columns (marked by asterisks) were added if a bait was unavailable. Both axes are arranged according to proteasome subunit classification, i.e., core (β and α) or regulatory (base and lid). Seven testis-specific subunits are highlighted in blue. "P" refers to Pomp.

Functional Implications of the DPiM

Slightly over half of the *Drosophila* protein-coding genes have associated experimental annotation (based on FlyBase release 5.23). Another 12% are annotated purely in silico (by inferred electronic annotation [IEA]), and the remainder (one-third of protein-coding genes) have no functional annotation. The DPiM provides empirical evidence and functional validation for 376 uncharacterized gene products and another 210 that were until now only annotated with IEA evidence. A total of 153 MCL clusters in the map show significant enrichment for GO terms, KEGG pathways, and Pfam/InterPro domains (multiple hypothesis-adjusted $p < 0.01$), indicating that members share common biological or functional attributes. These 153 annotation-enriched clusters include 167 proteins that lacked any annotation, for which the DPiM provides functional associations and biological context (Table S4). Inspection of individual protein complexes provides insights into specific as well as general functional aspects of the map. To illustrate this, six protein clusters with members sharing GO terms and pleiotropic cellular functions are described below (Figure 4).

The Hedgehog pathway is presumed to be "off" in the S2R+ cell line (Cherbas et al., 2011) but was represented by a few known pathway members (Pka-C1, Pka-R2, Cos, and Fu) as an autonomous cluster (Figure 4D). Interestingly, three of the four members of this cluster are protein kinases. Pka-R1 has only subthreshold HGSCore interactions with members of this cluster (Figure 4D). Pka-C1, known to interact with the transcription factor Costa, was not detected in our analysis of S2R+ cells.

Eukaryotic prefoldin is a multisubunit complex composed of two α and four β subunits that are required for stabilization of nascent proteins as they are translated and delivered to chaperonins for protein folding (Ohtaki et al., 2010). The complex is not well characterized in flies, and the subunits have been inferred from in silico approaches. This complex in the DPiM (Figure 4E) contains all six components (CG7770, CG6719, l(3)01239, CG7048, CG13993, and CG10635) as well as three additional putative complex members (CG9542, CG8617, and CG10252) (Figure 4E); essentially nothing is known about these proteins except for their sequences.

The complex related to Protein Phosphatase type 1 (PP1), one of the major classes of eukaryotic serine/threonine protein phosphatases (Dombrádi et al., 1990), includes all four known catalytic subunits, PP1c's, as well as the testis-specific subunit Pp1-Y1 (arrows in Figure 4F). In the DPiM, this complex includes the two inhibitory subunits (I-2 and CG12620) and two regulatory subunits (sds22 and A16). The two additional components CG15705 and CG13994 in this cluster were also found by Y2H analysis (Giot et al., 2003). Based mainly on Y2H interactions, it has been suggested that the *Drosophila* PP1c-interactome may include 40 putative PP1c-binding proteins (Bennett et al., 2006). Our coAP-MS analysis suggests that the PP1c complex in this cell type may be composed of fewer (12) proteins (Figure 4F).

The MCM (minichromosome maintenance 2–7) complex implicated in replication-associated helicase activity is suggested to be composed of six proteins in *Drosophila* (Forsburg, 2004). The DPiM defines a complex that contains all six as well as a seventh putative member, the uncharacterized protein CG3430 (Figure 4G).

The Augmin complex (Figure 4H), which is essential for spindle formation, has been defined through a series of biochemical studies, which in addition to the dgt protein core (dgt2–6), identified wac, msd1, and msd5 as members of the complex (Goshima and Kimura, 2010). The DPiM identified the Augmin complex in its entirety as a stand-alone cluster (Figure 4H).

Additional examples of known protein complexes with diverse biological and molecular functions are shown in Figure S3. The map also identified several IEA annotated proteins, which, although sharing GO terms, were not known to be members of a complex. For example cluster #166 (Table S4) is made up of three members (CG12171, CG31549, and CG31548) with a high average HGSCore (388). All three share a glucose/ribitol dehydrogenase domain, a NAD(P)-binding domain, and short-chain dehydrogenase/reductase (SDR)-conserved sites. DPiM results suggest that these previously uncharacterized proteins form a functional complex. In contrast, the DPiM also predicts the existence of complexes with members sharing experimentally derived annotation but no common GO terms (for example, cluster #27, Table S4).

Intercomplex Interactions and Functional Relationships

The predictive value of the DPiM for individual protein complexes is exemplified by the aforementioned analysis, but probing the interconnectedness of complexes within the map is far more challenging. On a global level, the interconnectedness of DPiM complexes is visualized in Figure S4. In numerous cases, we observed that functionally related complexes are well connected in the map. For a better understanding of protein function, it is important to examine possible functional relationships that involve not only immediate complex neighbors but also complexes that are associated with each other indirectly via intermediate protein assemblies.

(B) The proteasome cluster in the DPiM with subunits shaped according to Pfam/InterPro domains; circles represent nodes without domain enrichment. The thickness of each gray line is proportional to the HGSCore of interaction. Additional physical (red lines) and genetic (green lines) evidence from literature is shown, with line thickness proportional to number of sources.

(C) Clusters #7 and #162, the SNAP/SNARE complex, is connected by Syb to several members of cluster #22, the Flotillin complex.

(D) Cluster #117 includes proteins belonging to the Hedgehog-signaling pathway. Protein Pka-R1 has interactions with HGSCores below threshold (dotted lines).

(E) Cluster #42, the Prefoldin complex, in which all six predicted members are connected, along with three additional proteins, none of which is well studied.

(F) Cluster #26, the PP1 complex has multiple genetic and physical interactions described in the literature. The known subunits PP1α87B, PP1α13C, PP1α96A, and PP1β9C (blue arrows) and testis-specific subunit Pp1-Y1 (red arrow) are shown.

(G) Cluster #60, the MCM (helicase) complex, has all six known members along with CG3430 (connected to Mcm3 and Mcm5).

(H) Cluster #47, the Augmin complex, involved in mitotic spindle organization, is a stand-alone complex in the DPiM network.

See also Figure S3 and Table S6.

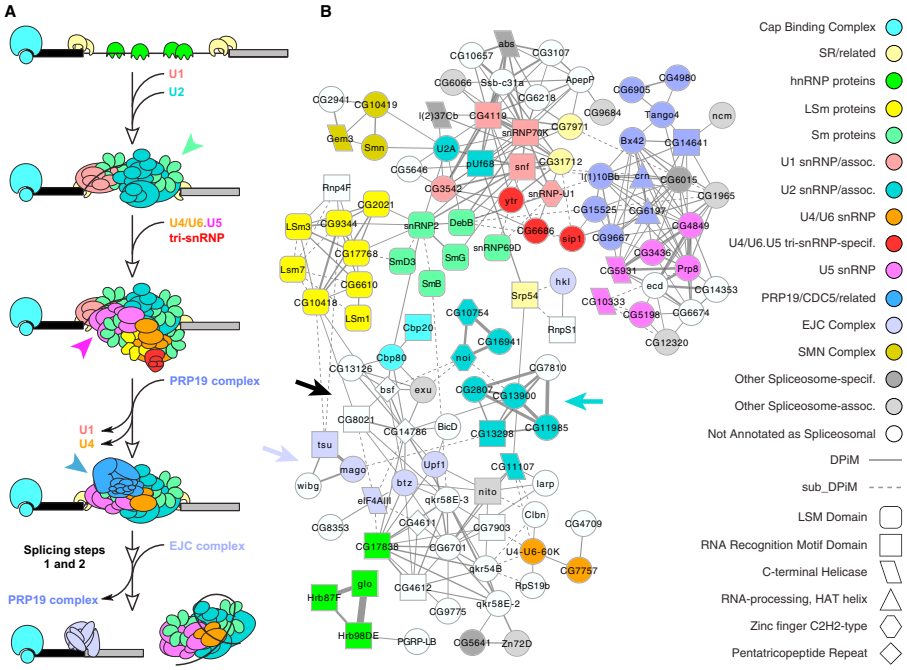

Figure 5. Modularity of the Spliceosome Subnetwork

(A) Schematic representation of stepwise interaction of snRNPs with pre-mRNA and other proteins in the process of splicing introns, as described in the literature.
(B) The spliceosome subnetwork in the DPiM consists of 12 clusters that are well connected. The ~80 nodes in this subnetwork constitute a very substantial portion of the spliceosome pathway as defined in KEGG (pathway: dme03040) and Herold et al. (2009). The major spliceosome subcomplexes are colored according to functional annotation (same as in A for comparison), and proteins are shaped according to Pfam domain enrichment. Protein interactions are shown as gray lines with thicknesses proportional to HGSCore, and those with scores below the statistical cutoff are shown as dotted lines. Other proteins that are not classified as spliceosome components in KEGG or elsewhere but connected to these complexes in the DPiM network are uncolored. A majority of such non-spliceosomal proteins have "mRNA binding" annotation. The modularity of this multisubunit molecular machinery is preserved in the DPiM in the form of subnetworks that cluster together. Colored arrows and arrowheads denote complexes referred to in the text.
See also Figure S4.

Given the level of functional characterization and modularity of the spliceosome, we chose it to examine whether functionally significant first- and second-degree neighboring interactions and clusters could be identified in the DPiM. The conformation and composition of the spliceosome are highly dynamic and are responsible for the accuracy as well as the flexibility of the splicing machinery. It is composed of several well-defined snRNPs that associate sequentially with pre-mRNA to guide intron splicing (Figure 5A). Each snRNP consists of one or two snRNAs, a common set of seven Sm (or LSm) proteins, and a variable number of unit-specific proteins (Will and Lührmann, 2011).

The spliceosome subnetwork in the DPiM (Figure 5B) is composed of 12 clusters containing most of the known spliceosome-related proteins. This clustering of spliceosome components in an unbiased systematic analysis of whole-cell lysates illustrates the power of our approach. Importantly, these spliceosome clusters are interconnected in the network, consistent with the notion that they share functionality, while remaining spatially and temporally modular. The complex defined by the six Sm proteins (green arrowhead, Figure 5A) is connected to other first-degree and second-degree neighboring clusters composed of specific U1-, U2-, U4-, U5-, and U6-related factors. Most Prp19/CDC5L complex members (magenta arrowhead, Figure 5A) are well connected to all U5-specific factors (blue arrowhead, Figure 5A and Figure 5B). Similarly, the U2 snRNP-specific factors (CG2807, CG7810, CG13900, CG13298, and CG11985; cyan arrow, Figure 5B) and members of exon junction complex (EJC, blue-gray arrow, Figure 5B) are connected to Sm/LSm

proteins via CG14786 (Figure 5B) and other members of cluster #62 (black arrow, Figure 5B). Although none of the cluster #62 members is classified as a spliceosome component, two are predicted as members of EJC (Upf1 and btz), and two others (CG8021 and bsf) have GO term annotation related to mRNA binding (not enriched at $p < 0.01$). Thus, a second-degree neighboring cluster defines functionally related protein assemblies in the DPiM.

Protein Complex Evolution

Examining the extent of conservation of individual protein subunits as well as the overall complex composition across organisms can shed valuable insight into their cellular roles. The most extensive manually curated annotations of protein complexes exist for yeast (MIPS, CYC2008) and human (REACTOME, CORUM). We aligned complexes defined by DPiM clusters with those described in yeast and human. Several complexes, for example MCM (Figure 4G, cluster #60), CCT (chaperonin-containing TCP1, Figure S3, cluster #32), and prefoldin (Figure 4E, cluster #42), showed almost complete conservation of composition between clearly orthologous subunits. Below, we focus on examples where orthology relationships are less obvious (Figure 6).

The eIF3 complex defines the largest eukaryotic initiation factor, which directs the multitude of steps essential for initiating translation. Comparison of the complexes from yeast and human to that of *Drosophila* (cluster #24, DPiM) reveals significant differences. The metazoan *Drosophila* and human complexes share seven interconnected proteins (Figures 6A–6C, within green-dotted region), which are not present in unicellular yeast, suggesting structural and functional remodeling specific to multicellular organisms. A group of four interconnected proteins is conserved in all three species (Figures 6A–6C, within blue-dotted region). Neither the raw data nor the HGSCore analysis supports Trp1 or Adam being part of the eIF3 complex, though their homologs are predicted to be members in other species. These findings allow us to raise the testable hypothesis that the role of yeast or human orthologs of Adam and Trip1 is not essential to the function of eIF3. We also compared Pfam domain compositions across the three species, revealing a gain of six domains in the metazoans in comparison to yeast and the loss of an unclassified domain in yeast with respect to metazoans (Table S7A). It is worth noting that none of the eIF3 complex members was used as bait; its recovery illustrates the power of our scoring approach.

The signalosome is a functionally conserved complex that catalyzes the deneddylation of proteins and promotes degradation through the cullin family of ubiquitin E3 ligases (Kato and Yoneda-Kato, 2009). Yeast proteins share surprisingly little sequence similarity with metazoan counterparts, despite the fact that the yeast complex has been shown to be functionally homologous to metazoan signalosomes (Wee et al., 2002) (Figure 6D). The eukaryotic signalosomes are composed of eight subunits (CSN1–8) as seen in the human complex (Figure 6F). The *Drosophila* signalosome has also been suggested to comprise eight subunits (Freilich et al., 1999), but our coAP-MS data raise the possibility that CSN1a, CSN1b, and CSN8 are not part of the complex, at least in S2R+ cells (Figure 6E). Domain analysis shows a linear growth in the number of PCI domains from yeast to humans, which cannot be attributed to the growth in the number of protein subunits (Table S7B).

The three-member ESCRT-I (endosomal-sorting complex required for transport) complex is well known in flies and humans (Michelet et al., 2010) (Figures 6G–6I). In the DPiM the ESCRT-I complex clustered with three other proteins that have no human homologs according to InParanoid (Figure 6H). The yeast complex shows some interesting characteristics. First, Vps28 is linked to STP22, a conserved interaction also evident in *Drosophila* and humans. On the other hand, MVB12, a multivesicular body-associated protein in yeast (arrow, Figure 6G), does not have a clear fly ortholog nor does it share a Pfam domain with any of the fly complex components. However, the *Drosophila* complex member CG7192, a protein of unknown function (arrow, Figure 6H), shares weak sequence similarity with the *Caenorhabditis elegans* protein C06A6.3, which has recently been shown to be functionally homologous to the yeast MVB12 (Audhya et al., 2007). Moreover, the yeast SRN2, whereas not identified as an ortholog of any metazoan gene, shares the Mod_r Pfam domain with fly CG1115 as well as human VPS37C (marked by asterisks, Figures 6G–6I), suggesting a weak evolutionary relationship.

Cluster #160 in the DPiM links four proteins associated with the UTP-B complex, a subcomplex of the SSU processome, a large ribonucleoprotein essential for RNA processing (Figure 6K). In yeast, two additional proteins (UTP6 and UTP18) are clearly part of this complex, but the corresponding proteins in *Drosophila* (CG7246 and l(2)kO7824) are not included in cluster #160 (Figures 6J and 6K). Both these proteins have been used as baits in the coAP-MS analysis, and they did not copurify other UTP complex members. Although the homologous proteins exist in humans, neither the interactions nor the complex has been extensively studied. The contrast of evolutionary information between yeast and fly provides an entry point for further investigation to see which of the interactions have been lost or retained in humans.

DISCUSSION

Understanding how functional units in the cell integrate their actions to control development and homeostasis defines a quintessential biological problem. Essential insights into this come from the definition of proteome architecture such as the map we present here, enabled by the knowledge of genome sequences and the development of sensitive mass spectrometry-based approaches. Although there are several studies focused on specific subproteomic spaces, no large-scale unbiased proteome map exists for higher eukaryotes (see review in Gavin et al., 2011). Our study defines a global metazoan protein complex network based on expression of a large library of affinity-tagged baits. The map includes a majority of proteins expressed in S2R+ cells and is based on the HGSCore, which includes a semiquantitative measure of protein abundance (TSCs), thus improving the sensitivity in comparison to other existing scoring methods. However, we note that several known interactions are detected in our analysis but fall below the statistical threshold (Table S3).

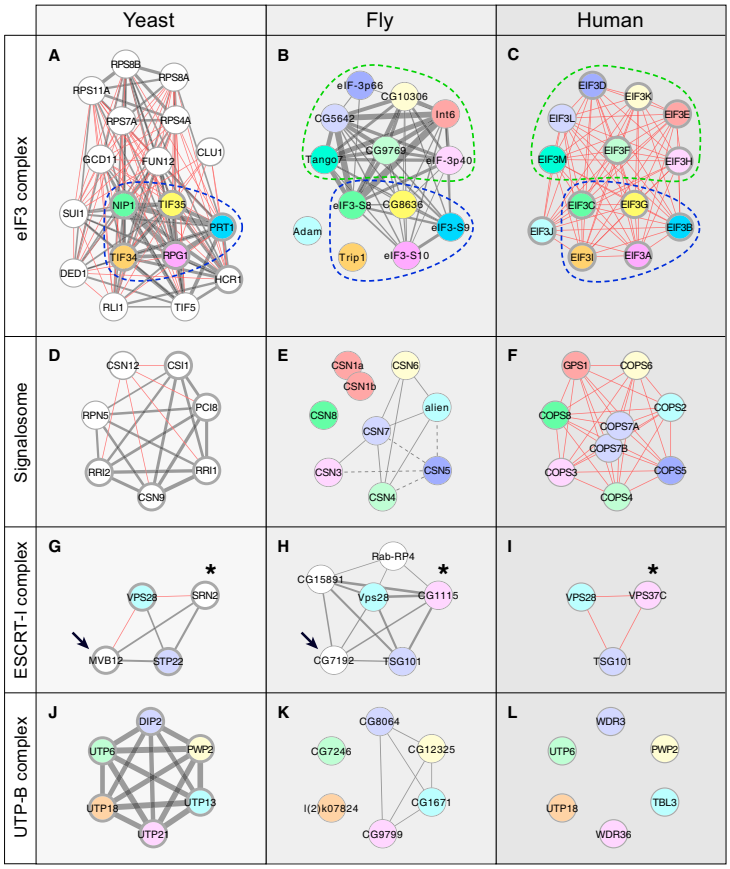

Figure 6. Examples of Protein Complex Evolution

Comparison of four complexes defined in fly by the DPiM (center panels) with yeast (left panels) and human complexes (right panels). Gray lines show physical interactions that have weighted scores, and red lines show interactions implied by the curated data sets. For comparison, InParanoid orthologs in all three species are depicted with identical colors. Proteins that do not have homologs in other species are shown in white. Complex members for which evidence exists in both high-throughput and curated data sets (yeast) or both REACTOME and CORUM databases (human) are distinguished by thicker nodes.

(A–C) The eIF3 complex (cluster #24). The fly and human complexes share seven interconnected proteins (within green-dotted region), which are not present in yeast. Five proteins are conserved in all three species (within blue-dotted region).

(D–F) The signalosome complex in yeast is composed of proteins sharing little sequence similarity with metazoan counterparts. The eukaryotic signalosome is composed of eight subunits (CSN1–8) as seen in the human complex (F), but CSN1a, CSN1b, and CSN8 are not part of the fly signalosome in S2R+ cells.

(G–I) ESCRT-I function is conserved from yeast to humans, but only VPS28 and STP22 in yeast and their respective fly and human orthologs are readily apparent. Additional analysis suggests a distant relationship between MVB12 in yeast and *Drosophila* complex member CG7192, a protein of unknown function (arrows). The yeast SRN2 also shares the Mod_r domain with CG1115 and VPS37C (asterisks).

(J) The yeast UTP-B complex involved in RNA processing has six well-connected members.

(K) In DPiM only four members are connected, but CG7246 and l(2)kO7824 are not included in the DPiM cluster #160.

(L) There is no evidence suggesting physical interaction among the complex members in human.

See also Table S7.

Several independent criteria indicate that the quality of the map is high, and clearly, the algorithms we use successfully clustered proteins that have been grouped previously as multimeric complexes. The broad recovery of known interactions and the remarkable enrichment of GO terms in individual clusters suggest that novel interactions predicted by the DPiM define important biological hypotheses as well as a powerful annotation tool. The analysis of the human protein orthologs we tested indicates that the DPiM reflects general features of metazoan proteomes and, thus, will be directly useful in probing protein interactions across species. We expect that the experimental and analytical resources we established will be useful as the proteome analysis is expanded to include additional *Drosophila* proteins and cells lines or tissues and provide a paradigm for proteomic studies in other organisms.

The DPiM, like its yeast counterparts (Gavin et al., 2006; Krogan et al., 2006), defines protein complex membership and suggests intercomplex relationships linking together functional units. Both issues are essential for understanding the network of functional relationships that govern the physiology of the cell. Experimentally probing such relationships is not trivial, but the availability of sophisticated genetic tools in *Drosophila* offers a unique opportunity to explore interactions using in vivo assays. Indeed, 118 of the DPiM direct interactions have been validated independently through genetic interactions involving mutant combinations (see FlyBase). Integration of protein and genetic interaction networks will afford us important insights that may provide a molecular basis for relationships only defined by genetics and, hence, generate mechanistic hypotheses.

The experimental approach we used has certain a priori limitations. We rely on the transient expression of epitope-tagged bait proteins, which are not expressed at normal levels, and tagging of the proteins may interfere with their functions. Nevertheless, the quality testing of the map indicates that despite these potential limitations, our experimental approach is generally reliable. We also note that several recent studies of subproteomic spaces using a similar experimental approach have produced valid results (Behrends et al., 2010; Sowa et al., 2009). Any cell type used will inevitably involve only a fraction of the predicted proteome, and expanding the analysis to different cell lines and tissues in the future will improve the overall proteomic coverage and define possible tissue-specific aspects of the map. We presume that some of the baits that failed to produce high-quality coAP-MS results may be due to interference of a C-terminal tag with protein function. For the future we note that the C-terminally tagged baits have also been tagged at the N terminus (Yu et al., 2011; http://www.fruitfly.org/EST/proteomics.shtml), possibly circumventing such inactivation.

The evolutionary comparisons illustrated in Figure 6 provide valuable means to explore gene function and to recognize functionally important protein interactions implied by the map. Examining the evolution of protein complex architecture across species can help establish or confirm distant orthologous relationships and improve annotation of orphan genes. The extent of protein conservation is linked to their ability to interact with other proteins, the nature of interactions, and how essential a protein function is for the cell (Mintseris and Weng, 2005; Wuchty, 2004). Our data support models of protein network evolution that are driven by the acquisition or loss of protein complex members rather than rewiring of existing components (van Dam and Snel, 2008; Yamada and Bork, 2009). A more detailed structural analysis will be necessary to examine the subunit interactions in those complexes where the level of conservation is low.

The DPiM establishes a singular resource and a baseline to explore dynamic properties of the protein interaction network in a metazoan proteome. It also enables the analysis of specific subproteomic spaces at greater depth. It is now possible to examine if and how the protein complex relationships derived from S2R+ cells change in different developmental or genetic backgrounds. To promote such studies, we are producing transgenic fly lines carrying the same FLAG-HA tagged version of the proteins under the control of a UAS promoter (https://interfly.med.harvard.edu/transgenic_info.php). The expression of tagged proteins can be spatiotemporally regulated by the use of different Gal4 drivers. Exploring the dynamic nature of the protein complex network defined here, enhanced through the use of quantitative mass spectrometry, will be of fundamental value and will likely provide system-wide insights into the molecular defects underlying pathogenic conditions. We expect that analogies of protein interaction relationships between *Drosophila* and humans will be helpful in the analysis of disease-related pathways and, indeed, the identification and evaluation of disease-related targets.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification

Open reading frames were transferred from the BDGP *Drosophila* melanogaster expression-ready clone set to the pMK33-C-FLAG-HA acceptor vector (Yu et al., 2011). Each clone was transiently transfected into a 54 ml culture of *Drosophila* S2R+ cells. Protein expression was induced with 0.35 mM $CuSO_4$ and whole-cell lysates prepared in lysis buffer (25 mM NaF, 1 mM Na_3VO_4, 50 mM Tris [pH 7.5], 1.5 mM $MgCl_2$, 125 mM NaCl, 0.2% IGEPAL, 5% glycerol, and Complete). Each clarified lysate was bound overnight to 75 μl of crosslinked HA immunoaffinity resin (Sigma). Unbound proteins were washed off with lysis buffer followed by PBS and then bound protein complexes were competitively eluted using synthetic HA peptide YPYDVPDYA (250 μg/ml) in PBS.

Mass Spectrometry and Data Analysis

The copurified proteins were precipitated using trichloroacetic acid, washed with acetone, dried, digested overnight with trypsin, and analyzed by LC-MS/MS. The spectral data were searched with SEQUEST (Eng et al., 2008) against a database of *D. melanogaster* proteins derived from FlyBase version 5.23. The LC-MS/MS identifications were filtered to, on average, a 1.2% protein FDR and 0.3% peptide FDR. The compiled data set was filtered to a combined 0.8% FDR, and further post-processing was used to correct for column carryover issues.

Bioinformatic Analysis

Both bait-prey and prey-prey protein interactions from coAP-MS data were analyzed and scored using HGSCore—a hypergeometric distribution error model, incorporating TSCs to improve the accuracy of co-occurrence prediction. A randomized data set of similar size was created to estimate FDR. Protein interactions were clustered using MCL (Enright et al., 2002). Other algorithms were implemented as described in original literature. Additional details are provided in the Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, seven tables, and one data file and can be found with this article online at doi:10.1016/j.cell.2011.08.047.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (NIH 5RO1HG003616) to S.A.-T. and a fellowship from the Deutsche José Carreras Leukämie-Stiftung e. V. to J.-F.R. Generation of the clone set was supported by a grant from the National Human Genome Research Institute (NHGRI P41HG3487 to S.E.C.). Special thanks to Anne-Claude Gavin, Bernhard Kuster, and Charlie Cohen, whose help was critical in the initiation of the project, as well as Gerry Rubin for help throughout. We thank Norbert Perrimon for S2R+ cells, Lucy and Peter Cherbas for help in cell culture, William Gelbart and the FlyBase team for making DPiM data available on FlyBase, and Alexey Veraksa, Ashim Mukherjee, Kadalmani Krishnan, Mathew Sowa, Dan Finley, Robin Reed, Angeliki Louvi, and members of the S.A.-T., S.E.C., S.P.G., and K.V. labs for helpful discussion and comments. We thank members of CCSB, Vidal and Harper labs, and David Hill and Eric Bennett in particular, for help with the human ORFeome collection. We also thank Manolis Kellis and Rogerio Candeias for their help.

Received: May 7, 2011
Revised: July 14, 2011
Accepted: August 19, 2011
Published: October 27, 2011

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185–2195.

Audhya, A., McLeod, I.X., Yates, J.R., and Oegema, K. (2007). MVB-12, a fourth subunit of metazoan ESCRT-I, functions in receptor downregulation. PLoS One *2*, e956.

Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the human autophagy system. Nature *466*, 68–76.

Belote, J.M., and Zhong, L. (2009). Duplicated proteasome subunit genes in *Drosophila* and their roles in spermatogenesis. Heredity *103*, 23–31.

Bennett, D., Lyulcheva, E., Alphey, L., and Hawcroft, G. (2006). Towards a comprehensive analysis of the protein phosphatase 1 interactome in *Drosophila*. J. Mol. Biol. *364*, 196–212, Erratum: (2006). J. Mol. Biol. *387*, 518.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., et al. (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat. Cell Biol. *6*, 97–105.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043–1046.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., et al. (2011). The transcriptional diversity of 25 *Drosophila* cell lines. Genome Res. *21*, 301–314.

Chintapalli, V.R., Wang, J., and Dow, J.A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. *39*, 715–720.

Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat. Methods *8*, 70–73.

Dombrádi, V., Axton, J.M., Brewis, N.D., da Cruz e Silva, E.F., Alphey, L., and Cohen, P.T. (1990). *Drosophila* contains three genes that encode distinct isoforms of protein phosphatase 1. Eur. J. Biochem. *194*, 739–745.

Eng, J.K., Fischer, B., Grossmann, J., and Maccoss, M.J. (2008). A fast SEQUEST cross correlation algorithm. J. Proteome Res. *7*, 4598–4602.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M., et al. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol. Syst. Biol. *3*, 89.

Forsburg, S.L. (2004). Eukaryotic MCM proteins: beyond replication initiation. Microbiol. Mol. Biol. Rev. *68*, 109–131.

Freilich, S., Oron, E., Kapp, Y., Nevo-Caspi, Y., Orgad, S., Segal, D., and Chamovitz, D.A. (1999). The COP9 signalosome is essential for development of *Drosophila melanogaster*. Curr. Biol. *9*, 1187–1190.

Fricke, B., Heink, S., Steffen, J., Kloetzel, P.M., and Krüger, E. (2007). The proteasome maturation protein POMP facilitates major steps of 20S proteasome formation at the endoplasmic reticulum. EMBO Rep. *8*, 1170–1175.

Gavin, A.C., Maeda, K., and Kühner, S. (2011). Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. Curr. Opin. Biotechnol. *22*, 42–49.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. (2003). A protein interaction map of *Drosophila melanogaster*. Science *302*, 1727–1736.

Goshima, G., and Kimura, A. (2010). New look inside the spindle: microtubule-dependent microtubule generation within the spindle. Curr. Opin. Cell Biol. *22*, 44–49.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. *36*, 59–74.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. Nature *471*, 473–479.

Guerrero, C., Milenkovic, T., Przulj, N., Kaiser, P., and Huang, L. (2008). Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. Proc. Natl. Acad. Sci. USA *105*, 13333–13338.

Hart, G.T., Lee, I., and Marcotte, E.R. (2007). A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. BMC Bioinformatics *8*, 236.

Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H., and Lührmann, R. (2009). Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. Mol. Cell. Biol. *29*, 281–301.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature *415*, 180–183.

Hölzl, H., Kapelari, B., Kellermann, J., Seemüller, E., Sümegi, M., Udvardy, A., Medalia, O., Sperling, J., Müller, S.A., Engel, A., and Baumeister, W. (2000). The regulatory complex of *Drosophila melanogaster* 26S proteasomes. Subunit composition and localization of a deubiquitylating enzyme. J. Cell Biol. *150*, 119–130.

Hu, P., Janga, S.C., Babu, M., Díaz-Mejía, J.J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., et al. (2009). Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. PLoS Biol. *7*, e96.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. Genome Res. *12*, 37–46.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. *38* (Database issue), D355–D360.

Kato, J.Y., and Yoneda-Kato, N. (2009). Mammalian COP9 signalosome. Genes Cells *14*, 1209–1225.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature *440*, 637–643.

Kühner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., et al. (2009). Proteome organization in a genome-reduced bacterium. Science *326*, 1235–1240.

Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P., et al. (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics *89*, 307–315.

Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H., and Finley, D. (2002). Multiple associated proteins regulate proteasome structure and function. Mol. Cell *10*, 495–507.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. (2004). A map of the interactome network of the metazoan *C. elegans*. Science *303*, 540–543.

Liu, H., Sadygov, R.G., and Yates, J.R., 3rd. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. *76*, 4193–4201.

Michelet, X., Djeddi, A., and Legouis, R. (2010). Developmental and cellular functions of the ESCRT machinery in pluricellular organisms. Biol. Cell *102*, 191–202.

Mintseris, J., and Weng, Z. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. Proc. Natl. Acad. Sci. USA *102*, 10930–10935.

Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G., 3rd, and Finley, R.L., Jr. (2011). DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. Nucleic Acids Res. *39* (Database issue), D736–D743.

Ohtaki, A., Noguchi, K., and Yohda, M. (2010). Structure and function of archaeal prefoldin, a co-chaperone of group II chaperonin. Front. Biosci. *15*, 708–717.

Rees, J.S., Lowe, N., Armean, I.M., Roote, J., Johnson, G., Drummond, E., Spriggs, H., Ryder, E., Russell, S., Johnston, D.S., et al. (2011). In vivo analysis of proteomes and interactomes using parallel affinity capture (iPAC) coupled to mass spectrometry. Mol. Cell. Proteomics *10*, M110 002386.

Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.O., et al. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. Genome Res. *14* (10B), 2128–2135.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173–1178.

Shaw, G., Morse, S., Ararat, M., and Graham, F.L. (2002). Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. FASEB J. *16*, 869–871.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389–403.

Stanyon, C.A., Liu, G., Mangiola, B.A., Patel, N., Giot, L., Kuang, B., Zhang, H., Zhong, J., and Finley, R.L., Jr. (2004). A *Drosophila* protein-interaction map centered on cell-cycle regulators. Genome Biol. *5*, R96.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell *122*, 957–968.

Südhof, T.C., and Rothman, J.E. (2009). Membrane fusion: grappling with SNARE and SM proteins. Science *323*, 474–477.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. *13*, 2129–2141.

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., and Zhang, H.; FlyBase Consortium. (2009). FlyBase: enhancing *Drosophila* Gene Ontology annotations. Nucleic Acids Res. *37* (Database issue), D555–D559.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature *403*, 623–627.

van Dam, T.J., and Snel, B. (2008). Protein complex evolution does not involve extensive network rewiring. PLoS Comput. Biol. *4*, e1000132.

Wang, X., Chen, C.F., Baker, P.R., Chen, P.L., Kaiser, P., and Huang, L. (2007). Mass spectrometric characterization of the affinity-purified human 26S proteasome complex. Biochemistry *46*, 3553–3565.

Wee, S., Hetfeld, B., Dubiel, W., and Wolf, D.A. (2002). Conservation of the COP9/signalosome in budding yeast. BMC Genet. *3*, 15.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, a003707. Published online December 20, 2010. 10.1101/cshperspect.a003707.

Williams, R.D., Hing, S.N., Greer, B.T., Whiteford, C.C., Wei, J.S., Natrajan, R., Kelsey, A., Rogers, S., Campbell, C., Pritchard-Jones, K., and Khan, J. (2004). Prognostic classification of relapsing favorable histology Wilms tumor using cDNA microarray expression profiling and support vector machines. Genes Chromosomes Cancer *41*, 65–79.

Wuchty, S. (2004). Evolution and topology in the yeast protein interaction network. Genome Res. *14*, 1310–1314.

Yamada, T., and Bork, P. (2009). Evolution of biomolecular networks: lessons from metabolic and protein interactions. Nat. Rev. Mol. Cell Biol. *10*, 791–803.

Yu, C., Wan, K.H., Hammonds, A.S., Stapleton, M., Carlson, J.W., and Celniker, S.E. (2011). Development of expression-ready constructs for generation of proteomic libraries. Methods Mol. Biol. *723*, 257–272.

Yu, J., Pacifico, S., Liu, G., and Finley, R.L., Jr. (2008). DroID: the *Drosophila* Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics *9*, 461.