Control of mRNA 3'-end formation during *Drosophila* neural development: mechanisms and biological roles

Raúl Alejandro Vallejos Baier

Submitted in fulfilment of the requirements for the

degree of Doctor of Philosophy at the University of Sussex

September 2016

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signed:……..

Raúl Alejandro Vallejos Baier

UNIVERSITY OF SUSSEX

Raúl Alejandro Vallejos Baier, DPhil Biology

Control of mRNA 3'-end formation during *Drosophila* neural development: mechanisms and biological roles

Summary

The process of alternative polyadenylation (APA) is a widespread gene regulatory mechanism that generates mRNAs with different 3' ends, allowing mRNAs to interact with different sets of RNA regulators such as microRNAs and RNA-binding proteins. Recent studies have shown that during development in both insects and mammals, mRNAs with extended 3' UTRs are restricted to the nervous system suggesting that extended 3' UTRs might play important roles during the formation and function of the nervous system. With its powerful genetics *Drosophila* emerges as an excellent system to study the molecular mechanisms and biological roles of APA within the physiological context of neural development. Much of the work is centred in the roles of the Cleavage Factor I (CFI) complex, because (i) is the complex with the highest evolutionary conservation between humans and *Drosophila*, (ii) it is expressed at very high level in neural tissues, and (iii) has a well-established structure and function in mammalian cells. Through the combination of genetic, molecular and genomic databases, I first show that the cleavage and polyadenylation (CPA) machinery in *Drosophila* is as complex as its human counterpart and shows an enrichment of expression in the nervous system. Secondly, using a suite of genetic and behavioural methods, I show that a mutation in the *Drosophila* orthologue of CFI25 affects feeding in the *Drosophila* larvae and is required for major developmental transitions. Third, I explore the mechanisms by which APA is controlled in the developing nervous system by CFI factor depletion. As a result of this, genes with reported neural 3' UTR extensions change their patterns of APA. Altogether, this work adds to the current understanding of the phenomenon of APA within the nervous system and gives new insights on the biological roles of CPA factors for behaviour and neural function.

Acknowledgements

First, I want to thank my supervisor, Claudio, for the opportunity I received to learn science and the way it works in his laboratory.

I also thank all members of the Alonso Lab, past and present – Joao, Pedro, Aalia, Agata, Sofia, Wan, Ines, Bhavna, Eddie, Clare, Buthaina, Casandra and Richard – for the constant support and positive environment. Special thanks to Joao, who constantly supported me towards the end of my PhD and shared with me part of his expertise and knowledge. As well as Sofia, Pedro, Aalia and Bhavna, who kindly helped me with experiments towards the end of my time in the lab.

To all members of the Couso Lab, who never turned their back when I required fly stocks or scientific advice. Special thanks to Unum, who also supported me greatly towards the end of my PhD and gave me valuable help and advice.

I would also like to thank all my friends here in Brighton and in Chile, especially Felipe, Diego, Angela, Tomas, David, Pato, Paula and Juan Pablo. With special thanks to Diego, whose knowledge in statistics helped me greatly during my PhD (Made me feel as an idiot every time he explained me something though). Special thanks also to my neighbours Tomas and Pato, for all the help, advice, wine and barbeques.

I would like to thank “Becas Chile”, for providing me the funding and therefore the opportunity to pursue my PhD project in the UK.

To all my family, especially my dad, mum and brother: Without them, all of this would have been impossible.

To Marcia, Edgardo and Andrés, my family

Table of Contents

List of Figures

List of Tables

Abbreviations

AS	Alternative Splicing
APA	Alternative Polyadenylation
ATP	Adenosine triphosphate
CFI	Cleavage Factor I
CFII	Cleavage Factor II
CNS	Central Nervous System
CPA	Cleavage and Polyadenylation
CPS	Cephalopharyngeal skeleton
CPSF	Cleavage and polyadenylation specificity factor
CSTF	Cleavage Stimulation Factor
DNA	Deoxyribonucleic acid
DSE	Downstream Sequence Element
FISH	Fluorescence in situ hybridisation
IgM	Immunoglobulin M
IPC	Insuling producing cells
L1	First instar larvae
L2	Second instar larvae
L3	Third instar larvae
miRNA	microRNA
PAP	Poly(A) Polymerase
PAS	Polyadenylation signal
PBS	Phosphate buffered saline
PBTx	Phosphate buffered saline with TritonX detergent

PNS	Peripheral Nervous System
Poly(A)	Poly-adenosine
RBP	RNA-Binding Protein
RNA	Ribonucleic acid
RNA PolII	RNA Polymerase II
RNAi	RNA interference
RNA-seq	RNA Sequencing
siRNA	Small Interference RNA
snRNP	Small nuclear ribonucleoprotein
UTR	Untranslated Region
VNC	Ventral nerve cord

Many people would accept that we do not really have knowledge of the world; we have knowledge only of our representations of the world. Yet we seem condemned by our constitution to treat these representations as if they were the world, for our everyday experience feels as if it were of a given and immediate world

Francisco Varela

In *The Embodied Mind*

Chapter 1

General Introduction

1.1 Preface

A central question in modern biology is how the nervous system is formed. Recent experiments have shown that the complexity of neural tissues is correlated to pervasive post-transcriptional modifications. For example, it has been shown that alternative splicing (AS), in which exons are included or excluded in a tissue-specific way and therefore different proteins are produced (Chen and Manley 2009), is particularly widespread in the mammalian nervous system, where neural-specific protein isoforms are expressed in this tissue to contribute to its functional complexity (Raj and Blencowe 2015). Furthermore, these specific patterns of AS within the nervous system can be achieved by RNA-Binding proteins (RBPs) that are also specifically expressed in this tissue (Licatalosi *et al.* 2012). Similarly, other regulators of gene expression, such as microRNAs (miRNAs) can also play neural-specific roles by restricted expression in this tissue and thus targeting neural genes (Kapsimali *et al.* 2007; Wheeler *et al.* 2006). Finally, transcription factors also play important roles for the patterning and functioning of the nervous system in both vertebrates and invertebrates. A good example of this are the *Hox* genes, which have an important role for cell-specification along the anteroposterior axis in bilateral animals (Krumlauf *et al.* 1993) and show neural-specific mRNA processing events, such as 3' UTR extensions achieved by alternative polyadenylation (APA), a phenomenon discovered in *Drosophila* (Thomsen *et al*. 2010; Rogulja-Ortmann *et al*. 2014). APA therefore can control mRNA localization, stability and translation efficiency by differential 3' UTR expression (Di Giammartino *et al.* 2011). Interestingly, as I will discuss in detail in this study, shortly after the first reports of *Hox*-neural 3' UTR extensions, this phenomenon was far from being an oddity of these category

of key developmental genes. Instead, hundreds of genes were shown to display drastically long 3' UTRs during neural development in *Drosophila* (Thomsen *et al.* 2010; Hilgers *et al.* 2011; Smibert *et al.* 2012) and thousands of genes in mammals (Miura *et al.* 2013). Implying that the nervous system, perhaps contrary to other tissues, relies enormously in RNA processing events to be able to develop and function in a reproducible way.

Although the general-consensus explanation for the biological meaning of these 3' UTR extensions is based on differential targeting by trans-acting factors - such as miRNAS and RBPs - it is evident that this is hardly the full reason when considering the extent at which RNA PolII can bypass termination sites that in theory should fulfil these conditions (with modest extensions of less than 1 kb (Patraquim *et al.* 2011)) but instead produce extensions of up to 12 kb (Hilgers *et al.* 2011). These observations point to the question: How is alternative polyadenylation controlled during neural development and what are its biological roles?

The following work addresses the molecular mechanisms that control APA and its biological roles during *Drosophila* neural development. For this, I identify the factors that control mRNA 3' processing in *Drosophila,* and then analyse their expression in different tissues and developmental stages in order to test a prevalent model that proposes the abundance of cleavage and polyadenylation (CPA) factors within cells trigger the selection of alternative polyadenylation sites (PAS) in pre-mRNAs (Takagaki *et al*. 1996).

Given that most studies that have tested the aforementioned model were performed *in vitro* in cells in culture, I explore the expression of CPA factors during the formation of the nervous system to look for the mechanisms that

control neural APA. Further, I focus in the most conserved cleavage and polyadenylation complex subunit between *Drosophila* and humans: Cleavage factor I (CFI), analysing its role not only for APA within the nervous system, but also for behaviour.

In this chapter, I will first introduce the concept of RNA 3' termination and its discovery; second, I will present the discovery of factors that control both cleavage and polyadenylation and describe them in terms of structure and function; third, I will describe the concept of alternative cleavage and polyadenylation and reported cases that showcase its biological roles; fourth, I will discuss the models that have been proposed to explain the mechanisms controlling APA; and finally, I will discuss in detail the phenomenon of extensive 3' UTR extensions in the nervous system.

1.2 Discovery of mRNA polyadenylation in eukaryotic cells

The first pieces of evidence about enzymes that were able to synthetize polynucleotides came from the 1950s by Severo Ochoa and colleagues. They reported an enzyme from the microorganism *Azotobacter vinelandii*, a gram-negative bacterium used to study nitrogen fixation that was able to synthetize highly polymerized polynucleotides from 5'-nucleoside diphosphates in a reaction that required Mg^{++}. Furthermore, these polynucleotides were shown to be made of 5'-nucleoside units linked to one another through 3'-phosphoribose ester bonds. Interestingly, the synthetized polynucleotides had a biochemistry indistinguishable from that of natural RNA (Grunberg-Manago *et al*. 1956; Grunberg-Manago *et al*. 1955). Following this discovery, Arthur Kornberg and colleagues revealed that this process was also reversible and showed this in the

model microorganism *Escherichia coli* (Littauer and Kornberg 1957). Although the work of Severo Ochoa and Arthur Kornberg during the late 1950s about the mechanisms of the biological synthesis of ribonucleic and deoxyribonucleic acids led them to win the Nobel Prize in Physiology or Medicine in 1959, the biological meaning of RNA polynucleotides within the cell remained unclear.

During the next decade it was reported for the first time that an enzyme extracted from calf thymus nuclei was able to synthetize a sequence of adenylate units from ATP, which was acid-insoluble and required Mg^{++} for its polymerization activity. Apart from being the first report of this process in mammalian cells, this was the first study that proposed the existence of these poly(A) polymers at the end of endogenous RNA molecules and that they were tightly bound to proteins (Edmonds and Abrams 1960). Nonetheless, the potential relevance and roles of mRNA polyadenylation as post-transcriptional modifications in 3'-ends of most transcripts of eukaryotic cells were not uncovered until another decade later. In the 1970s it was confirmed by several groups that poly(A) sequences were found in RNA extracted from HeLa cells and mouse sarcoma 180 cells. These long poly(A) sequences made this fraction of the RNA resistant to RNAse treatment and were not present in the translated proteins. Reasonably, it was proposed that these regions were relevant for translation and could act as binding sites for other proteins (Darnell *et al*. 1971; Lee *et al*. 1971; Edmonds *et al*. 1971). It was also proposed that a Poly(A) Polymerase (PAP), such as the one described 10 years earlier, could be the enzyme catalysing these reactions *in vivo* (Edmonds and Abrams 1960). Finally, the first mechanisms by which poly(A) tails were added to mRNA 3'-ends during transcription were elucidated. To the surprise of the scientific field, in which it was generally assumed that 3'end termination occurred

simply by transcriptional termination *per se*, it was shown by experiments with viral RNA that RNA PolII proceeds past the signals for 3'-end termination to be subsequently cleaved and polyadenylated. Thus, CIS elements within the RNA molecule were postulated to exist that are interpreted by the cell for cleavage and addition of poly(A) by a PAP and this was shown to occur before mRNA splicing (Ford and Hsu 1978; Manley *et al*. 1982; Nevins and Darnell 1978). Remarkably, the presence of more than one signal within RNA molecules to be selected for 3'-end termination was already proposed in these studies from the 1970s and early 1980s. However, the factors that control 3'-end processing, the specific CIS elements within RNAs that control cleavage and polyadenylation and the concept of alternative cleavage and polyadenylation with its associated implications still remained unknown.

Nowadays it is known that RNA polyadenylation is shared among Eukarya and the only genes which are cleaved, but not polyadenylated, are replication-dependent histone mRNAs (Dávila and Samuelsson 2008; Marzluff *et al*. 2008). Instead of having a poly(A) tail, these pre-mRNAs contain an RNA stem-loop structure close to their 3' end, which is recognized by the U7 snRNP and proteins such as Stem-loop binding protein (SLBP), together with a subset of factors from the canonical 3' end machinery (Dávila and Samuelsson 2008), which will be discussed in the next section. On the other hand, the primary structure and biochemical characterization of the bovine PAP was done in 1991 by both James Manley and colleagues and by Elmar Wahle independently (Manley *et al*. 1991; Wahle 1991b). It was not until the year 2000 that the crystal structure of the mammalian PAP was resolved (Martin *et al*. 2000). As a result, the key domains for its function were described: It has an N-terminal catalytic domain which

polymerises adenosines using ATP as a substrate and an RNA-Binding region that overlaps with a nuclear localization signal (NLS) near the C-terminus. Additionally, vertebrates have an extra region towards the C-terminal domain which is rich in serines and threonines, making it a target for post-translational modifications, such as phosphorylation at multiple sites (Martin *et al*. 2000).

During recent years, studies addressing the question on the biological meaning of mRNA polyadenylation within cells have made enormous progress since the first reports on this process in the 1960s. For example, it has been shown that the median length of the poly(A) tail is a property conserved among different species, with mammalian cells having 69-96 nucleotides, plants (*Arabidopsis thaliana*) and *Drosophila* S2 cells having 50-51 nucleotides, and budding (*Saccharomyces cerevisiae*) and fission (*Schizosaccharomyces pombe*) yeast having 27 and 28 nucleotides. What is more, the length of the poly(A) tails are conserved between orthologous mRNAs (Subtelny *et al.* 2014). In terms of function, it has been shown that the length of the poly(A) tail is tightly regulated during the somatic cell cycle and controls translation efficiency. For example, the cell-cycle regulatory genes *CDK1*, *TOP2A* and *FBXO5* present a dramatic decrease in the length of their poly(A) tail which represses their translation during M-phase (Park *et al.* 2016). In this same study, the authors show that for genes to be able to escape this translational repression observed in M-phase, a terminal oligopyrimidine (TOP) tract is required at the end of their poly(A) tail. Lastly, the length of the poly(A) tail has been shown to be coupled to translational efficiency during the early development in zebra fish (*Danio rerio*) and in frogs (*Xenopus laevis*) embryos. However, after gastrulation, this coupling diminishes and goes to the point of being absent in non-embryonic tissues, highlighting the role of

poly(A) tails to work as a developmental switch for translational control during early embryogenesis (Subtelny *et al.* 2014).

1.3 Protein factors and RNA CIS elements that control mRNA 3'-end cleavage and polyadenylation

In the previous section we discussed the discovery of mRNA polyadenylation as a general post-transcriptional modification observed across all eukaryotes. Also, I showed that the first pieces of evidence about one of the members of what is now known as the "Cleavage and Polyadenylation" (CPA) machinery was discovered almost 60 years ago: the Poly(A) polymerase (PAP) (Edmonds and Abrams 1960). In this section I will discuss all the other factors that control 3'-end formation and briefly their structure and how they work, according to the latest studies in this emerging field in modern biology.

The first members of the mammalian cleavage and polyadenylation (CPA) machinery were initially discovered by biochemical fractionation approaches in the late 1980s (Gilmartin and Nevins 1989; Takagaki *et al*. 1989). In a study conducted by James Manley and colleagues in 1988 (Takagaki *et al*. 1988), the authors provided evidence for the existence of a "cleavage/specificity" factor (CSF) that could efficiently cleave SV40 late pre-mRNAs at a poly(A) addition site and that this factor could be separated chromatographically from PAP. Interestingly, while isolated PAP could perform its functions *in vitro* only in a non-specific manner, addition of CSF caused it to function in a sequence-specific way by requiring the signal "AAUAAA". This sequence was shown several years earlier by Nick Proudfoot and George Brownlee to be present in a variety of RNA

molecules about 20 nucleotides from the poly(A) tail (Proudfoot and Brownlee 1976) and now is known as the Polyadenylation signal (PAS).

After the description of CSF in this study from 1988, James Manley and colleagues showed that CSF could be further fractionated into four different subunits. The first one, known as “Specificity factor” (SF), was required for both specific cleavage and for specific polyadenylation. Two other factors, coined as “cleavage factors I and II” (CFI and CFII) were sufficient for accurate cleavage of pre-mRNAs when mixed with SF. The last factor, coined as “Cleavage stimulation factor” (CSTF), enhanced the efficiency of the cleavage reaction significantly when added to a mixture with the other three factors. Furthermore, none of these factors contained or required RNA components for their function, apart from the substrate pre-mRNA itself. Finally, PAP was shown to be a necessary component to cleave several other pre-mRNAs, apart from the one tested (Takagaki *et al*. 1989). Since the early 1990s, it was shown that each one of these factors discovered in 1989 was made of more subunits: CSF was re-named to “Cleavage and polyadenylation specificity factor” (CPSF) and shown to be made of four subunits of 160, 100, 73 and 30 kDa after purification and fractionation from calf thymus. This multisubunit CPA complex was also shown to bind to the poly(A) signal “AAUAAA” and be required for both cleavage and polyadenylation (Bienroth *et al*. 1991; Murthy and Manley 1992). More than a decade later, another member of CPSF was described in human cells (HeLa), Fip1 (Kaufmann *et al.* 2004). Intriguingly, although the presence of an endonuclease was required to be part of the CPA machinery to cleave the pre-mRNA substrates, its identity was not uncovered until 2006, when it was revealed to be also member of CPSF (CPSF73) (Mandel *et al.* 2006), a result consistent with the evidence for CPSF

as required for cleavage. Additionally, another factor from CPSF was discovered in even more recent years, such as WDR33 in 2009 (Shi *et al*. 2009).

To summarise, the current information about CPSF shows that is composed of six subunits: CPSF160, CPSF100, CPSF73, CPSF30, Fip1 and WDR33. CPSF160 is the largest subunit. In mammals, it is composed of tandem WD40 repeats clustered into three major β-propellers (Neuwald and Poleksic 2000) and can be UV-crosslinked to pre-mRNAs in a poly(A) signal (PAS)-dependent manner (Keller *et al.* 1991). The mammalian CPSF100 is similar to CPSF73 in structure, both belonging to the β-CASP family (Callebaut *et al.* 2002), but while CPSF73 acts as a hydrolase in coordination with metal ions (Mandel *et al.* 2006), CPSF100 does not have functional motifs to bind zinc ions, making this factor incapable of catalysis (Dominski *et al.* 2005). CPSF30 is the smallest CPSF subunit and in mammals consists of five CCCH zinc finger motifs and a CCHC zinc knuckle motif at its C terminus, which is absent in its yeast homolog Yth1 (Puck *et al.* 1997). Fip1, as mentioned earlier, was discovered in 2004 and stably associates with all other members of CPSF, being required for both cleavage and polyadenylation (Kaufmann *et al.* 2004). The human version is almost twice as large as its yeast counterpart and is similar in domain organization only towards its acidic N-terminus (Preker *et al*. 1995). This version also has an arginine-rich RNA-Binding Motif towards its C-terminus that binds preferentially to U-rich sequence elements within pre-mRNAs, which yeast do not have (Kaufmann *et al.* 2004). Lastly, WDR33, as mentioned earlier, was the last component of CPSF discovered (Shi *et al*. 2009). Although its exact molecular function is not well understood, its structure has been described in mammals as having an N-terminal WD40 domain, a middle collagen-like domain, and a C-terminal glycine-

proline-arginine (GPR) domain (Ito *et al.* 2001). A representation of the domain architecture of the members of CPSF in humans is shown in Figure 1.1.

Of the four previously described original components required for cleavage and polyadenylation presented in 1989 by James Manley and colleagues (CSF, CSTF, CFII and CFII) (Takagaki *et al*. 1989), CSTF was shown to be composed of three subunits of 77, 64 and 50 kDa and was required for efficient cleavage of pre-mRNA substrates (Takagaki *et al*. 1990). CSTF77 is the largest subunit of CSTF and bridges CSTF64 (which binds RNA) with CSTF50 (which interacts with other proteins). The molecular structure of CSTF77 shows a HAT (Half a TPR) domain towards its N-terminal, which has been involved in RNA processing (Preker and Keller 1998). The molecular structure of this domain in mice shows that CSTF77 can dimerize through the HAT domain (Bai *et al.* 2007), which makes the whole CSTF complex work as a dimer, with two of each one of its components. In the aforementioned study that described the CSTF complex in 1990 (Takagaki *et al*. 1990) the authors show that CSTF64 could be UV-crosslinked with “AAUAAA” containing RNAs, similarly as the case for CPSF160 (Keller *et al.* 1991). CSTF64 was later shown to bind pre-mRNAs in U-rich sequences downstream of the PAS and be able to influence the site of cleavage by CPSF (MacDonald *et al*. 1994).

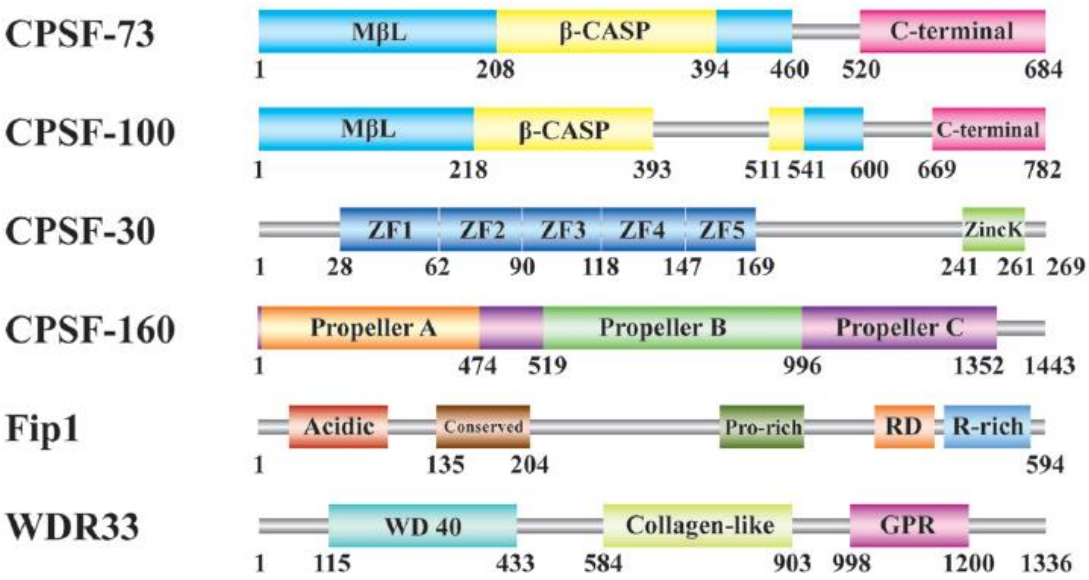


Figure 1.1 Domain organization of the human CPSF subunits

Diagram of the organization of domains of the six different members of CPSF in humans. N-terminal is to the left and C-terminal is to the right. “MβL” stands for “metallo-β-lactamase domain”. “ZF” stands for “Zinc Finger domain”. “ZincK” stands for “Zinc Knuckle Motif”. “RD” means “Arginine and aspartate rich domain”. “R” means arginine-rich domain. And “GPR” means “glycine-proline-arginine domain”. Diagram taken from (Xiang *et al*. 2014).

Three years later, it was shown that this U-rich sequence, which could also be GU-rich, was the binding site for CSTF64 downstream of the PAS and was known as the “Downstream sequence element” (DSE) (Takagaki and Manley 1997). The molecular structure of CSTF64 showed an RNA Recognition Motif towards its N-terminus and a repeated structure in its C-terminal region in which a pentapeptide sequence (MEARA/G) is repeated 12 times. Also, a segment of approximately 270 amino acids surrounds this repeat and is highly enriched in proline and glycine (Takagaki *et al*. 1992). CSTF50 is the smallest subunit of CSTF and is only existent in metazoans. Similarly to CSTF77, it contains an N-terminal dimerization domain and seven WD40 repeats in its C-terminal, further supporting the role of the CSTF complex as a dimer (Takagaki and Manley 2000). A representation of the domain architecture of the members of CSTF in humans is shown in Figure 1.2.

The CFI complex was purified and biochemically characterized in 1996 (Ruegsegger *et al*. 1996). In mammals, it is composed of four subunits: CFI25, CFI68, CFI59 and CFI72. While CFI59 is a mammalian paralog of CFI68, CFI72 is an isoform of CFI68 (Ruepp *et al*. 2011). The CFI complex, which is also only present in metazoans, also acts as a dimer, with two CFI25 units and two CFI68 units (or CFI59 or CFI72). Nonetheless, both CFI59 and CFI72 are functionally redundant with CFI68 (Ruegsegger *et al*. 1998). Although we will discuss the structure and roles of CFI in detail in subsequent chapters in line with experimental results, we can mention that CFI binds to RNA even in the absence of the already mentioned “AAUAAA” hexamer, which works as the canonical PAS. Furthermore, CFI was described as a determinant of poly(A) site recognition in an PAS-independent manner by recruiting Fip1 and PAP to the

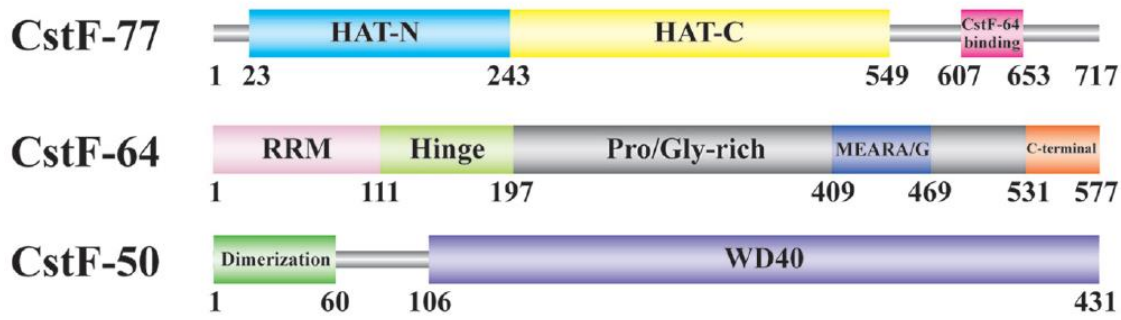


Figure 1.2 Domain organization of the human CSTF subunits

Diagram of the organization of domains of the three different members of CSTF in humans. N-terminal is to the left and C-terminal is to the right. “HAT” stands for “Half a TPR”. “RRM” stands for “Rna Recognition Motif”. Diagram taken from (Xiang *et al*. 2014).

RNA substrate and acting during the initial stages of transcription (Venkataraman *et al*. 2005). This binding site for CFI was discovered by SELEX experiments and revealed to be UGUAN (N = A > U >= C/G) (Brown and Gilmartin 2003), which is located upstream of the PAS (contrary to the CSTF binding site “DSE” located downstream from the PAS). After the crystal structure of CFI25 in association with RNA was resolved, the authors showed that two “UGUA” elements are bound by each one of the CFI25 subunits at the same time, highlighting the potential regulatory roles of CFI25 for PAS selection, as we will soon discuss (Yang *et al*. 2010). The molecular architecture of the human CFI25 shows a central Nudix domain, which is a widespread family of proteins that act as pyrophosphohydrolases. CFI25 nonetheless presents changes in this catalytic domain causing the protein to be unable to function as a hydrolase. Instead, its Nudix domain is used to bind RNA, as well as CFI68 (Trésaugues *et al.* 2008; McLennan 2006). CFI25 also has an RNA-binding domain towards its N-terminal from the Nudix domain. Given that – as mentioned earlier - CFI25 acts as a dimer within the CFI complex with two UGUA sequences bound in an anti-parallel fashion (Yang *et al*. 2010), a model was proposed in which CFI25 binds UGUAs neighbouring different PAS selectively and loops out the pre-mRNA molecule, proposing a mechanism by which the process of alternative polyadenylation, which we will discuss in the next section, occurs at the molecular level (Yang *et al*. 2011). CFI68 on the other hand has a different architecture than CFI25: it has a proline-rich region in the middle and an RS domain towards its C-terminal. Interestingly, this features resemble the structure of splicing factors (Ruegsegger *et al*. 1998). What is more, CFI68 has been shown to interact with factors from

the splicing machinery including U1 snRNP and U2AF 65 (Awasthi and Alwine 2003; Millevoi *et al*. 2006), bridging 3' end processing with mRNA splicing.

In addition to the domains present in CFI68 and absent in CFI25, CFI68 has an RNA Recognition Motif (RRM) in its N-terminal. Nonetheless, the affinity of this domain for RNA is rather weak, but it is enhanced after interaction with CFI25, in which each one of the CFI25 units binds one CFI68 through its RRM, forming a tetramer with both CFI25 and CFI68 proteins facing each other (Li *et al.* 2011). A representation of the domain architecture of the members of CFI in humans, together with the proposed mechanism for PAS selection, is shown in Figure 1.3.

CFII is composed of two factors, Pcf11 and Clp1, and although this complex has not been fully characterized in mammals (because it has been mostly studied in yeast), there are studies covering their roles in human cells. For example, depletion of Pcf11 in HeLa cells was shown to reduce 3'-end termination efficiency. Also, evidence was provided for Pcf11 being required for degradation of the 3' product after cleavage has occurred (West and Proudfoot 2008). The molecular structure of this complex in humans is not well understood, although it is known to be twice as large as its yeast counterpart and to have sequence homology at its N-terminal, where it has a "CTD Interacting Domain" (CID) (de Vries *et al.* 2000). Clp1 in humans is better studied than its partner Pcf11 and surprisingly, it shows diverse biological roles beyond 3' end formation in humans. For example, Clp1 acts as a 5'-OH polynucleotide kinase and has been involved in the activation of siRNAs; these last molecules that are incorporated into the "RNA-induced silencing complex" (RISC) for gene silencing (Weitzer and Martinez 2007).

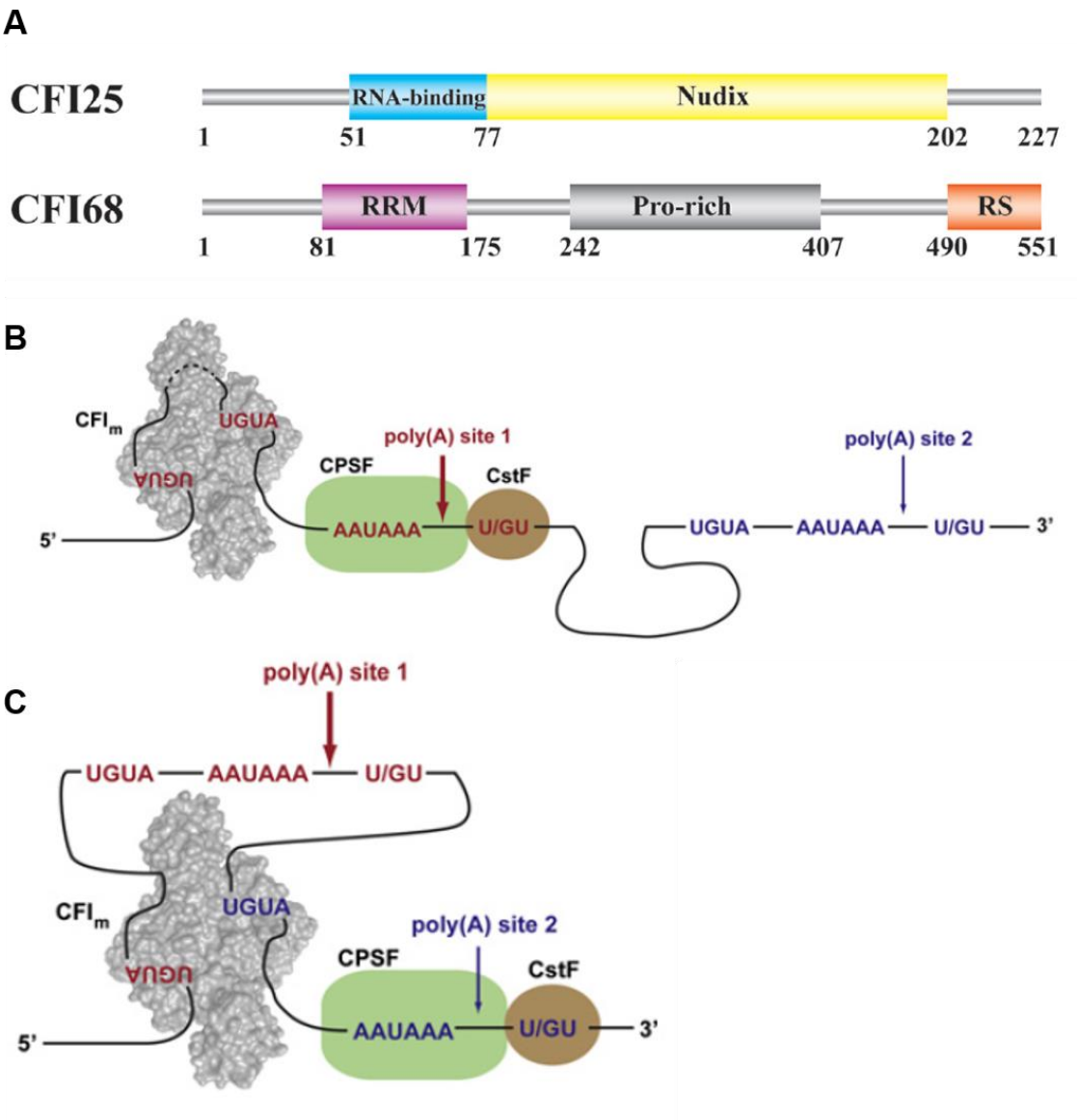


Figure 1.3 Domain organization of the human CFI subunits and model for PAS selection

(Legend on the following page)

Figure 1.3 Domain organization of the human CFI subunits and model for PAS selection

(A) Diagram of the domain organization in the two different members of CFI in humans. N-terminal is to the left and C-terminal is to the right. “RRM” stands for “Rna Recognition Motif”. “RS” means “arginine and serine domain”. Diagram taken from (Xiang *et al.* 2014). **(B-C)** Proposed model for alternative PAS selection by CFI **(B)**. When the “Poly(A) site 1” is selected, the two closest UGUA elements are bound simultaneously by CFI in an antiparallel fashion, allowing CPSF to bind to the “AAUAAA” PAS sequence and CSTF to the DSE within this poly(A) site. **(C)** When “Poly(A) site 2” is selected, only the first UGUA element is bound by CFI, while the UGUA element neighbouring the Poly(A) site 2 is used (looping out all the mRNA elements between these two sequences, including the “Poly(A) site 1”). Thus, both the PAS and the DSE (to be bound by CPSF and CSTF) are used within “poly(A) site 2”. Diagram taken from (Yang *et al.* 2011).

Furthermore, Clp1 has also been shown to be involved in tRNA splicing (Ramirez *et al*. 2008). Regarding 3' end processing, Clp1 interacts with components of CPSF and CFI, although its role as a kinase is not required for cleavage and polyadenylation (de Vries *et al.* 2000). Human Clp1 shows high sequence homology with its yeast counterpart, which has a central ATPase domain and two other smaller domain in its termini (de Vries *et al.* 2000). A representation of the domain architecture of the members of CFII is shown in Figure 1.4. The yeast members are shown because their structure has been fully described (which is not being the case for the human counterparts).

So far, I have described the components that form part of each one of the complexes involved in cleavage and polyadenylation: CPSF, CSTF, CFI and CFII. I have also described briefly their function and structure in human cells, with the exception of CFII, which is much better studied in yeast. During the 1990s, more members of the CPA machinery were discovered, which do not form part of any of the above-mentioned complexes. In this work, we will categorize them as the "Non-Complex" group. One of the member of this group is PAP, which was already described in this Chapter.

In 1996, a protein named "Symplekin" was described as being localized in the cytoplasmic face of the plaque associated with the tight junction-containing zone of polar epithelial cells and Sertoli cells from testis (Keon *et al.* 1996). Its association with the CPA machinery only came four years after, when it was shown to interact with the CSTF complex, more specifically with CSTF64 (Takagaki and Manley 2000). Symplekin acts as a scaffolding protein, which bridges different components of the CPA machinery together (its name comes

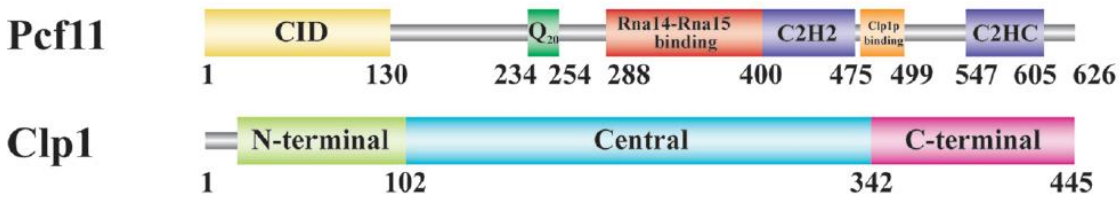


Figure 1.4 Domain organization of the yeast CFII subunits

Diagram of the organization of domains of the two different members of CFII in yeast. N-terminal is to the left and C-terminal is to the right. “CID” stands for “CTD Interacting Domain”. “Q_{20}” means a consecutive glutamines domain. “C2H2” and C2HC” are Zinc Finger domains. Diagram taken from (Xiang *et al*. 2014).

from a Greek word that means “to tie together, to weave, to be intertwined”, coined in the cited study from 1996). Its structure presents binding domains for CPSF73 and CSTF64. Moreover, Symplekin binds to CSTF64 by competing with CSTF77 in a mutually exclusive manner, and while its interaction with CSTF64 is limiting for histone pre-mRNA processing, it is relatively unimportant for general cleavage and polyadenylation (Ruepp *et al*. 2011).

The last core CPA factor that forms part of the “Non-Complex” group is PABPN1, which binds to the poly(A) tail and has a cytoplasmic counterpart named PABP. This last was discovered in 1973 (Blobel 1973). PABPN1 was instead discovered nearly 20 years later by Elmar Wahle (Wahle 1991a) (whom as mentioned earlier also characterized PAP independently from James Manley (Wahle 1991b)).

In the first of these studies by Wahle, PABPN1 was shown to interact with CPSF and bind to the poly(A) tail in the nucleus. More specifically, a transition from a slow initiation phase of polyadenylation to rapid elongation occurred when the poly(A) tail was long enough to act as a binding platform for PABPN1. Thus, PABPN1 controls PAP efficiency. Also, Elmar Wahle correctly speculated that PABPN1 could control the length of the poly(A) tail (Wahle 1991a). Four years later, this same author demonstrated that the addition of either CPSF or PABPN1 separately to a mix only gave moderate processivity to *in vitro* RNA polyadenylation reactions. However, when they acted together, a rapid addition of poly(A) was observed, reaching a limit of 200 to 300 nt (Wahle 1995).

The molecular structure of PABPN1 shows that in contains a single RNA-recognition motif (RRM), contrary to its cytoplasmic counterpart, which contains four. Furthermore, this RRM separates an arginine-rich C-terminal domain from an acidic N-terminal domain, which are essential for poly(A) binding and

stimulation of PAP, respectively (Ge *et al.* 2008). A representation of the domain architecture of the “Non-Complex” CPA factors in humans is shown in figure 1.5.

Although it plays a much more general role in RNA transcription, the largest subunit of RNA PolII participates (among other processes) in 3’-end formation through its C-terminal domain (CTD) by interacting with CPA complexes such as CPSF and CSTF both *in vivo* and *in vitro* (Hirose and Manley 1998; McCracken *et al*. 1997). Pcf11, as I mentioned earlier, has a CID domain which also interacts with the CTD of RNA PolII, bridging 3’end processing with transcription (de Vries *et al.* 2000). A diagram representing the CPA machinery bound to RNA at its key CIS regulatory elements is shown in Figure 1.6.

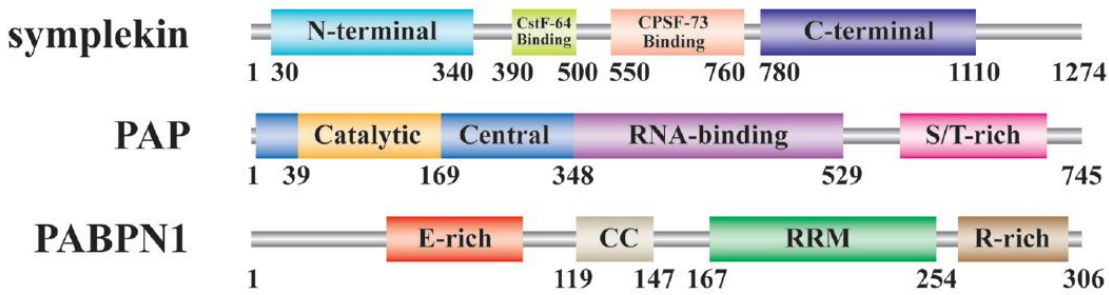


Figure 1.5 Domain organization of the "Non-Complex" CPA factors in humans

Diagram of the organization of domains of the three "Non-Complex" CPA factors in humans. N-terminal is to the left and C-terminal is to the right. "S/T-rich" means "Serine and threonine rich domain". "E-rich" means "Glutamate rich domain". "CC" stands for "Coiled-coil" domain. "RRM" stands for "RNA Recognition Motif" and "R-Rich" means "Arginine reach domain". Diagram taken from (Xiang *et al*. 2014).

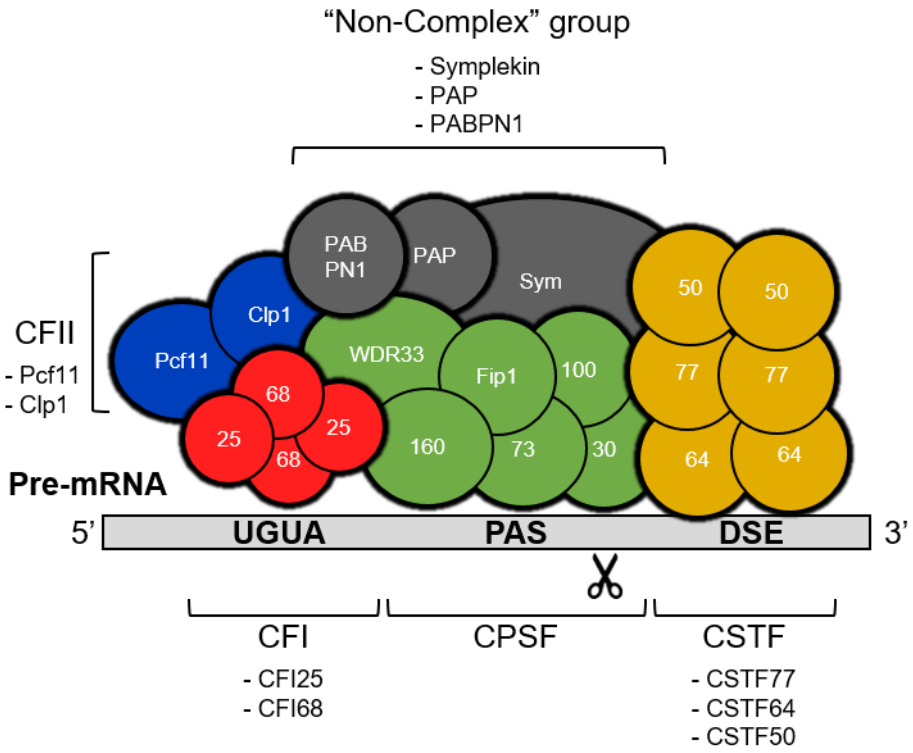


Figure 1.6 Diagram of 3'-end cleavage and polyadenylation machinery and CIS regulatory elements in pre-mRNA

Diagram of the CPA machinery with its complexes distinguished by colour code as described in section 1.3. The brackets indicate the name of each complex and the specific name of their subunits. At the bottom a pre-mRNA is represented as a grey rectangle: UGUA is the CFI binding site, PAS is the Poly(A) site bound by CPSF and DSE is the downstream sequence element bound. The cleavage site is depicted as scissors and located between the PAS and the Downstream Sequence Element (DSE). Note that both CSTF and CFI are shown with their subunits as dimers.

1.4 mRNA alternative polyadenylation and biological roles

In the previous section I described the components that form part of the CPA machinery and their functions, as well as the main CIS regulatory elements within pre-mRNAs that dictate the site of cleavage and polyadenylation. As I showed in Figure 1.3 for CFI, pre-mRNAs can have more than one PAS with all its associated regulatory elements. Therefore, the CPA machinery will need to choose one of them to perform its function. As a result of this, mRNAs with different 3' ends will be generated. This process is known as "Alternative Polyadenylation" (APA), and it will be discussed in this section, together with its biological roles.

APA is a widespread phenomenon in eukaryotes, with nearly 70% of all human and mammalian genes displaying alternative 3'-ends isoforms (APA) (Derti *et al.* 2012; Hoque *et al.* 2013). In addition to this, approximately half of all genes in vertebrates such as the zebrafish and invertebrates such as *C. elegans* and *Drosophila*, also display alternative 3'-ends isoforms (Ulitsky *et al*. 2012; Smibert *et al.* 2012; Jan *et al.* 2011). Crucially, misregulation of 3'-formation and APA can lead to catastrophic consequences in the cell, and to diseases in humans such as cancer, thrombophilia and some thalassemias (Danckwardt *et al*. 2008; Erson-Bensan and Can 2016). Thus, the correct orchestration of this process in eukaryotes is vital.

As mentioned in the previous section, one of the main CIS regulatory elements required for cleavage and polyadenylation is the Poly(A) signal or PAS, whose canonical sequence is the hexamer "AAUAAA", discovered in 1976 (Proudfoot and Brownlee 1976). This signal can be located within the 3'UTR region of a pre-mRNA, generating isoforms with different 3'UTR length but coding for the same

protein. The signal can also be located in the coding sequence or introns of a pre-mRNA molecule, generating different protein isoforms (similarly to alternative splicing (Pan *et al.* 2008; Black 2003)). A representation of the two general types of APA is shown in figure 1.7.

In the former case, it has been shown that 3' UTRs act as binding platforms for trans-acting factors such as miRNAs and RBPs. In this way, a longer 3'UTR is able to undergo more regulation by such factors. For example, this was shown to be the case for the *Hox* genes in 12 *Drosophila* species, where the long and the short 3'UTRs of these key developmental genes contained very different miRNA target sites. Furthermore, although the level of sequence conservation in 3' UTRs among the 12 species studied was low, the authors show similar RNA topology by *in silico* RNA folding simulations, indicating that the structure of 3' UTRs was under high selective pressure (Patraquim *et al*. 2011).

In the latter case, that of coding-sequence APA or APA within introns and exons, Poly(A) sites located upstream of the 3'UTR can change the protein coded in the affected genes, diversifying the transcriptome and proteome. For example, the gene Cyclin D1 is key for the cell cycle by regulating the progression from G1 to S phase. As a result of polymorphisms in human populations, one of the introns of Cyclin D1 can contain a PAS, and when this element is used for 3'-UTR termination it generates a shorter protein, which is constitutively nuclear and expressed at high levels. This short protein isoform has been related to several human cancers, including breast and prostate cancer (Knudsen *et al.* 2006; Burd *et al.* 2006; Wang *et al.* 2008). The selection of either short or long 3' UTRs by APA can have a repertoire of functions in biology. For example, two cold-induced

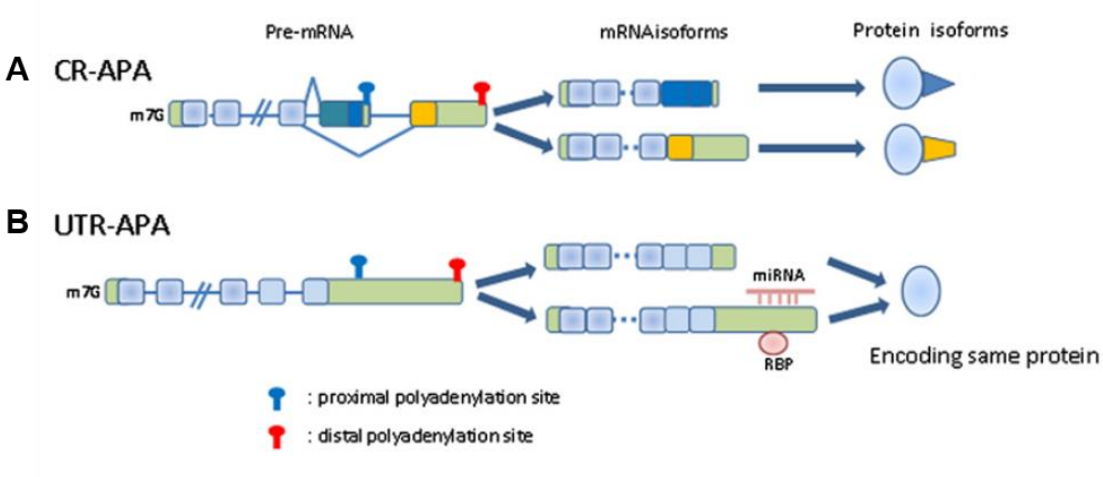


Figure 1.7 Diagram of the two main categories of APA

(A-B) Diagrams representing Coding Region APA **(A)** and 3' UTR APA **(B)**. Proximal and distal poly(A) sites are indicated by blue and red pins, respectively. Light green boxes represent UTRs. Light blue boxes represent shared coding regions. Dark blue and yellow boxes represent unshared coding regions and lines represent introns. Diagram taken and modified from (Di Giammartino *et al*. 2011).

RBPs, Cirbp and Rbm3, were shown to be important regulators for temperature-entrained circadian gene expression cycles through APA. More specifically, the depletion of these proteins triggered the use of proximal PAS and shorter 3' UTRs, whereas low temperature, which triggers and upregulation of both Cirbp and Rbm3, triggered the use of distal PAS and longer 3' UTRs. Furthermore, the authors found that the use of either proximal or distal PAS in several genes regulated by these factors showed strong circadian oscillations. Thus, this study reveals an interesting connection between cyclic environmental cues and the control of gene expression through APA in mice (Liu *et al.* 2013).

At the molecular level, the selection of either short or long 3' UTRs can have an impact on properties such as protein amounts and RNA localization. Regarding the former case, it has been shown that short 3' UTRs tend to produce higher amount of proteins, partly because of having fewer miRNA target sites (Ransom *et al*. 2008; Mayr and Bartel 2009). Regarding the latter case, it has been shown that 3' UTRs can regulate RNA localization. For example, the brain-derived neurotrophic factor (BDNF) expresses the short 3' UTR in the somata of neurons, whereas the long 3' UTR is also localized in dendrites (An *et al.* 2008). Interestingly, 3' UTRs can also regulate protein localization independently of RNA localization. For instance, the long 3'UTR of the human CD47 gene enables expression of this protein in the cell surface, whereas the short 3'UTR RNA isoform enables protein localization to the endoplasmic reticulum. Interestingly, the mechanism underlying this decision was shown to be post-translational, given that the long 3'UTRs binds to proteins such as HuR and SET, that interact with the site of translation for this isoform only (Berkovits and Mayr 2015).

In summary, I have shown that the selection of alternative 3'-end sites by APA within pre-mRNAs can have a variety of biological roles, and that this process, when misregulated, leads to catastrophic effects related with relevant aspects of human health. From an evolutionarily point of view, it is interesting to note that there is a clear correlation between 3' UTR length and morphological complexity, defined in this context as the number of cell types present in each organism. A study in 2012 addressed this fascinating question by using mature mRNA sequences and 3' UTRs from 15 different organisms, going from yeast, as to have a reference from a unicellular eukaryote, through tunicates, nematodes, insects, frogs, fish, birds, dogs, cattle, rodents, chimpanzees and humans. The authors found that the median 3' UTR length increased as the number of cell types did in an exponential fashion (Chen *et al*. 2012). As expected, yeast had the shortest median 3' UTR length with nearly 100 nucleotides, while humans and chimpanzees were in the top with median 3' UTR lengths of approximately 800 nucleotides. Moreover, this increase in cellular diversity was also correlated with an accumulation of miRNA genes and targets. The authors suggest that an expansion of post-transcriptional regulatory circuits can contribute to the emergence of new cell types during animal evolution, thus placing APA as a relevant actor for the evolution of life on earth.

Intriguingly, the expression of long 3' UTR isoforms in metazoans can show biases in different tissues. As will be discussed in this chapter, the nervous system has emerged as an important actor because it can express these long 3' UTRs in an exclusive way when compared with other tissues during development (Thomsen *et al.* 2010; Hilgers *et al.* 2011; Smibert *et al.* 2012), relating the

molecular complexity of post-transcriptional regulation with the inherent complexity of the brain and nervous tissue.

1.5 Models and evidence on alternative polyadenylation control

As discussed in the previous sections, APA is a widespread phenomenon with a repertoire of biological roles. Still, how APA is controlled remains only partly understood. One model that can explain why some tissues show radically different profiles of APA in comparison with others (to be further discussed ahead) proposes that the presence of specific factors other than core the CPA factors (such as the ones described in section 1.3) can force transcripts to bypass proximal PASs in the cells where they are expressed. Evidence for this model was provided in 2014 by our laboratory. In this study, the authors showed that the RNA-binding protein ELAV, which is a neural-specific RBP, controlled *Ubx* RNA-processing. More specifically, removal of ELAV leads to a shortening in the 3'UTR of the Hox gene *Ubx*, while the ectopic expression of ELAV during germ band extension stage (during which the short isoform of *Ubx* is expressed) leads to an increase in the long isoform (Rogulja-Ortmann *et al*. 2014). These interesting experiments showed that ELAV is sufficient to control Hox RNA-processing *in vivo*.

Yet, examples of more neural-specific factors that can explain the drastically different profiles of APA observed in the nervous system are scarce. Also, other mechanisms can be involved in a non-exclusive way. One such mechanism that can explain how different PAS are selected involves the same core CPA factors described previously. This time, it is not their presence or absence which acts a switch on 3' UTR length, but their abundance. Evidence for this model was

presented back in 1996 by the group of James Manley and colleagues concerning CSTF64 (See figure 1.2 and 1.6 for its structure and function). In this study, the authors showed that during differentiation of mouse B-lymphocytes there is a switch in immunoglobulin M (IgM) from a membrane-bound (μm) form to a secreted (μs) form. Interestingly, this switch is caused by effects in IgM mRNA processing instead of post-translational modifications, through coding-region APA (See figure 1.7). For the secreted form to be expressed, an upstream μs PAS is selected, which excludes the last two exons of the IgM pre-mRNA.

On the other hand, for the membrane-bound form to be expressed, a more distal μm PAS is selected, which includes these exons. The authors show that this switch is triggered by CSTF64 amounts, where high levels of CSTF64 lead to the use of the μs site (secreted form), while low levels of CSTF lead to the use of the μm form (membrane-bound) (Takagaki *et al.* 1996). Moreover, the authors show *in vitro* that CSTF64 shows higher affinity for the μm PAS and that this PAS is stronger than the μs one. Thus, CSTF64 expression is repressed in mouse primary B-cells, triggering the use of the μm PAS and keeping IgM from being secreted.

This study is interesting for two main reasons, the first one is that it was the first study providing experimental evidence for the “CPA abundance” model of APA control, the second one is that the idea of “strong” and “weak” poly(A) sites was used in the context of this model. Although the concepts of “strong” and “weak” PAS is somehow ambiguous given the complexity of RNA sequences and structure in mammals. Progress in clarifying its meaning was being made while the CPA factors were being discovered in the late 1980s. For example, in 1989, a study showed that when compared with the canonical PAS hexamer

“AAUAAA”, described by Nick Proudfoot in 1976 (Proudfoot and Brownlee 1976), the variant “AUUAAA” had approximately 80% the processing efficiency *in vitro* from its canonical counterpart, by experiments with viral SV40 RNA (Wilusz *et al*. 1989). During the next year, a systematic dissection of the “AAUAAA” hexamer by mutating the nucleotides in each one of the positions showed that all 18 changes significantly reduced the efficiency of cleavage and polyadenylation, with the exception of “AGUAAA”, which showed an efficiency close to 30% when compared with its canonical counterpart, also by experiments with viral SV40 RNA (Sheets *et al*. 1990).

During the years it was also shown that apart from the PAS signal itself, the previously discussed motifs upstream and downstream of the PAS that are bound by CFI and CSTF, respectively, also contribute to the strength of a PAS by either using canonical or less efficient, related sequences (Zhao *et al*. 1999; Bagga *et al*. 1995). Interestingly, it has been shown that distal PASs tend to be “stronger” than their proximal counterparts within the same mRNA in humans (Legendre and Gautheret 2003), as also in the case of CSTF64 binding to IgM pre-mRNAs in mouse lymphocytes (Takagaki *et al.* 1996). Most likely, the reason for this is that a distal PAS will have to outcompete the proximal one to be used, given that RNA PolII will first confront the proximal PAS during transcription.

Another factor that will affect the apparent usage of either proximal or distal PAS is the stability of each one of these transcripts. This can be an important consideration, given that most techniques are based on measuring steady-state mRNA levels (Moore 2012).

More recent experiments using siRNA knockdown of CPA factors in human cells have also shown that depletion of core CPA factors can affect APA, supporting

the notion that CPA factor levels are an important cue to dictate PAS usage. As we will discuss in subsequent chapters in more detail, CFI has emerged as a key factor controlling 3'UTR length by experiments of this kind (Kubo *et al.* 2006; Masamha *et al.* 2014). Other factors have also been tested *in vitro* by knockdown experiments and shown to control APA in different general modes. For example a study using siRNA knockdowns in C2C12 cells proposed a subset of principles for PAS site selection. For instance, Pcf11 and Fip1 enhanced the use of proximal PAS, while CFI25, CFI68 and PABPN1 enhanced the use of distal PAS (Li *et al*. 2015).

Thus although pieces of evidence supporting different proposed models as to explain how APA is controlled have been developed since the 1990s, a unifying model that can accurately explain these decisions is missing. To achieve this, future research will need to be focused in both the CIS regulatory elements and the dynamics of CPA factor expression in different cell types, and to consider more deeply the molecular mechanisms of APA in metazoans, given that most of our current knowledge on how APA is controlled comes from experiments *in vitro*. Thus, one of the main aim of this study is to advance the understanding on how APA is controlled during the formation of the central nervous system by using *Drosophila melanogaster*, as we will discuss in the next section.

1.6 Tissue-specific patterns of alternative polyadenylation in the nervous system

In 2010, a study conducted at the Alonso Lab in the University of Sussex showed for the first time that during *Drosophila* embryogenesis, the *Hox* gene *Ultrabithorax* (*Ubx*) showed a peculiar characteristic at the level of mRNA

expression: during early embryogenesis, short 3'UTR isoforms were expressed in the epidermis, while during late embryogenesis, long 3' UTR isoforms were observed to be restricted to the nervous system. Moreover, these extensions displayed more binding sites for miRNAs *iab4/8* (Thomsen *et al.* 2010), which is a miRNA locus located in the bithorax complex (BX-C) that generates miRNAs from both DNA strands and regulate expression of these posterior *Hox* genes (Tyler *et al*. 2008; Bender 2008).

As mentioned before, the *Hox* genes are key developmental genes that pattern the anteroposterior axis of all bilateral animals and also play a role in the formation of the nervous system (Miller *et al*. 2001; Mallo and Alonso 2013; Rogulja-Ortmann and Technau 2008). This key observation is shown in Figure 1.8. Instead of being an unusual characteristic of *Ubx* mRNA processing, the authors showed that all other Hox genes with alternative 3' UTR isoforms: *Antennapedia* (*Antp*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*) also displayed the same trend (Thomsen *et al.* 2010), as shown in Figure 1.9. Interestingly, as we mentioned previously from a study one year later also from the Alonso Lab (Patraquim *et al*. 2011), the distal 3' UTRs of these genes showed an expansion in binding sites for miRNA *iab*4/8, which were evolutionarily conserved among different *Drosophila* species. The authors, therefore, proposed that APA works as a "Context-dependent" mechanism that is able to modulate visibility to miRNAs.

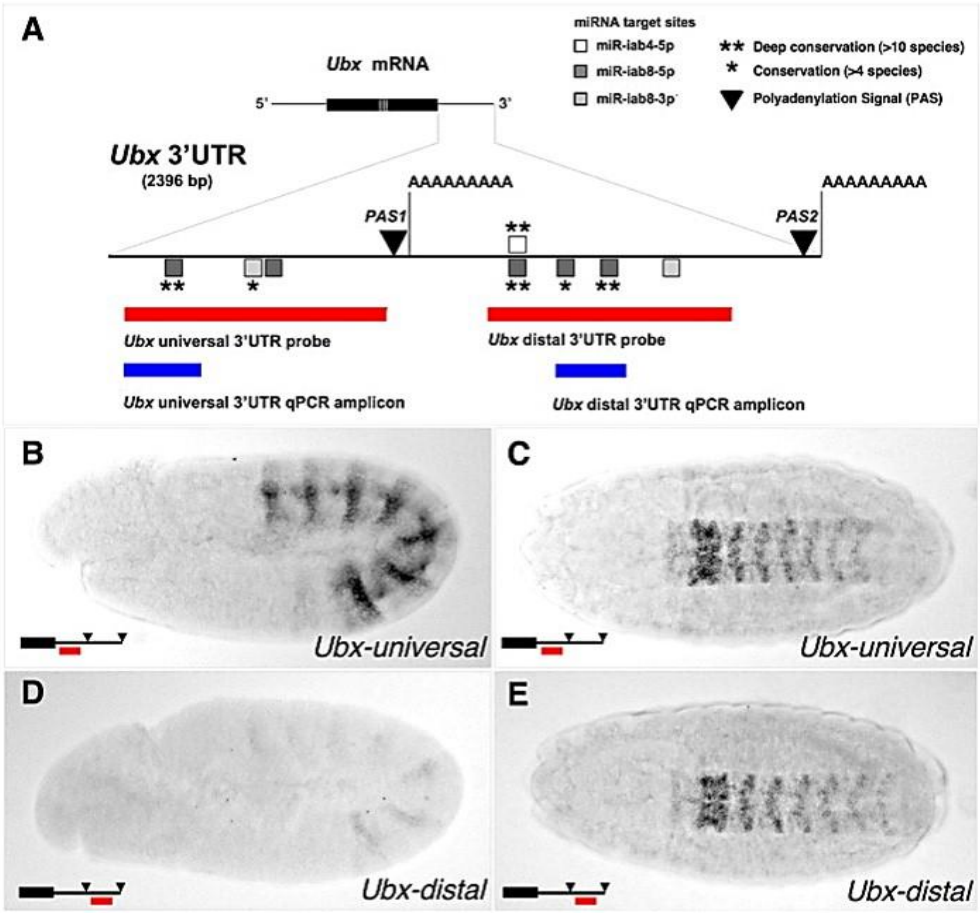


Figure 1.8 The *Hox* gene *Ubx* produces different 3' UTR isoforms with the distal 3' UTR expressed in the nervous system

(Legend on the following page)

Figure 1.8 The *Hox* gene *Ubx* produces different 3' UTR isoforms during embryogenesis, with the distal 3' UTR expressed in the nervous system

(A) Diagram of the 3' UTR of *Ubx* indicating the two different PAS by inverted triangles and miRNA *iab4/8* target sites by boxes, considering each strand in detail (5p and 3p). Note that miR-iab4-5p only targets the distal 3' UTR and that target sites in the distal 3' UTR show more evolutionary conservation than the ones in the universal 3' UTR. The level of conservation of these miRNA target sites is indicated by stars **(B and C)** mRNA localization of the Universal 3' UTR of *Ubx* in stage 11 **(B)** and stage 16 **(C)** embryos. Note that the chromogenic signal is strong in both the epidermis in stage 11 embryos **(B)** as well as in the nervous system in stage 16 embryos (ventral view) **(C)**. **(D and E)** mRNA localization of the distal 3' UTR of *Ubx* in stage 11 **(D)** and stage 16 **I** embryos. Note that the chromogenic signal is very weak in the epidermis of stage 11 embryos **(D)**, while is strong in the nervous system (ventral view) of stage 16 embryos **I**. Figure taken from (Thomsen *et al.* 2010).

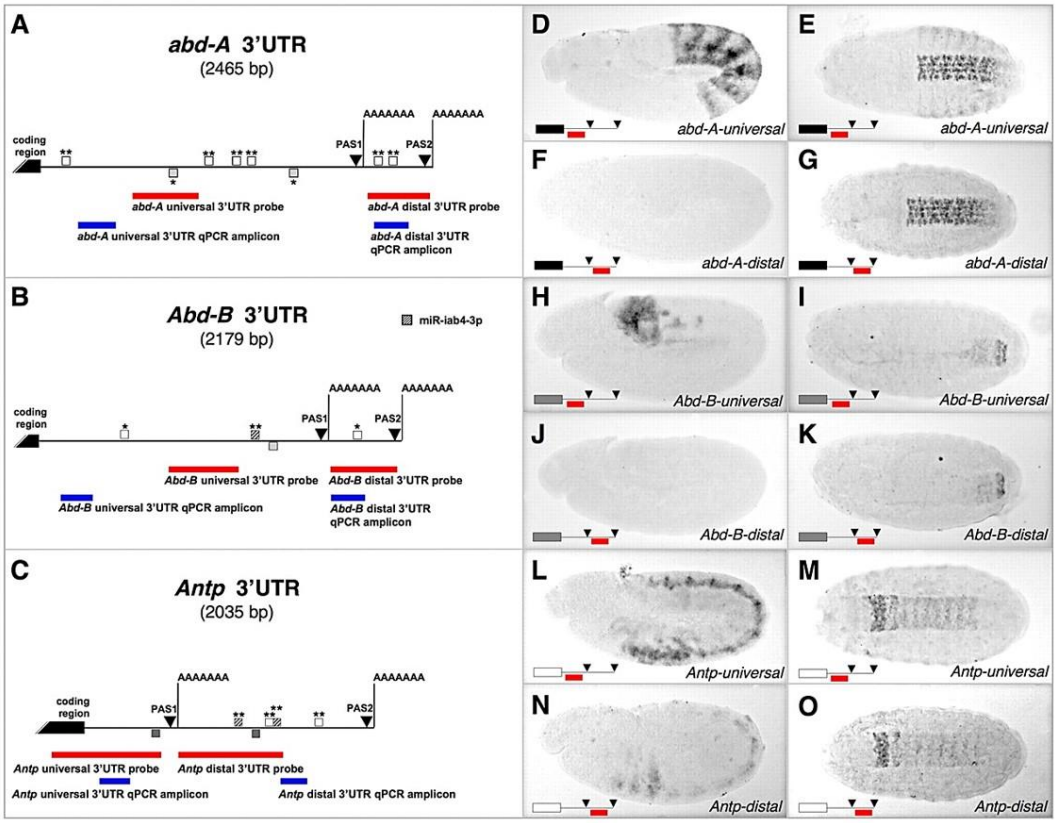


Figure 1.9 *abd-A*, *Abd-B* and *Antp* also express the distal 3' UTR in the nervous system

(Legend on the following page)

Figure 1.9 *abd-A*, *Abd-B* and *Antp* also express the distal 3' UTR in the nervous system

(A-C) Diagrams of the 3' UTR of *abd-A* **(A)**, *Abd-B* **(B)** and *Antp* **(C)** as shown in figure 1.8 panel A. The location of the RNA *in situ* hybridization probes used in this study to detect these isoforms, as well as primers for qPCR are shown (See Thomsen *et al.* 2010). **(D and E)** mRNA localization of the Universal 3' UTR of *abd-A* in stage 11 **(D)** and stage 16 **I** embryos. Note that the chromogenic signal is strong in both the epidermis in stage 11 embryos **(D)** as well as in the nervous system in stage 16 embryos (ventral view) **I. (F and G)** mRNA localization of the distal 3' UTR of *abd-A* in stage 11 **(F)** and stage 16 **(G)** embryos. Note that the chromogenic signal is absent in the epidermis of stage 11 embryos **(F)**, while is strong in the nervous system (ventral view) of stage 16 embryos **(G). (H-K)** Analysis for *Abd-B* as done for *abd-A* **(D-G)**, note that the same trend is observed. **(L-O)** Analysis for *Antp* as done for *abd-A* **(D-G)**, note that the same trend is also observed, although this time there is more expression of the distal 3'UTR in the epidermis of stage 11 embryos **(N).** Figure taken from (Thomsen *et al.* 2010).

During the next year (2011), another group led by Valérie Hilgers and colleagues showed that this property of “Neural-extended” 3-UTRs was not exclusive to the *Hox* genes and they identified other 30 genes that displayed the same trend (Hilgers *et al.* 2011). Remarkably, this time the extensions in some genes were shown to be more than 10 times longer than the ones observed In the *Hox* genes. As we will discuss in chapter 5, a comparison of the 3’ UTR length reported in this study versus the current databases shows that some extensions can be even longer than annotated in the study (Hilgers *et al.* 2011).

During 2012, another group led by Eric Lai showed that this property of “Neural-extended” 3’ UTRs was even broader, concerning almost 400 genes and life stages other than embryogenesis, such as larvae, pupae and adults. Interestingly, while the nervous systems showed a bias towards 3’ UTR lengthening, the testis showed a bias towards 3’ UTR shortening (Smibert *et al.* 2012).

During the next year in 2013, it was shown that this well-established observation in *Drosophila* was far from being an exception of insects and that the complex mammalian brain also showed extensive 3’ UTR lengthening, with 2035 genes in the mouse and 1847 genes in humans using substantially distal novel 3’ UTRs according to RNA-seq data (Miura *et al.* 2013). In line with what was observed in the Hilgers study, Northern Blot analysis of selected genes showed transcripts displaying exceptionally long 3’ UTRs, with lengths of more than 10 kb and some of more than 18 kb. The authors report thousands of conserved miRNA sites present in these extensions which are strongly enriched for well-studied neural miRNAs (Miura *et al.* 2013). Yet, it is difficult to ascribe miRNA binding sites and potentially RBP binding sites as the only or main biological reasons for such

extensions. Bioinformatics analyses have already shown that only modest extensions of 1 kb or less are enough for differential miRNA targeting (Patraquim *et al*. 2011). Thus, the meaning of these extremely long 3' UTRs in neural tissues remains unclear and will be discussed in subsequent chapters.

1.7 Aims and outcomes of this thesis

As discussed in the previous sections, the phenomenon of APA and its bias for producing long 3' UTRs in the nervous system of both vertebrates and invertebrates is a recent field of study that has grown rapidly since 2010. Despite this, the complete molecular identity and characteristics of CPA factors, let alone their biological functions beyond 3'-end processing in metazoans, together with the mechanisms by which APA is controlled, remain not fully understood. Thus, as it was the case for the biological meaning of RNA poly(A) polymers within living cells in the early 1960s, the biological meaning of extremely long 3' UTRs in the nervous system is similarly enigmatic nowadays.

This dissertation aims at addressing the following questions. In chapter 3, I ask about evidence to test the model of "CPA factor abundance" for PAS selection in the context of neural development. To achieve this, I first interrogate the identity of core CPA factors in *Drosophila* and compare them with their known human and yeast counterparts for protein similarity and redundancy. I then address their expression levels during *Drosophila* embryogenesis by using databases to look for patterns that can suggest changes in CPA factor abundance as development progresses. Then, I analyse the expression patterns of *Drosophila* CPA factors in life stages other than embryos in different tissues by using databases and I explore differences within the nervous system in terms of tissue-specific CPA

factor expression levels. I then analyse the expression pattern of the members of CFI, the most conserved CPA complex between humans and *Drosophila* and the one with the highest expression levels in the nervous system, and I show that in late stages both CFI25 and CFI68 are highly expressed and restricted to the developing nervous system, correlated with the observed trend of extended 3' UTRs reported 6 years earlier.

In Chapter 4, I ask about the roles that the *Drosophila* orthologue of CFI25 has for nervous system function. I show that a mutation in this gene affects feeding behaviour in the *Drosophila* larvae and is required for developmental transitions between larval stages. I address the reasons of the observed phenotype by identifying candidate target genes that can connect larval feeding with the molecular function of CFI25. I also show that a subset of the identified genes are affected in APA patterns, as well as mRNA expression levels. And I discuss the potential molecular mechanisms that can explain the observed phenotype.

In Chapter 5, I address the question of the molecular mechanisms to achieve 3' UTR extensions within the developing nervous system. I show that the reported 3' UTR lengths of genes that were described in 2011 by Valerie Hilgers are not accurate when compared with the current databases and that the extensions of 3' UTRs can go below or above than published measurements, resulting in 3' UTR lengths comparable as those observed in human tissues. I then address the question on the roles of CFI factor abundance within the developing nervous system in relation with these extensions. Finally, I show that neural-specific knockdown of CFI factors can affect APA in a subset of the genes reported in 2011 for 3' UTR extensions, as well as in a subset of the *Hox* genes.

The reasons for differential sensitivity to CFI depletion is explored by considering both 3' UTR length and composition. Consequently, a bioinformatic approach is used to scan for enriched motifs in the 3' UTRs of these neural-extended genes to look for a molecular mechanism that can explain the different sensitivity shown by these genes to depletion in CFI factors.

Altogether, this work shows that the *Drosophila* CPA machinery is as complex as its human counterpart and that neural 3' UTR extensions are achieved by CFI factor abundance, making *Drosophila* an excellent system to address the molecular mechanisms that control APA during the formation of the nervous system. Also, this work shows that the members of CFI are key for nervous system function, because a mutation in one of these factors (CFI25) can affect feeding and impair larval developmental transitions, unravelling unprecedented roles for core CPA factors when considered at the organismal level, and suggesting other potential biological roles of the CPA machinery that remain to be discovered.

Chapter 2

Materials and Methods

2.1 Stocks and fly husbandry

Fruit flies (*Drosophila melanogaster*) were cultured using molasses food following standard procedures at 25°C on a 12 hour light and dark cycle. Oregon R was used as a wild type strain. All *Drosophila* stocks used in this study are depicted in table 2.1

Table 2.1 *Drosophila* stocks

ID	Genotype	Origin
Oregon R	Wild Type	Host laboratory
w^{1118}	w^{1118}	Host laboratory
TM3, Dfd-GMR-nvYFP	w*; ry^{506} Dr^{1}/TM3, P{Dfd-GMR-nvYFP}3, Sb1	Bloomington Stock Center #23231
CG7185 GFP Tag	y^{1} w^{*}; CG7185[38575]::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG	Vienna Drosophila Resource Center #318105
$CFI25^{G19200}$	y^{1} w^{*}; P{EP}$CG3689^{G19200}$ / TM3, Sb^{1} Ser^{1}	Bloomington Stock Center #28422
$CFI68^{CC00645}$	w* ; P{w[+mC]=PTT-GC}CG7185[CC00645]	Bloomington Stock Center #51539
UAS – CFI25 RNAi	y^{*} w^{1118} ; UAS–CG3689 RNAi	Vienna Drosophila Resource Center #105499/KK
UAS – CFI68 RNAi	y^{1} sc^{*} v^{1} ; P{TriP.HMS00113}attP2	Bloomington Stock Center # 34804
Elav>Gal4	P{GAL4-elav.L}2 / CyO	Bloomington Stock Center #8765
$CFI25^{Revertant}$	Excision of P{EP} from stock #28422	This study
$CG4022^{GS12916}$	y^{1} w^{67c23} ; P{GSV6}GS12916 / TM3, Sb^{1} Ser^{1}	Kyoto Stock Center #204369
$CG4022^{c05627}$	w^{1118} ; Pbac{PB}$CG4022^{c05627}$	Bloomington Stock Center #17716
Δ2-3 transposase	w^{*} ; P{Δ2-3} e^{1}/ Tm6	Donated by Juan Pablo Couso Lab

2.2 Embryo collection and fixation

Flies were kept in small collection cages at 25°C with apple juice agar plates supplemented with yeast paste for embryo laying. Collected embryos were dechorionated in 50% bleach for about 3 minutes and then fixed for 20 minutes in 3.5 mL heptane / (1.813 mL 10% ultrapure formaldehyde and 1.687 mL 1XPBS) at room temperature. The fixative was removed and the embryos were devitellinized in methanol with vigorous shaking for at least 2 minutes. Embryos were rinsed three times with 100% methanol, then rinsed with 100% ethanol and stored in 100% ethanol at -20°C for later use.

2.3 Generation of CFI25 mutant revertants

The *Drosophila* CFI25 mutant stock y^1 w^* ; P{EP}CG3689^{G19200} / TM3, Sb^1 Ser^1 was isogenized by backcrossing it into a standard w^{1118} genetic background with a balancer third chromosome (TM3, P{Dfd-GMR-nvYFP}3, Sb1). Then, isogenic P{EP}CG3689^{G19200} mutants, in which feeding was tested beforehand, were crossed with the stock w* ; P{Δ2-3} e1/ Tm6. Subsequently, P{EP}CG3689^{G19200}/ P{Δ2-3} males were crossed to the w*; ry^{506} Dr^1/TM3, P{Dfd-GMR-nvYFP}3, Sb1 balancer stock to establish new revertant lines from white-eyed males. 32 independent revertant stocks were generated in which confirmation of the excision of the P{EP} element was done by PCR and sequencing, using Oregon R wild types as a negative control and the original CFI25 mutants as a positive control. CG3689$^{Revertant\ 4}$, in which the excision does not damage CFI25 and the whole P{EP} element was excised, was used as the control for feeding for CFI25 mutants.

2.4 Antibody staining

Antibody stainings were performed using a standard protocol as follows. Fixed embryos were rehydrated in 50% PBTx / Ethanol (1XPBS 0.3% Triton X-100) and then washed several times in PBTx. Primary antibodies (diluted in PBTx) were incubated overnight at 4°C. Secondary antibodies were incubated for two hours at room temperature and washed with PBTx. After secondary antibody incubation, embryos were incubated with DAPI for 15 minutes to label the nuclei, washed with PBTx and mounted in Vectashield (Vector Laboratories). All antibodies and respective concentrations (v/v) used in this study are listed below (Table 2.2 and 2.3). Fluorescent imaging was carried out using a Leica DM600 fluorescent microscope and a Leica SP8 confocal microscope. All images were processed and analysed by ImageJ.

Table 2.2 Primary antibodies used in this study

Name	Host	Concentration	Origin
Anti-elav	Mouse	1:100	Developmental studies hybridoma bank
Anti- futsch	Mouse	1:100	Developmental studies hybridoma bank
Anti – NUDT21 (CFI25)	Rabbit	1:1000	Abcam
Anti-GFP	Rabbit	1:500	Life Technologies
Anti-tubulin	Mouse	1:500	Developmental studies hybridoma bank

Table 2.3 Secondary antibodies used in this study

Name	Concentration	Origin
Anti-mouse Alexa 488	1:750	Life Technologies

Anti-rabbit Alexa 488	1:750	Life Technologies
Anti-mouse HRP	1:3000	Cell Signalling Technology
Anti-rabbit HRP	1:3000	Dako

2.5 RNA probes

Templates of RNA probes for RNA *in situ* hybridisation were obtained from PCR amplified genomic fragments or cDNA (see table 2.4) and cloned into pGEM-T easy (Promega) according to the manufacturer's instructions. Plasmids were linearized with a unique restriction site, purified by QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions and their concentration was measured using the Nanodrop 2000 spectrophotometer (Thermo scientific). Sense and Antisense probes were synthesised using digoxigenin (DIG) RNA labelling mix (Roche) according to the manufacturer's instructions with either T7 or SP6 RNA polymerase (Roche), depending on the orientation of the insert. After *in vitro* transcription, the DNA template was removed with DNAse I (New England Biolabs) and RNA probes were precipitated with 2.5 µl 4M lithium chloride and 75 µl pre-chilled 70% ethanol solution at -80°C overnight. RNA probes were centrifuged at 4°C for 30 minutes and air-dried at room temperature. RNA pellets were re-suspended in 50 µl hybridisation buffer (50% formamide (ACROS), 5X SSC (Sigma), 100 µg/ml salmon sperm DNA (Invitrogen), 0.1% PBTween (1XPBS RNAse-free, 0.1% tween 20 (Sigma)), aliquoted and stored at -80°C.

Table 2.4 Primer sequences and RNA probes length

Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')	Probe length (bp)	Source
CFI25	CGTCCAGCCGGT TAATTT	GTTAGGTAGCGC TATCGTTG	955	This study

abd-A-universal	CCCACCATCAACCAACTTTC	TACTTGCGCAATTGTTTTGC	428	(Thomsen *et al*. 2010)
abd-A distal	GTTTTACTCCGCCTGGGAAG	AATCCCCTTGGCTGAAATCT	403	(Thomsen *et al*. 2010)

2.6 Fluorescent RNA in situ hybridisation (FISH)

In situ hybridisations were performed similarly to that described by (Beckervordersandforth *et al*. 2008) but with slight modifications. Fixed embryos were rehydrated in PBTween (1XPBS RNAse-free 0.1% Tween 20), pre-treated with 3% H_2O_2 in Ethanol for 20 minutes to quench endogenous HRPs and with sodium borohydride (0.001% in PBTween) for 10 minutes to reduce auto-fluorescence. Embryos were then pre-hybridised in hybridisation buffer for two hours at 55°C. 200-300 ng of DIG-labelled RNA probes in hybridisation buffer were denatured at 80°C for 5 minutes. RNA probes were then incubated with the pre-treated embryos at 55°C overnight. All steps were carried out in RNAse-free conditions.

For detection of the DIG-labeled RNA probes, embryos were blocked in TNB buffer (0.1 M Tris PH 7.5 (Fisher), 0.15 M NaCl (Fisher), 0.5% blocking reagent (Roche)) for 30 minutes and incubated with 1:500 anti-DIG-POD in TNB buffer for 2 hours at room temperature. The fluorescence signal was detected using the Cy3 TSA amplification kit (Perkin Elmer) according to manufacturer's instructions.

2.7 RNA extraction

RNA was extracted from staged embryos using TRI reagent (Sigma) following the manufacturer's instructions. 50-100 staged embryos were homogenized in 50 µl of TRI reagent using a sterile RNAse-free pestle in 1.5 ml Eppendorf tubes.

After homogenization, 450 µl of TRI reagent were added and the tubes were incubated for 5 minutes at room temperature to dissociate nucleoprotein complexes. The samples were centrifuged for 10 minutes at 4°C to precipitate the insoluble material and the high molecular weight DNA. After centrifugation, the supernatant was transferred to a new 1.5 ml Eppendorf tube. RNA was separated from DNA and proteins by adding 100 µl of RNAse-free chloroform (Sigma Aldrich), mixing and incubating for 15 minutes at room temperature. The different phases – aqueous phase (RNA), interphase and organic phase (DNA and proteins) – were separated by 15 minutes of centrifugation at 4°C and the aqueous phase was transferred to a new tube. The RNA was precipitated with 250 µl RNAse-free isopropanol (Sigma Aldrich) at -20°C for one hour, followed by centrifugation at maximum speed for 30 minutes at 4°C. The precipitated RNA was washed twice with RNAse-free 75% ethanol, resuspended in nuclease-free water and stored at -80°C until use. The concentration of RNA was measured by Nanodrop 2000 spectrophotometer (Thermo Scientific). All steps were carried out in RNAse-free conditions.

2.8 Reverse transcription (RT)

1 µg of total RNA was treated with DNAse I (New England Biolabs) at 37°C for 10 minutes to eliminate genomic DNA. Treated RNA samples were used for cDNA synthesis using oligo(dT) primers (Invitrogen) and MuLV Reverse Transcriptase (Invitrogen). The same amount of RNA was used when comparing different genotypes.

Total RNA was mixed with 3 µl of oligo(dT) and water to a final volume of 12 µl, denatured at 75°C for 3 minutes and placed on ice. Then, the remaining RT

components were added – 2 µl of 10X RT buffer (Invitrogen), 4 µl of 2.5 M dNTP mix (Invitrogen), 1 µl of RNAse inhibitor (Invitrogen) and 1 µl of MuLV Reverse Transcriptase (For the No-RT controls, 1 µl of nuclease-free water was used instead of MuLV Reverse Transcriptase) – and incubated at 44°C for one hour for cDNA synthesis. An additional incubation at 92°C for 10 minutes inactivated the Reverse Transcriptase. The cDNA samples were stored at -20°C until use.

2.9 Semi-quantitative RT-PCR

PCR reactions were prepared on ice to a final volume of 25 µl as follows: 2.5 µl of 10X PCR buffer (New England Biolabs), 0.5 µl of 10 mM dNTP mix (New England Biolabs), 1 µl of 10 µM forward/reverse primer (See table 2.5), 0.25 µl of standard *Taq* DNA polymerase (New England Biolabs), 1 µl of cDNA and 18.75 µl of nuclease-free water. PCRs were performed using an Eppendorf PCR machine with the following conditions:

1 cycle:

Extended DNA denaturation at 95°C for 5 minutes

25 to 35 cycles (depending on each pair of primers):

Denaturation at 95°C for 30 seconds

Primer annealing at 55°C for 30 seconds

Extension at 72°C for 30 seconds

Hold: 4°C

Each primer pair was optimised to ensure that the reaction was on the exponential phase of amplification.

Expression values were normalised using reference gene RpA1. At least three independent biological replicates were done for each experiment.

All experiments included two negative controls: (i) PCR with 1 µl of No-RT reaction to control for genomic DNA contamination from each RT reaction and (ii) PCR with nuclease-free water to control for PCR mix contamination. All experiments also included one positive control: Genomic DNA template to control for PCR reaction and primer binding.

Table 2.5 Semi-quantitative RT-PCR primers

Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')	Amplicon length (bp)	Source
Ubx Universal	GAAATGACGCG GAGACAGAT	AATCTGCGCTC CTTCCACTA	236	(Thomsen *et al*. 2010)
Ubx Distal	GAACGAAGGCA GATGCAAAT	GGTAAGTGGTC GGATGCAGT	225	(Thomsen *et al.* 2010)
abd-A Universal	CGGGTTTTATTG CTGTGGAT	CGTTGGCCCAG AGACTCTAC	193	(Thomsen *et al.* 2010)
abd-A Distal	CCTTTTCGATGA GGTCCAAA	CGGTTTCGGTC GGTCTAATA	219	(Thomsen *et al.* 2010)
Antp Universal	ACGGAGTCTAC CCACTTAAA	GATCTGAGGTC ACATGAGTTG	336	This study
Antp Distal	GAGGACGGAAT GGCAAACTA	GTCTTTTCACCT GGGATTGG	165	(Thomsen *et al.* 2010)
Abd-B Universal	CGTATTTCTCTC AACGCTCTC	CGGAGTGTGTC TTCTTGTTT	300	This study
Abd-B Distal	TCCGTACAACAC CATTTTCG	AGTGGCGATTA CGAGCTGAT	229	(Thomsen *et al.* 2010)
Elav Universal	AGTAGCAGGCA GGAGAAA	GACTGTGCCAA CCTTTGA	303	This study
Elav Distal	GACGAACTGCT CCGATTT	CGCTCTTCTCC GATTACTTAC	284	This study
ADAR Universal	TGTATATGCTAA GTTCAGTTTACG	GCTTAAAGTGCT TGTTTATAATGT G	223	This study

ADAR Distal	CCCGCTAAACC AGTGATAAG	GCGTTTAGCCC AGAATGT	319	This study
Nrg Universal	GAACAACAAGC AACACAACA	GAGCGGGACAA AGATATACAG	320	This study
Nrg Distal	CGAATCGGTTC GGCTTTAT	GAGGCTGGGTA TTGGTTATTC	300	This study
Pum Universal	GCATACACCCA CACAATGA	TTGGCTTACTTG GCTAACAG	318	This study
Pum Distal	GCAGGGCTCGG TATTATTT	TTCGCTGGCTTA CACTAAC	330	This study
Imp Universal	AGCACCCACCA CAATTTAC	GCGCGCTGCTT TCTATTA	320	This study
Imp Distal	GGAACGAAACG AAACGAAAC	GCTCAGTCTCC AGTTGATTAC	282	This study
Ago1 Universal	CCACTTCCTTCC CTCAAATC	CAAACTTGTGCC TGACATTC	281	This study
Ago1 Distal	ATGCGAGTTTGT GAAATATGC	GGGTACATTTC GTGGGTTTA	244	This study
Brat Universal	CGGTCTCTCCA GCTCTAAT	CGAGGGTTTGA AGTGAGAAG	341	This study
Brat Distal	CTTGAGGATGT GTGTGCATAG	CCGTGTGGCTT TCGTATTT	294	This study
Wdb Universal	CCAAGATCAGTA AGAGCGTAAG	ATTAACGCGGA CACACAC	305	This study
Wdb Distal	CTAAGCGACGT GTGTGTAAG	CAAACAGGTCG AGTCGATAAG	307	This study
Nmo Universal	AGAACATGGAG GAGAGGAG	TACCGCTGCTG CTTTAAC	298	This study
Nmo Distal	AAACACTCGATA CGCTAACC	CTTTGTTGCGTG CCTTTAC	290	This study
Fne Universal	AGATGAGCCAG ACGACAA	GAGTTATGCTG GTAGTTCCTAAA	296	This study
Fne Distal	GCCCAGCAGCT AATGAAA	GGGTGTGTAAG TGTGAACTG	302	This study
Nej Universal	CAGCTACAATG GTTGGTAGG	GTTGGTCTTCGT CGTCATC	103	This study
Nej Distal	CATAGGGATCG GGATTAGGA	GCGTCGTTGTT GTGTTTG	311	This study
Gβ13F Universal	GAAACAGAAAC AGCAGCATAAG	GTTGTTGTGGTC TACGTTCTA	304	This study

Gβ13F Distal	GGCCAGTCAGTCAGTTAATC	GGTTTCCTCCATCTTCATCTT	303	This study
Shep Universal	ACCCAGCATCCAGAATCTA	CTCACTTGCCGCTGTTT	299	This study
Shep Distal	ACCCACACCAAATAGTTTCC	GCGTTCATTCCTCCTCTATG	306	This study
Step Universal	CAGCTCGGCGAATCTTT	TGATGGCTTGTTTCGAGTC	227	This study
Step Distal	CGGCTTACTGACGTCTAATC	TTACCGCCGTCCTTTATATTC	311	This study
Hrb27C Universal	CAGCACTCTCACCCATTTAG	TGGTATTTCGCGCTCTATTC	286	This study
Hrb27C Distal	GATGCGCCAAATGCAAAC	ACTCTCGTTCTGAGGGATTAG	317	This study
MeiP26 Universal	CAAAGCGCGCAACAATC	GCTGTCGATGAGGCAAAT	298	This study
MeiP26 Distal	CGCAAACGGCAGACTATT	CGAGGGATTGATGGACTATTG	303	This study
Kurtz Universal	TACCAATAGCCATGGTAAACAG	CAGGTGTGGCCGAAATAC	196	This study
Kurtz Distal	CCATTTCCGTCTGTCTGTATAA	ACCACATTTCGCCATTACA	202	This study
RanBPM Universal	GTTCAGACCAGCAAACGA	CCGATGAAGATGAGCTGAC	217	This study
RanBPM Distal	CACATTCGGCAAACATGAATAG	ACACATTGTCAGAATGGCATA	138	This study
S6k Universal	CGCCAATCGAAACAGACA	CGTTGCAGTTGTCCCTAAA	208	This study
S6k Distal	CCGTTCAGGGCTCATATTG	CCTGTGTGGCTTTCTGTT	223	This study
For Universal	GAGAATCAGAACCCGTTTCCT	TTCGATGCGAGCTGCTG	150	This study
For Distal	GGTCTGTGACTCTGTTTCAG	CAGTAGAGAGGCCACATAGA	215	This study
Lov Universal	GACCAGGAAATGAACCCTAC	TGCTTACGGGACAAGGA	196	This study
Lov Distal	GGAAGCAAGTTGAGGGAAA	GGAACGTGCACAACTATGA	198	This study
Itp-r83A Universal	TCAATGCGGGATGAACAAT	TTTAACAGCCCTAAGTTCTCTG	201	This study

Itp-r83A Distal	GAAGCGAGTTG TGTTGGTTA	CAAAGATATATT CTCCACTTTGCA C	124	This study
Klu Universal	CCAGCTTAGTG CTACAGAAA	TGGAATTGCTAG CTGTATCG	313	This study
Klu Distal	AGAAACGAACG CTACACAAA	CACGAAAGGTG AGGTGATTC	320	This study
Shi Universal	GGGCAGATGCT TAGTGACGA	CGATAAGCGAA AGCAACGCC	338	This study
Shi Distal	TTCGAATCGCA GTGCAGGAG	GCGGACATTGC GTTGCTAAA	249	This study
NPFR Universal	AGAGCTCGAAG CCTGTAA	TCCTAGGAACT GTTGAGAGAA	195	This study
NPFR Distal	TGAGGTCTGGT CTCGTGTCT	CAGCCAGAGTG TTTCCCGAT	196	This study
RanBPM	GGCCATCGAAC ATACACTAC	TGTGCTTGAAC GTCTTGG	261	This study
CFI25	CCCTTACGAACT ACACATTCG	GTGTCCTCAACA ATCCACTC	383	This study
CFI68	CACTGGTCACA GCCATTT	GCGTTCACGTT CTCTACTG	351	This study
RpA1	AAGAGCATCGA CGACCTGAT	GCCACATTCAAC CGCTTATT	213	Host Laboratory

2.10 Agarose gel electrophoresis

RNA and PCR products were visualized by agarose gel electrophoresis. All agarose gels were made at 1% (w/v) by dissolving nuclease-free agarose (Fisher) in 1X SB buffer (2.25 g Boric Acid (Fisher) and 0.4 g Sodium hydroxide (Sigma Aldrich) in 1 L of double distilled water). The mixture was heated in a microwave until it was completely homogenised. After cooling, 250 µl of 0.1 mg/ml ethidium bromide was added to the liquid agarose solution before pouring into the gel cast system. Samples were prepared in 1X loading buffer (Thermo Scientific), loaded into the wells of the gels alongside a 100 bp DNA ladder (New

England Biolabs) and subjected to electrophoresis in 1X SB buffer for 30 minutes at 150 V.

Gel pictures were taken using a Uvidoc gel documentation system (Uvitec Cambridge) and UviPhotoMW image analysis software. Quantification of the gels was done with ImageJ.

2.11 Western Blotting

20 – 50 embryos were homogenised in 20 µl of 2X Laemmli buffer (4% SDS, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris HCl, PH 6.8) in a 1.5 ml Eppendorf tube, 2 µl of β-mercaptoethanol were added to the samples and boiled at 95°C for five minutes. Proteins were separated on a 12% SDS-PAGE gel at 100V for 2 hours, then electrophoretically transferred onto a 0.45 µm nitrocellulose membrane (GE Healthcare Life Sciences). After protein transfer, membranes were blocked in 5% milk PBST (1XPBS 0.1% Tween 20) for one hour and then incubated with primary antibodies overnight at 4°C (See table 2.2 and 2.3 for antibodies). Membranes were washed and incubated with 1:3000 anti-mouse-HRP or anti-rabbit HRP secondary antibodies in 5% milk PBST for 1 hour. Detection was performed using Clarity Western ECL Substrate (Bio-Rad) according to manufacturer's instructions.

2.12 Larval feeding behaviour

Freshly hatched first instar larva (< 30 minutes post hatching) were placed in blue yeast paste, a modified recipe from (Zinke *et al*. 1999) (5 g yeast (Saf-Levure), 5 ml water and 0.1 gr Bromophenol Blue (Sigma Aldrich)) in agar plates and left to feed on this medium for one hour at 25°C. Then, they were washed from the

yeast paste with 1XPBS and scored for the presence of food in their guts by using a Leica MZ75 dissecting Scope. For visualization of the digestive system, yeast paste was mixed with DAPI (20 µg/ml DAPI in yeast paste) and the larvae were fed as previously described, with the only difference that after one hour of feeding the larvae were mounted in 70% glycerol 1XPBS and visualized on a Leica DM600 fluorescent microscope to observe the stained nuclei.

2.13 Statistical analysis

Statistical analyses were carried out in Prism GraphPad 6.0 software package (http://www.graphpad.com/scientific-software/prism/). Unpaired, two-sided t-tests or non-parametric, unpaired Mann-Whitney tests to compare ranks were used, depending on the features of the data. Significance level was binned according to p-values' probability: Non significant (n.s) $p > 0.05$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$, **** $p < 0.0001$.

Chapter 3

The cleavage and polyadenylation machinery in *Drosophila*

3.1 Chapter overview

More than 60% of all human and mammalian genes display alternative polyadenylation (APA) (Derti *et al*. 2012; Hoque *et al*. 2013). APA allows the generation of mRNAs with different 3'UTRs, which can control mRNA localization, stability and translation efficiency, therefore modulating protein function (Di Giammartino *et al*. 2011). In addition to this, approximately half of all genes in other vertebrates, such as zebrafish, and invertebrates, such as *C. elegans* and *Drosophila*, also display reported APA isoforms (Ulitsky *et al*. 2012; Smibert *et al*. 2012; Jan *et al*. 2011), supporting the notion that APA is a pervasive phenomenon for controlling gene expression. Although the mechanisms by which APA is controlled are not well understood, one prevalent model proposes that the relative abundance of core cleavage and polyadenylation (CPA) factors within the cell and differential strength of polyadenylation sites (PAS) can modulate PAS selection (Takagaki *et al*. 1996). In this chapter, I address the strength of this model in the context of neural development in *Drosophila melanogaster*, a system where the expression of long 3'UTR isoforms have been shown to be restricted to the nervous system (Thomsen *et al*. 2010; Hilgers *et al*. 2011; Smibert *et al*. 2012). I identify the *Drosophila* orthologues of core human CPA factors and I observe that their expression levels decrease over the course of embryogenesis. I also observe an enrichment of CPA expression in the nervous system, suggesting that the proposed model for APA selection may operate in this developmental context. I also focus on Cleavage Factor I (CFI), which is the most evolutionarily conserved CPA complex and most highly expressed in neural tissues. The expression patterns CFI components CFI25 and CFI68 show strong enrichment in the nervous system and are also essential for viability to adulthood.

3.2 Results

3.2.1 The *Drosophila* cleavage and polyadenylation machinery is as complex as its human counterpart and has lower gene redundancy

As discussed in Chapter 1, since the discovery of CPA factors in the late 1980s, there has been evident progress in the exploration of mechanisms that control APA. For example, one key experiment which provided experimental evidence for the current model of APA controlled by CPA factor abundance is the IgM heavy chain switch controlled by the abundance of CSTF64 in B-Lymphocytes (Takagaki *et al*. 1996). However, this and other similar studies conducted on cells in culture present limitations, as they cannot completely reproduce what is observed in multicellular organisms. Complex processes cannot be recapitulated in cell culture, for example the formation of the nervous system, a tissue of great interest in this context, given its property of producing extensively long 3' UTRs (Thomsen *et al*. 2010; Hilgers *et al*. 2011; Smibert *et al*. 2012). A similar issue arises in yeast, which was widely used for biochemical identification of CPA factors in the first place. This is compounded by this unicellular organism also having low conservation at the protein level in CPA factors, when compared to mammals.

Thus, in order to consider the nervous system in its complexity and to address the question on how APA is controlled within this tissue, a system is required that allows genetic manipulations using the whole organism. Because of this, the fruit fly *Drosophila melanogaster* emerges as an excellent system to study this question given that it is a well-established model organism that is widely used for genetic manipulations including availability of mutants, a variety of genetic tools,

a rapid life cycle, low costs of maintenance and large numbers of individuals that can be generated (Venken and Bellen 2005; Jeibmann and Paulus 2009; Brand and Perrimon 1993). In addition, fruit flies possess a nervous system that, although not as complex as its mammalian counterpart, still undergoes pervasive 3' UTR lengthening (Thomsen *et al.* 2010). However, in order to use *Drosophila* to address the question on conserved CPA factor expression control during neural development, it is required to identify and evaluate the presence of the known mammalian CPA factors in this system.

Shortly after the sequencing of the *Drosophila* genome in 2000 (Adams *et al*. 2000), a study used these earliest sequences to look for the presence of orthologues of mammalian splicing factors (Proteins and small nuclear RNAs) in *Drosophila*, as well as proteins related with mRNA processing in general (Mount and Salz 2000). In this study, the authors found evidence for the presence of a subset of members of the 3'-end machinery, although in many cases functional data for the identified *Drosophila* gene was not available. For example, *CG5222* was labelled as "Related to CPSF-100 and -73", although the latest genome assembly of *Drosophila* (BDGP6) shows no relation of this gene with either CPSF100 or CPSF73 in terms of paralogs and orthologues. Therefore, in order to be able to use *Drosophila* to address the question on how APA is controlled during neural development, I used the latest release of both the human (Yates *et al*. 2016) and *Drosophila* (Attrill *et al*. 2016) genomes to identify the *Drosophila* orthologues of all the known core CPA factors by BLAST (BLASTN version 2.5.0) (Johnson *et al*. 2008). I found that indeed all core members of the human CPA machinery have orthologues in *Drosophila*. Moreover, when also comparing the known yeast proteins against their human counterparts, I found that most

Drosophila CPA factors have much higher similarity at the protein level, as shown in Figure 3.1. When we consider other studies that also analysed the presence of *Drosophila* orthologues of human genes including general transcription factors (Aoyagi and Wassarman 2000) and genes involved in human diseases (Bier 2005), these results are expected. This high level of similarity observed between the human and the *Drosophila* CPA machinery also suggests that the process of cleavage and polyadenylation is not only ancient, but also has been under considerable evolutionary selective pressure, which in itself is another advantage for this study, meaning that our results can be extrapolated to humans. It has been recently shown that other molecular complexes involved in chromatin modifications and association with RNA PolII, like the PAF complex (Shi *et al*. 1996) can also play roles in APA in mammalian cells (Yang *et al*. 2016). The orthologues of the PAF complex have been investigated in *Drosophila* (Adelman *et al*. 2006) and striking differences were shown in terms of subunit composition and function between these species. Because of this, we will only consider the core CPA factors (Figure 1.6) in this study. Shortly after our own analysis of *Drosophila* orthologues of CPA factors, a study in 2012 revealed orthologues of members of the CPA machinery in 14 different species, including *Drosophila* using BLAST (Darmon and Lutz 2012). In this study, the authors compared the protein domains of CPA factors in these 14 species to assess their structural similarity. Because of this, we did not go on to study the specific protein domain similarity between humans and *Drosophila* since we expected, from our protein sequence similarity analysis, that they are highly analogous between the two species.

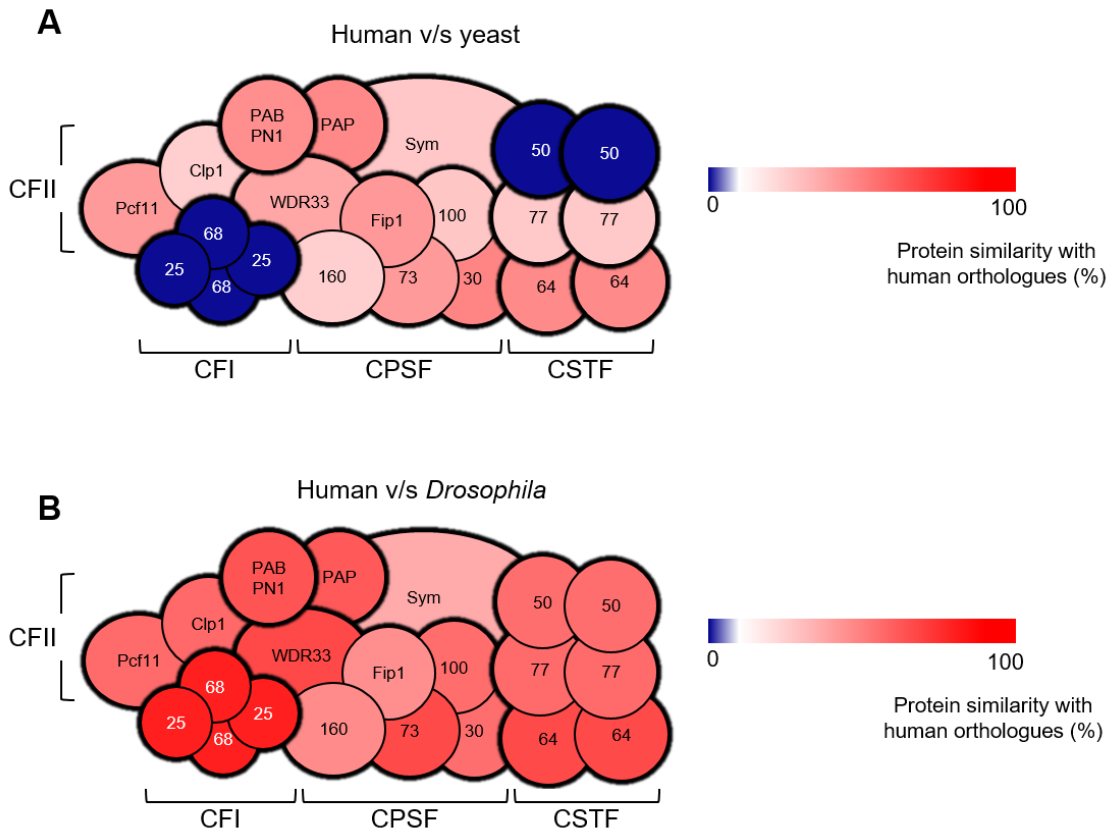


Figure 3.1 The *Drosophila* CPA machinery is as complex as its human counterpart

(Legend on the following page)

Figure 3.1 The *Drosophila* CPA machinery is as complex as its human counterpart

Diagram of the human CPA machinery indicating the similarity with its yeast **(A)** and *Drosophila* **(B)** orthologues by colour code. Please note that the architecture of the yeast CPA machinery is different, but the human architecture was used in all diagrams for simplicity. The blue colour indicates the absence of orthologues. Note that all factors have orthologues in the *Drosophila* genome and they show higher similarity than their yeast counterparts. The sequences of the human proteins were extracted from “Uniprot” (Bateman *et al*. 2015), the sequences of the yeast proteins were extracted from “Saccharomyces Genome Database” (Cherry *et al*. 2012) and the sequences of the *Drosophila* proteins were extracted from “Flybase” (Attrill *et al*. 2016), when more than one protein isoform was present in each species, I used the most similar ones for this diagram, taking in account the covered region. Protein sequence comparison was done using the NCBI protein-protein BLAST (BLASTP version 2.5.0) (Altschul *et al*. 1997).

Another consideration for the use of *Drosophila* to study the process of APA, is that this model organism has been shown to possess low gene redundancy when compared with mammals. Thus, genetic alterations in *Drosophila* can be informative because of lack of redundant genes to cover their function. A good example of this phenomenon are the Hox genes. These key developmental genes have only one copy each in the *Drosophila* genome, the model organism in which they were discovered (Lewis 1978; Mallo and Alonso 2013) but they can have up to four copies each, known as paralogs, in the human and mouse genome, with redundant functions (Favier and Dollé 1997; Soshnikova *et al*. 2013). To address these considerations, I asked the question of how redundant are the CPA factors in the *Drosophila* genome in comparison with the mammalian CPA machinery. This analysis will give valuable information for subsequent interpretation of experimental data and will also indicate if, similar to the Hox genes, *Drosophila* is a good system to study these factors to uncover gene function and biological roles.

To achieve this, I analysed both human and *Drosophila* CPA factor genes in "Ensembl" (Yates *et al*. 2016) to scan for paralogues. I observe that only 3 *Drosophila* factors: CPSF73, CPSF30 and CSTF50 have one paralogue in the genome. In contrast, 7 human factors: CFI68, CPSF73, CPSF30, CSTF64, CSTF50, PAP and PABPN1 had paralogues in the human genome, with PAP, PABPN1 and CPSF30 having two paralogues each, as shown in Figure 3.2. These results confirm our hypothesis that while the *Drosophila* CPA machinery is as complex as its human counterpart, it also displays lower gene redundancy. Studies with the mammalian CPA machinery have shown that paralogs of these factors can have specialized functions that are tissue-specific. For example, the

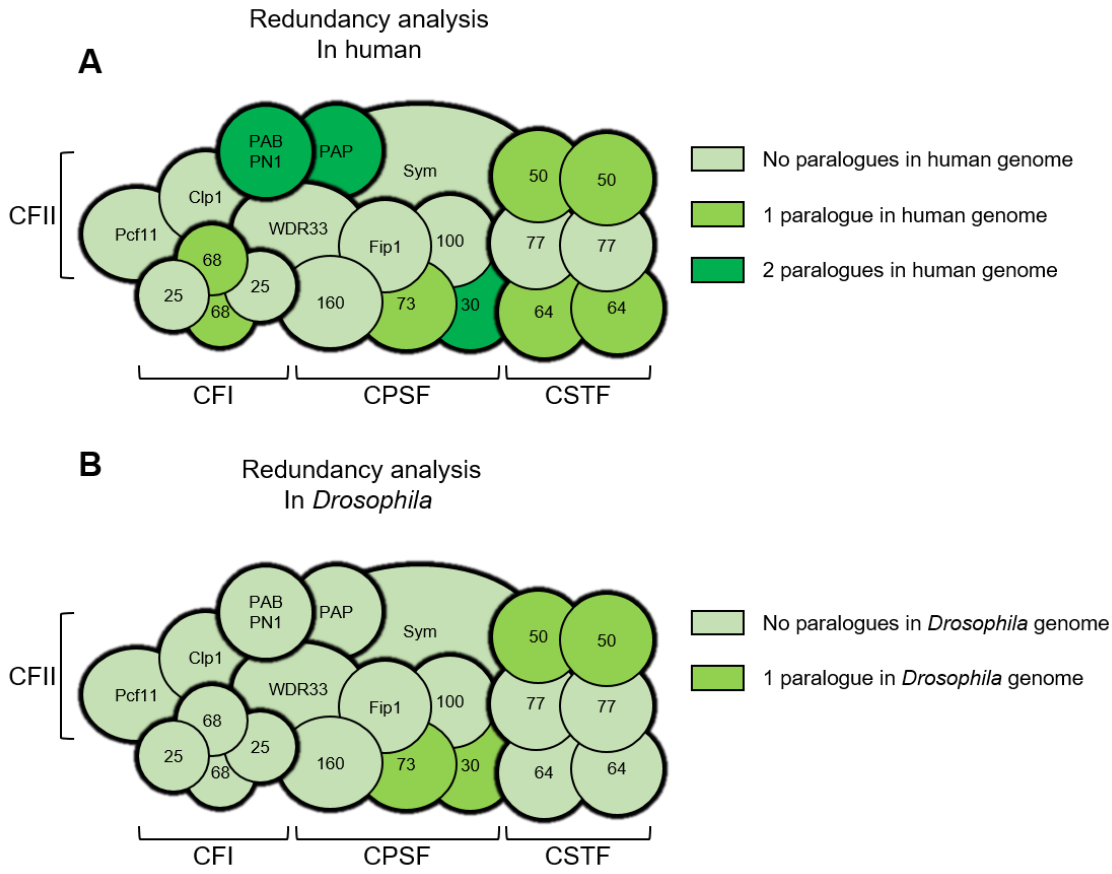


Figure 3.2 The *Drosophila* CPA machinery has lower gene redundancy than its human counterpart

Gene redundancy analysis for the human **(A)** and *Drosophila* **(B)** CPA machinery. Note that the human complex has more paralogues in its genome, with two copies for PABPN1, PAP and CPSF30. The identification of paralogues in the human and *Drosophila* genome was done using “Ensembl” (Yates *et al*. 2016).

paralog of CSTF64 is known as τCSTF64 (tau-CSTF64). This paralogue has been shown to mediate testis-specific PAS selection in the mouse, given that the normal CSTF64 gene is inactive during male meiosis (Dass *et al*. 2001).

Interestingly, *in vitro* experiments show that CSTF64 and τCSTF64 display different binding affinities for RNA (Monarez *et al*. 2007), pointing towards the complexity of APA regulation in mammalian tissues. Another example of this is PAP, which displays 2 paralogs: “PAP β” and “PAP γ”: PAP γ, also known as “Neo-PAP”, was discovered in 2001 from human tumour cDNA samples. While normal PAP is phosphorylated throughout the cell cycle, Neo-PAP did not show signs of phosphorylation, even though this paralogue is also found in the nucleus, similar to the canonical PAP (Topalian *et al*. 2001). These studies show that even tissues with abnormally high rates of proliferation, such as human tumours, also make use of tissue-specific CPA factor paralogues to achieve their functions. Similar to the Hox genes, functional redundancy and back-up genes are an established feature of the mammalian CPA machinery. Thus, the low redundancy observed in *Drosophila* also makes it an excellent system to study the molecular mechanisms controlling APA during neural development.

3.2.2 Analysis of evolutionary conservation v/s molecular function of CPA factors between humans and *Drosophila*

To understand the key steps during 3’ end termination and APA, I decided to observe the evolutionary conservation of the most conserved CPA factors between *Drosophila* and humans and in this way relate their conservation with their biological functions. From the analysis shown in Figure 3.1, I observe that the members of CFI complex: CFI25 (*Drosophila* CG3689) and CFI68

(*Drosophila* CG7185) are the most conserved proteins between *Drosophila* and humans, with 77% and 79% similarity, respectively. Furthermore, their protein domains are also conserved, which will be discussed later in more detail, as it indicates functional similarities. These factors were shown to bind to UGUA sequences upstream of the polyadenylation sequence (PAS). As I mentioned earlier, the crystal structure of CFI suggests that an RNA-looping mechanism works to control APA (Figure 1.3) (Yang *et al*. 2011). These observations regarding CFI, as well as its reported effects in APA in cells in culture (Masamha *et al*. 2014; Kubo *et al*. 2006), make this molecular complex an interesting candidate to study further in *Drosophila* by using known information regarding its function in mammalian cells to test if they operate within the nervous system to control APA in flies as well.

The next factors in the similarity ranking are two components of CPSF: CPSF73 and WDR33, and one component of CSTF: CSTF64, all of them showing 67% similarity at the protein level between humans and *Drosophila*. As described earlier, CPSF73 is the endonuclease that cleaves pre-mRNA substrates (Banerjee *et al*. 2006) and WDR33 (*Drosophila* CG1109) was the final component of CPSF discovered (Shi *et al*. 2009), which has been shown to bind pre-mRNAs, specifically the PAS, both *in vitro* and *in vivo* (Schonemann *et al*. 2014). This suggests that WDR33 is the CPSF member responsible for PAS binding, and not CPSF160, as has been postulated in previous studies (Murthy and Manley 1995). CSTF64 binds to pre-mRNAs and participates in the recognition of the GU-rich downstream element (Perez-Canadillas and Varani 2003).

PAP and PABPN1, showing 61% and 63% protein similarity in *Drosophila* and humans, respectively, are also considered in our ranking study. PAP (*Drosophila*

hrg) is the polymerase that synthesises the poly(A) tail of pre-mRNAs. Since its discovery (Wahle 1991b), its structure and function has been well established in the literature (Martin *et al*. 2000; Balbo and Bohm 2007). PABPN1 (*Drosophila* Pabp2) binds to the poly(A) tail and controls its length and PAP processivity (Kerwitz *et al*. 2003; Wahle 1991a). The next factors in this ranking are the members of CFII: Pcf11 and Clp1 (*Drosophila* cbc), showing 56% and 57% protein similarity, respectively. CFII is the least characterized complex in mammals, and has been studied more thoroughly in yeast (Noble *et al.* 2007). Pcf11 has been shown to interact with the C terminal Domain (CTD) of RNA Polymerase II (RNA Pol II), controlling transcription termination (Hollingworth *et al*. 2006). Clp1 binds adenosine triphosphate (ATP) and interacts with Pcf11, acting as a bridge between RNA Pol II and the CPA machinery (Haddad *et al*. 2012).

A summary of the aforementioned ranking of protein similarity between the *Drosophila* and human CPA machinery, together with diagrams of their biological functions, is shown in Figure 3.3. These observations suggest that both CFI25 and CFI68 perform key roles for *Drosophila* 3' end processing, given their extremely high similarity at the protein level with their human counterparts. What is more, the observation that all core CPA factors are present in the *Drosophila* genome (Figure 3.1) and that the redundancy observed is much lower than that of the mammalian counterparts (Figure 3.2) show us that *Drosophila* is an excellent system to pursue study of the control of APA during neural development. In order to explore this, I address the expression levels of CPA

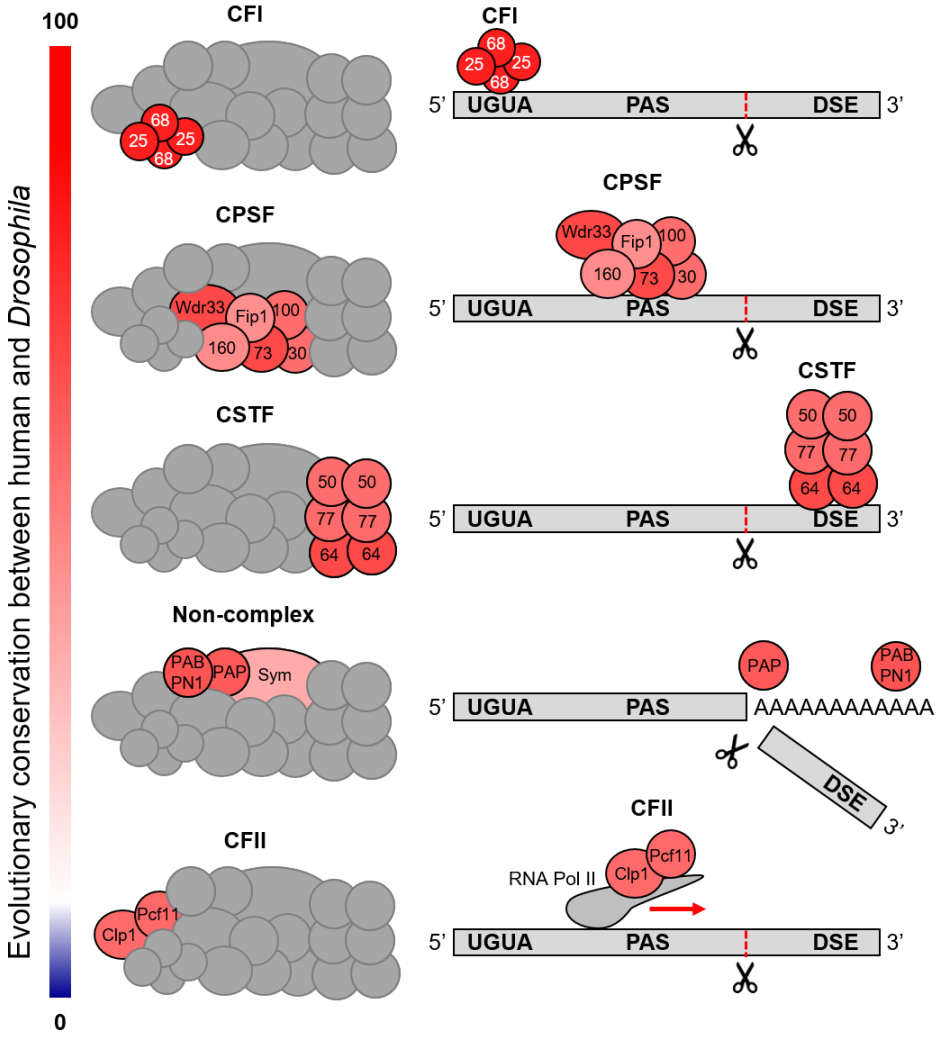


Figure 3.3 Protein similarity ranking between the *Drosophila* and human CPA machinery and biological functions

(Legend on the following page)

Figure 3.2 Protein similarity ranking between the *Drosophila* and human CPA machinery and biological function

Diagrams of the CPA complexes colour coded as in figure 3.1, ordered from top to bottom according to their level of protein similarity between *Drosophila* and humans. In each case, a diagram of their biological function is shown, where the grey rectangle represents a pre-mRNA from 5' to 3' with the main cis-regulatory elements indicated. UGUA is the binding site for CFI, PAS is the binding site for CPSF and DSE is the binding site for CSTF. The dotted red lines with the scissors represents the cleavage site. An already cleaved and polyadenylated mRNA is represented for the Non-complex group, to show that PAP and PABPN1 interact with the poly(A) tail. CFII are represented as interacting with RNA Pol II.

factors during embryogenesis, to see if there is a noticeable change that can be related to the onset of 3' UTR extensions observed during late embryogenesis. This experiment will also allow us to test whether 3' UTR neural extensions are achieved by a mechanism in line with the "CPA factor abundance" model by using genetic tools available in *Drosophila*. Further, I examine the expression patterns of *Drosophila* CPA factors in different tissues, to also address whether the nervous system displays differential CPA factor expression that can be related to the observed 3' UTR extensions. These experiments and their results will be described in the following section.

3.2.3 The *Drosophila* cleavage and polyadenylation factors show similar expression levels during embryogenesis and an enrichment in neural tissues

As described in the previous section, all human core CPA factors have orthologues in *Drosophila,* with high levels of protein similarity. Furthermore, I show that the genetic redundancy of the *Drosophila* CPA machinery is considerably lower than in mammals.

Experiments *in vitro* have shown that during differentiation of C2C12 cells into myotubes, there is a pervasive lengthening of 3' UTRs. During this process, a downregulation is observed in the expression levels of all CSTF factors: CSTF50, CSTF64 and CSTF77 (Ji *et al*. 2009). These observations suggest that changes in the expression levels of core CPA factors can be used as a cue for differential PAS selection during cellular differentiation, as shown for the IgM heavy chain switch system mentioned previously (Takagaki *et al*. 1996). Therefore, we asked the question: Do expression levels of CPA factors change throughout *Drosophila*

embryogenesis? This is an important question because as we mentioned previously, during embryogenesis there is a pervasive lengthening of 3' UTRs within the nervous system (Thomsen *et al*. 2010; Hilgers *et al*. 2011). Thus, observation of changes in expression levels of CPA factors during this process will confirm a trend seen in experimental conditions *in vitro* and will also allow us to precisely manipulate this *in vivo* in further experiments.

Data were retrieved from the modENCODE project (Graveley *et al*. 2011) to evaluate the expression levels of all CPA factors during embryogenesis. In this database, the authors use RNA-seq to analyse gene expression in thousands of *Drosophila* transcripts during different life stages, including embryogenesis. Accordingly, gene expression levels are measured as "Reads Per Kilobase of transcript per Million Mapped Reads" (RPKM), which are the measurement units given by the RNA-seq method employed in modENCODE (Illumina poly(A) + RNA-Seq), these measurements are proportional to the number of cDNA fragments that originate from each gene mRNA and are normalized by both sequencing depth and gene length. The RPKM measurements were plotted against 2 hours-time windows from 0 to 24 hours, covering the full period of embryogenesis. To represent these data I grouped CPA factor expression by complex, as shown in Figure 3.4. From this analysis, it can be observed that the members that form part of the same complex show comparable expression levels across embryogenesis, with high expression levels during early embryogenesis and low expression levels during late embryogenesis. To show this trend, I transformed the RPKM values to relative levels, using the highest expression point as 100% and modifying the rest accordingly, with the average of the relative expression levels of all CPA complexes shown as a red curve. Rpl32, a ribosomal

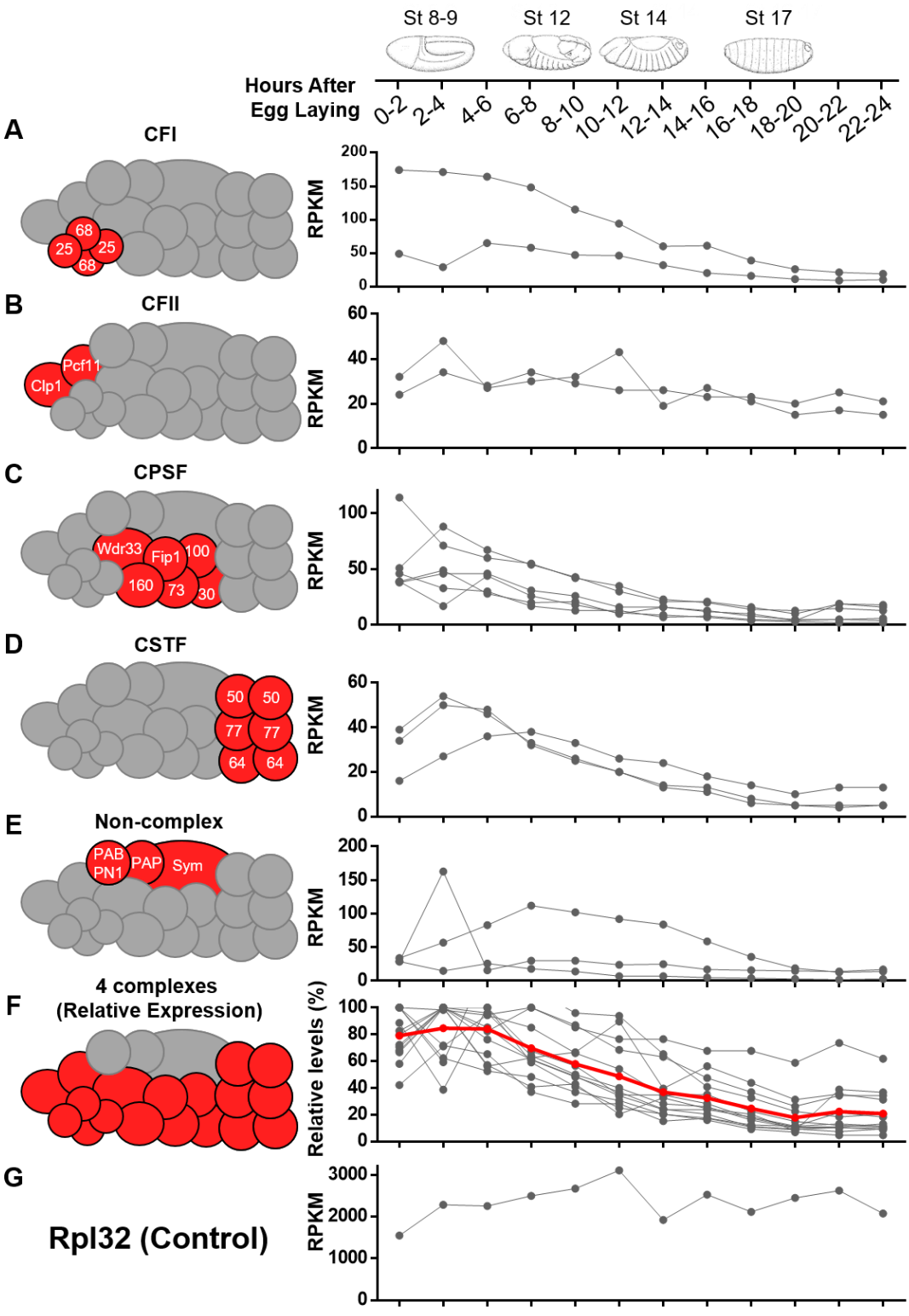


Figure 3.4 The *Drosophila* CPA factors show similar expression levels throughout embryogenesis

(Legend on the following page)

Figure 3.4 The *Drosophila* CPA factors show similar expression levels throughout embryogenesis

(A-G) Graphs representing the RNA expression levels, as measured by RPKM values (Reads Per Kilobase of transcript per Million mapped reads), of CFI **(A)**, CFII **(B)**, CPSF **(C)**, CSTF **(D)**, the Non-Complex group **(E)**, the average of A-D **(F)** and Rpl32 **(G)** throughout embryogenesis, separated in 2-hours windows. Each curve in the diagram represents a member of that complex. At the top, embryo diagrams for four representative stages in the indicated time windows (Modified from Hartenstein, 1993). Note that for the four complexes CFI, CFII, CPSF and CSTF **(A-D)** the expression levels of their members show a comparable trend, having high expression levels during early embryogenesis and low expression levels towards late embryogenesis. In contrast, the members that do not form part of any complex (Non-Complex group) **(E)** do not show this trend. **(F)** Relative expression levels of all CPA complexes, using the highest expression point as 100% and modifying the rest accordingly. **(G)** Rpl32 is used as a negative control to show that this trend is not a biased generated by the technique employed.

gene used as a negative control, does not follow this same trend in expression. Thus, this observed downregulation in expression of CPA factor as development progresses, follows the trend observed during 3' UTR lengthening in C2C12 cells, suggesting that the biological meaning of this decrease in expression, in line with the "CPA factor abundance" model, could be related with the shift from proximal PAS usage to distal PAS usage in the nervous system. However, the observed decrease in expression can be interpreted in another way, given that as development progresses, more cells are produced and the tissues become more complex. Thus, this could be because there is a general reduction in CPA factor expression levels in all cells, or by progressive restricted expression in the nervous system. To address whether this is an actual downregulation of CPA expression or a consequence of development, we need to analyse the expression patterns of CPA factors in detail, which I will explore later in this study.

I show that similar to studies performed *in vitro*, a downregulation in the expression of CPA factors is observed over the course of *Drosophila* embryogenesis. A question that needs to be addressed for the purpose of this investigation is whether this phenomenon is related to the specific control of expression within the developing nervous system. By using the model on CPA factor abundance for the control of APA as our reference, we expect that neural-specific CPA factor expression levels are used in order to achieve its extensive 3' UTR extensions, resulting in levels being higher or lower when compared to tissues in which this pervasive 3' UTR lengthening is not observed. To address this question using available data, I analysed the expression levels of CPA factors in *Drosophila* in different tissues and life stages to see if the nervous system, when compared with other tissues, is different in terms of CPA factor expression,

therefore offering an indication of the mechanisms used for tissue-specific 3' UTR lengthening.

Tissue expression data was retrieved from *Drosophila* third instar larvae and adults, which are the available life stages from the Fly Atlas project (Chintapalli *et al*. 2007). This project employs *Drosophila* microarray chips (GeneChip *Drosophila* Genome 2.0 Array, Affymetrix) mapping the expression of 18770 transcripts, including all the CPA factor genes. The same amount of RNA per tissue (1500 ng) is used for amplification and hybridisation and 4 biological replicates are performed. For each gene and tissue, the mRNA SIGNAL value was extracted, which represents how abundant the mRNA is in that particular tissue and stage. This dataset allowed the analysis of the expression levels of the CPA factors in larval and adult tissues, modifying the error bars as to represent the calculated confidence intervals of the mean (95%) (SEM given in original dataset). Subsequently, a horizontal line is drawn from the lower limit of the CNS category for larva, and from the Brain category for Adults. In this way, statistical significance is revealed when there is no overlap of the horizontal line with the values coming from the other tissues.

The life stages studied within the *Drosophila* life cycle, as well as the tissues sampled for CPA factor expression levels and their embryonic origin in comparison with their human counterparts are shown in Figure 3.5. Although humans and *Drosophila* diverged more than 700 million years ago (Nei *et al*. 2001), the tissues used in this study all have a direct homolog in humans. Thus, these observations could also be extrapolated in the context of human biology. The results of this analysis are shown in figures 3.6, 3.7 and 3.8 and show that all members of CFI and CSTF, together with PAP and symplekin show

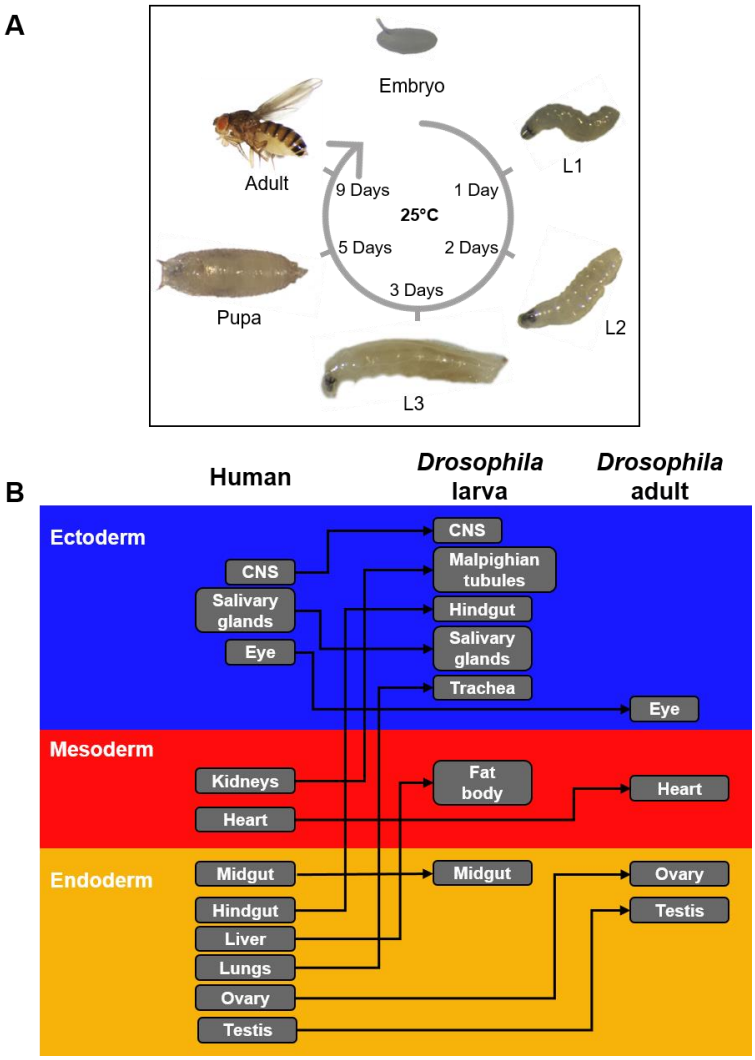


Figure 3.5 *Drosophila* life cycle and stages used for tissue expression analysis with their developmental origin compared between humans and *Drosophila*

(Legend on the following page)

Figure 3.5 *Drosophila* life cycle and stages used for tissue expression analysis with their developmental origin compared between humans and *Drosophila*

(A) Life cycle diagram of *Drosophila*. For CPA factor tissue expression analysis the category “Larva” means third instar larva (L3) and “Adult” means 7 days-old males and females, both from the “Canton S” wild type strain. A female is depicted in the diagram. The time that each developmental stage takes since egg laying at 25°C is shown. **(B)** Diagram of the larval and adult tissues used in this study and their human analogues. Organised by the embryonic layer from which they form. The arrows connect analogous tissues between the two species. Note that all larval and adult tissues used have a corresponding homolog in humans and mammals in general. The origin of *Drosophila* larval and adult tissues was retrieved from (Hartenstein, 1993). The origin of human tissues was retrieved from (Sadler, 2011).

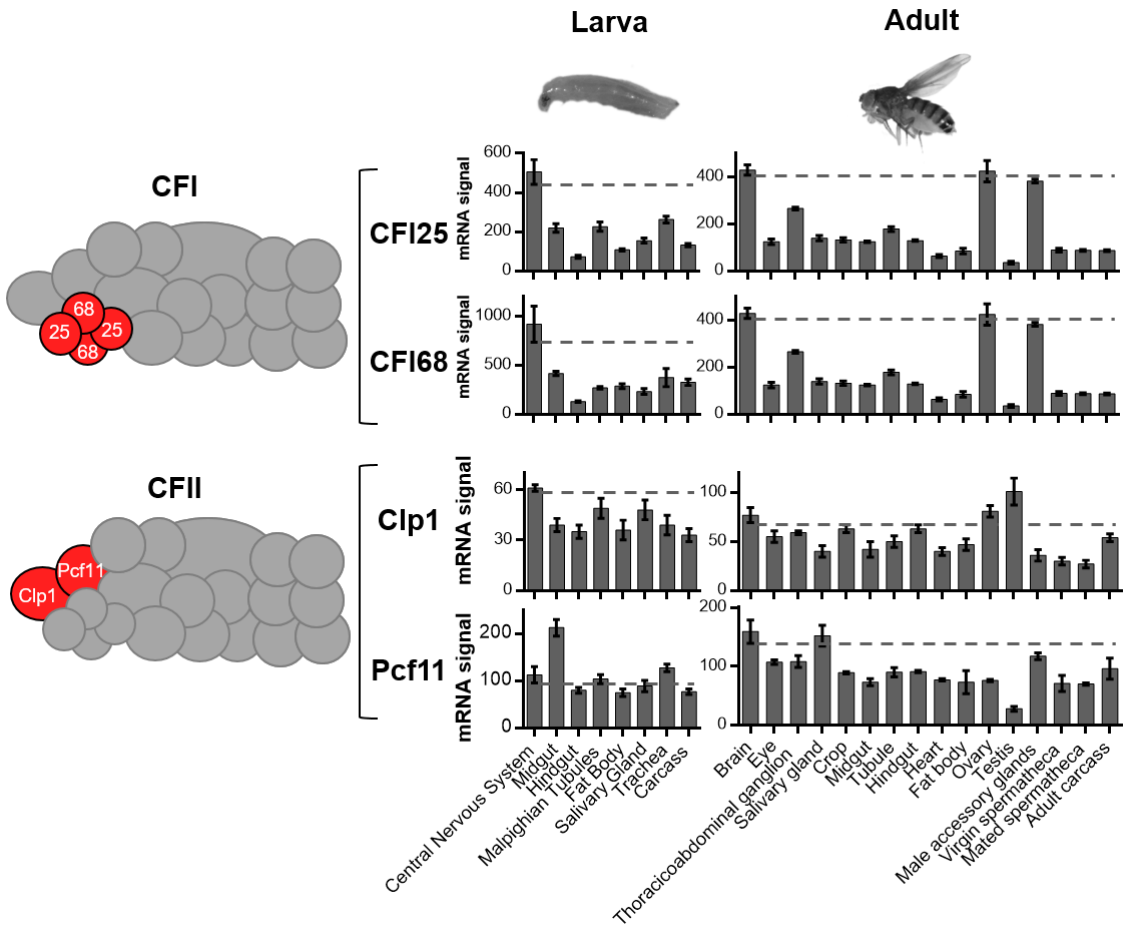


Figure 3.6 mRNA expression analysis of CFI and CFII factors in *Drosophila* larval and adult tissues

(Legend on the following page)

Figure 3.6 mRNA expression analysis of CFI and CFII factors in *Drosophila* larval and adult tissues

mRNA expression levels for CFI and CFII factors in larval and adult tissues. Error bars represent the calculated confidence intervals of the mean (95%) (SEM given in original dataset). A horizontal line is drawn from the lower limit of the CNS category for larva, and from the Brain category for Adults to display statistical significance when compared with other tissues. In larvae, CFI25, CFI68 and Clp1 show significantly higher expression levels in the CNS when compared with the rest of larval tissues. This trend is not seen in adult tissues, in which usually the ovaries, testis and male accessory glands show the higher expression levels. “T-Ab ganglion” stands for Thoracicoabdominal ganglion.

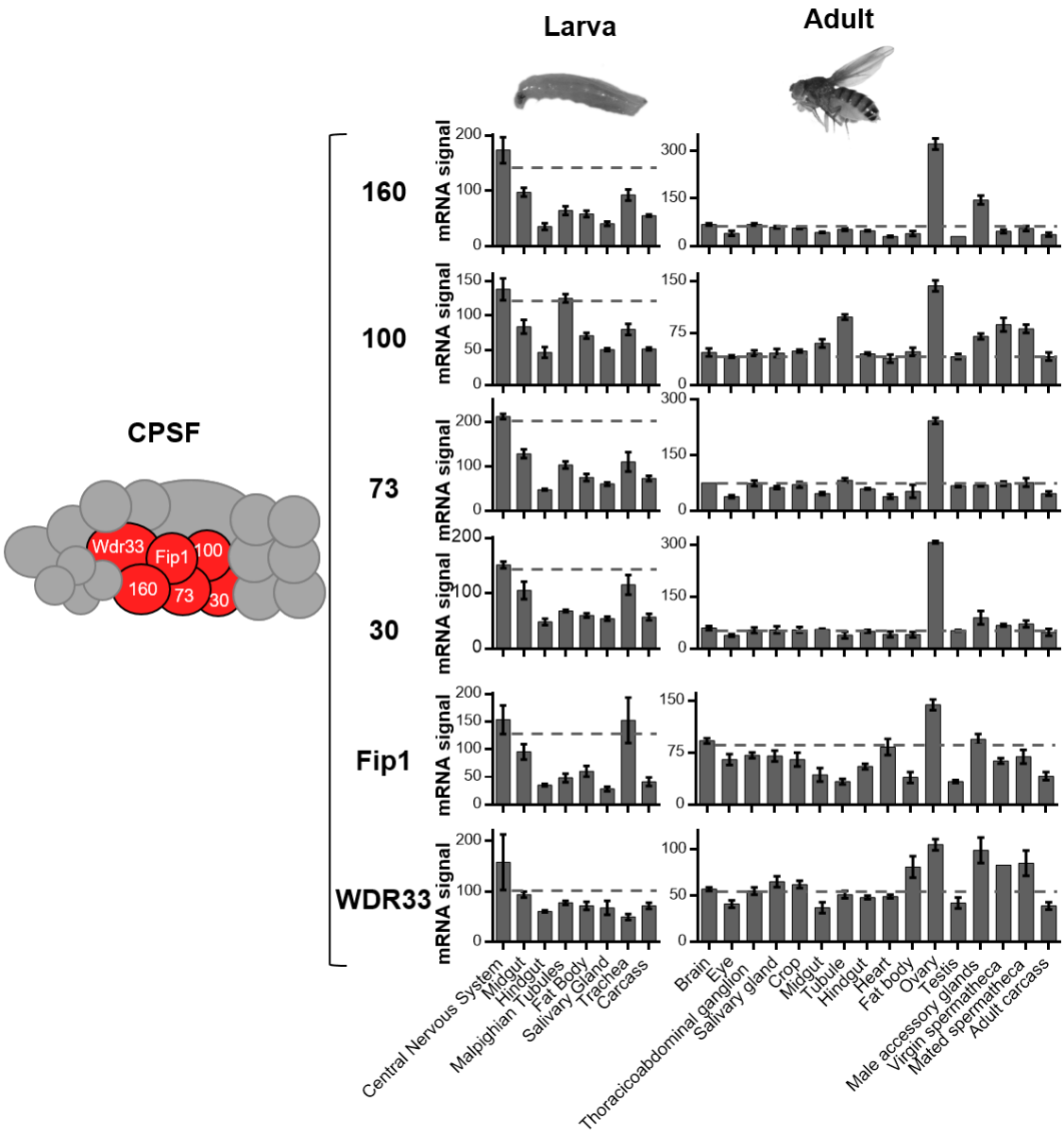


Figure 3.7 mRNA expression analysis of CPSF factors in *Drosophila* larval and adult tissues

mRNA expression levels for CPSF factors in larval and adult tissues. In larval tissues, CPSF160, CPSF73 and CPSF30 show significantly higher expression levels in the CNS when compared with the rest of larval tissues. This trend is not seen in adult tissues, in which ovaries show much higher expression levels compared with the rest. Error bars calculated as in Fig 3.6

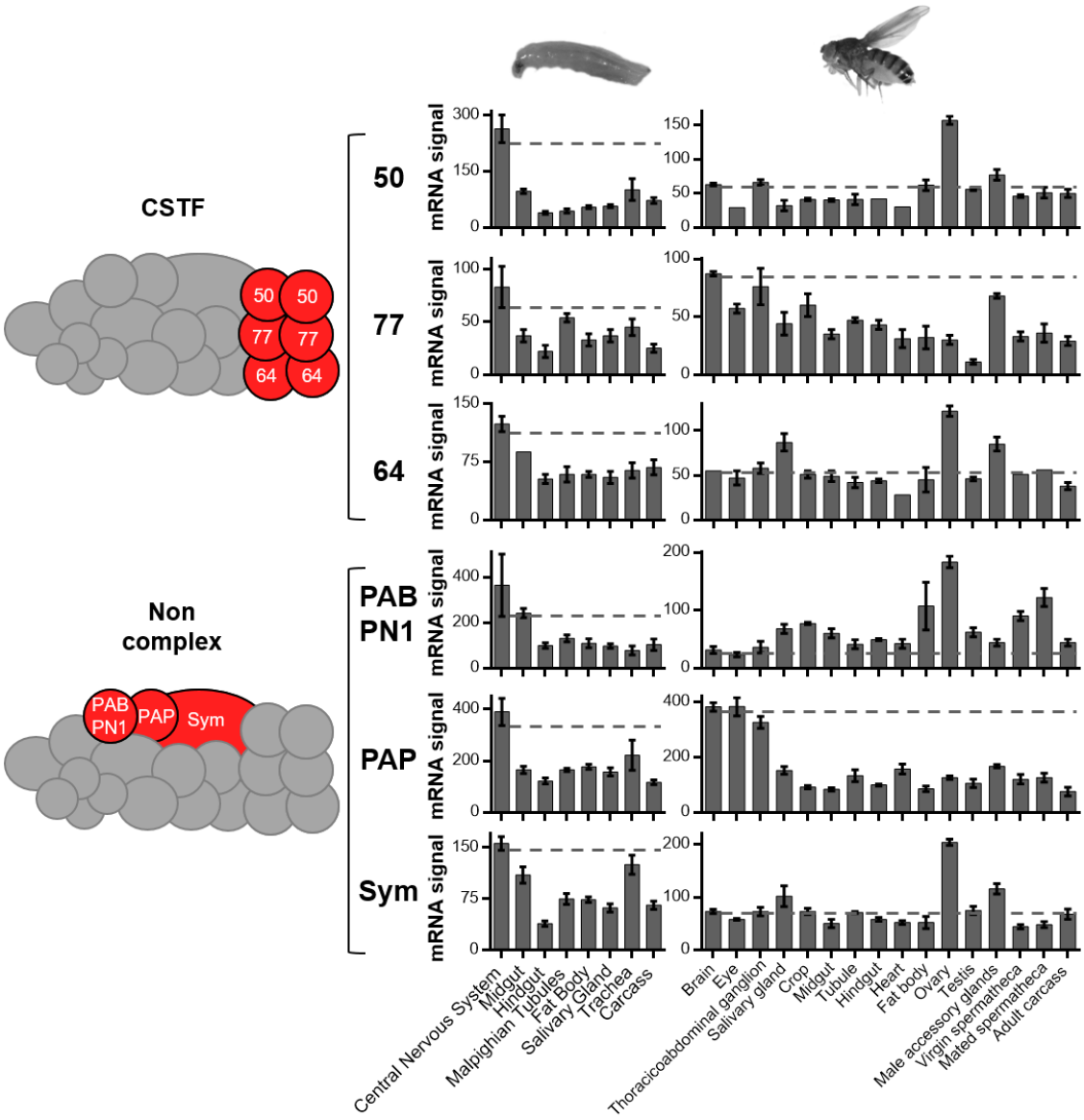


Figure 3.8 mRNA expression analysis of CSTF and "Non-Complex" factors in *Drosophila* larval and adult tissues

mRNA expression levels for CSTF and "Non-Complex" factors in larval and adult tissues. In larval tissues, CSTF50, CSTF77, CSTF64, PAP and Sym show significantly higher expression levels in the CNS when compared with the rest of larval tissues. This trend is not seen in adult tissues, in which ovaries show much higher expression levels compared with the rest. Error bars calculated as in Fig 3.6

significantly higher expression levels in the larval nervous system when compared with the rest of the tissues in larvae and adult flies. Conversely, CFII, CPSF and PABN1 do not show an enrichment in the larval nervous system, although some specific members of these complexes still show neural enrichment in larvae, including Clp1, CPSF160, CPSF73, CPSF30 and WDR33. On the other hand, none of the CPA factors showed any enrichment in the adult brain when compared with other adult tissues.

The tissues with the highest expression levels of CPA factors in adults are actually the testis and ovaries, together with their accessory glands and structures. These observations suggest that CPA factors may play crucial roles during the formation and rewiring of the nervous system in larvae, possibly requiring the orchestration of APA patterns in genes involved in this process. Once the developed adult organism is formed, CPA factors may play only a physiological role while their developmental one required during earlier stages is not strongly required anymore. Thus, the high expression levels observed in the adult germline might reflect the importance of the maternal transmission of CPA factors to the next generation. These results suggest that the biological meaning of this neural enrichment in CPA factor expression can be related to the extensively long 3' UTRs that are observed in this tissue, a hypothesis that is in line with the discussed model of CPA factor abundance for PAS selection (Takagaki *et al.* 1996). Thus, in order to have an indication of which CPA factors may be key for neural APA to then be experimentally tested, I compare only the neural expression levels among them, to then study their expression patterns during embryogenesis and address whether the changes in expression levels

observed during this process (Figure 3.4) are caused by restriction of expression within the nervous system.

3.2.4 The members of the CFI complex show the highest expression levels in neural tissues in larvae and adults and show neural expression patterns during late embryogenesis

From the analysis on the expression levels of CPA factors in different tissues in *Drosophila* in section 3.2.2, it can be observed that nearly 70% of all CPA factors show higher expression levels in neural tissues than in other larval tissues. If we compare only the neural expression levels across the 16 different CPA factors, we see that the members of CFI show the highest expression levels, together with PAP and PABPN1, as shown in figure 3.9. Both PAP and PABPN1 are factors that act after the APA decision has been made. As will be further discussed in section 3.2.4, their biological function in *Drosophila*, as analysed by means of mutations, has been covered in the literature (Murata *et al*. 2001; Murata *et al*. 1996; Benoit *et al*. 2005). However, the members of CFI remain uncharacterised in *Drosophila*. As mentioned previously, CFI members have been implicated in the control of alternative polyadenylation (Kim *et al*. 2010). In the case of CFI25, it has been shown to be directly implicated in the development of glioblastoma and in the shortening of 3'UTRs in glioblastoma patients (Masamha *et al*. 2014). Because of these reasons, our work focused primarily on the members of CFI in more detail. First, we analysed the expression patterns of CFI25 and CFI68 during *Drosophila* embryogenesis to see whether they showed an enrichment in the nervous system, the logic underlying this experiment is to investigate whether the tissue-specific enrichment observed in larval and adult

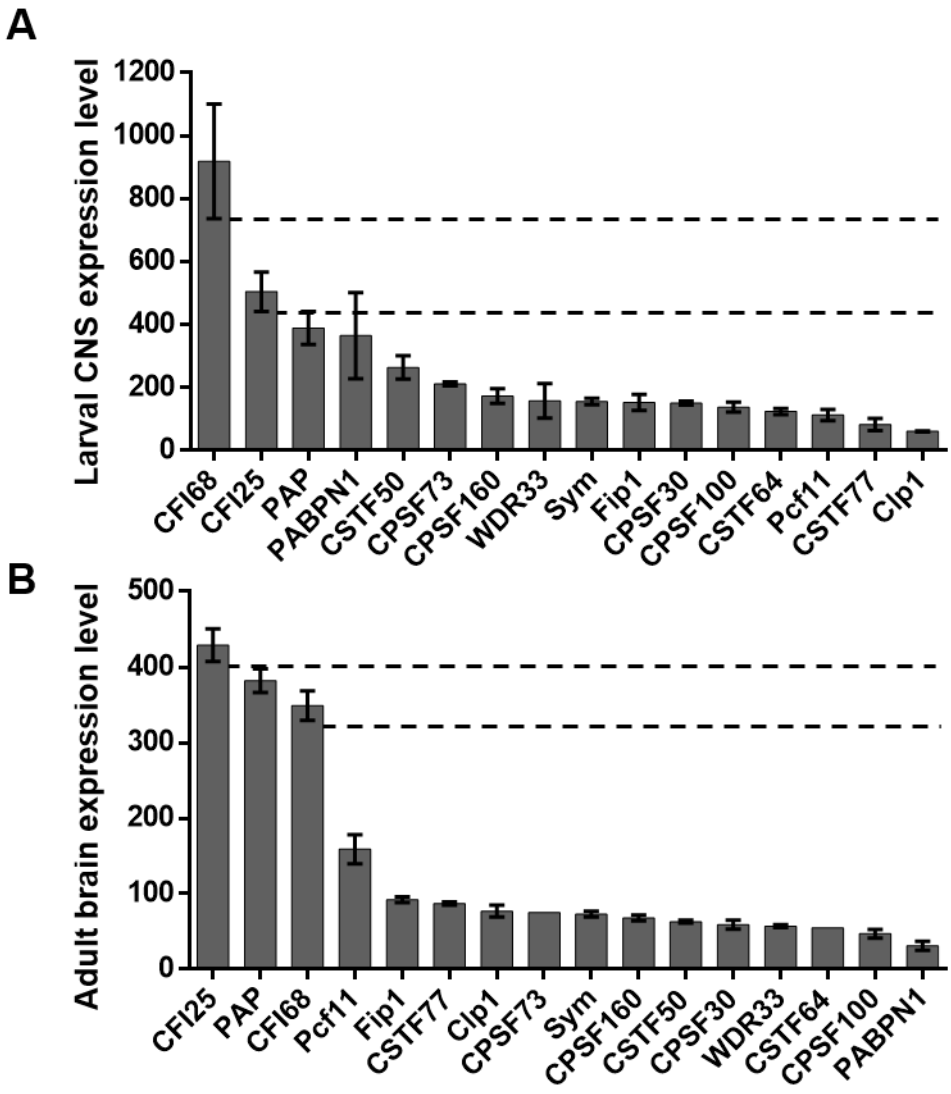


Figure 3.9 Expression levels of *Drosophila* CPA factors in larval and adult neural tissues

(Legend on the following page)

Figure 3.9 Expression levels of *Drosophila* CPA factors in larval and adult neural tissues

(A-B) Graphs representing the expression levels of CPA factors in larval CNS **(A)** and adult brains **(B)**, according to Fly Atlas (Chintapalli *et al*. 2007), ranked from higher to lower values. Error bars calculated as in Fig 3.6. A horizontal line is drawn from the lower limit of CFI25 and CFI68 to show statistical significance when compared with other CPA factors. Note that the CFI members are among the most abundantly expressed genes in these tissues, together with PAP and PABPN1.

neural tissues by the FlyAtlas project also occurs in the neural tissues during embryogenesis. To analyse the expression pattern of CFI25, I designed and synthesised fluorescently labelled anti-sense RNA probes to detect CFI25 mRNA transcripts by *in situ* hybridization (See chapter 2 materials and methods), as a control, I used sense probes from the same sequence in order to consider unspecific signal. From these experiments it can be seen that during early stages of embryogenesis there is ubiquitous signal of CFI25 transcripts. Interestingly, an enrichment of expression can be observed within the central nervous system during late embryogenesis from stage 14 onwards (Figure 3.10). Furthermore, during stage 16, the signal can be seen mostly from the embryonic ventral nerve cord and the brain. These results, when compared with the RNA-seq data from modENCODE during embryogenesis (Figure 3.4) suggest that the hypothesis of a general reduction in CPA factor expression levels in all cells does not account for the observed reduction in CPA factor levels during embryogenesis, but rather it is a result of the progressive restriction of expression within the nervous system, as shown for CFI25.

To analyse the expression pattern of CFI68, we used a GFP reporter line generated recently by Pavel Tomancak's group (Sarov *et al*. 2016) (See table 2.1). The expression pattern of CFI68 as reported by this GFP line, shows the same dynamics as CFI25 (Figure 3.11), by which we observe strong neural expression during late stages of embryogenesis. These results suggest that CFI factors may have a key role for the formation and functioning of the nervous system in *Drosophila*. Surprisingly, CFI factors have not been studied in this context even though the nervous system in the late *Drosophila* embryo has been shown to be a tissue that selectively expresses long 3'UTR isoforms of several

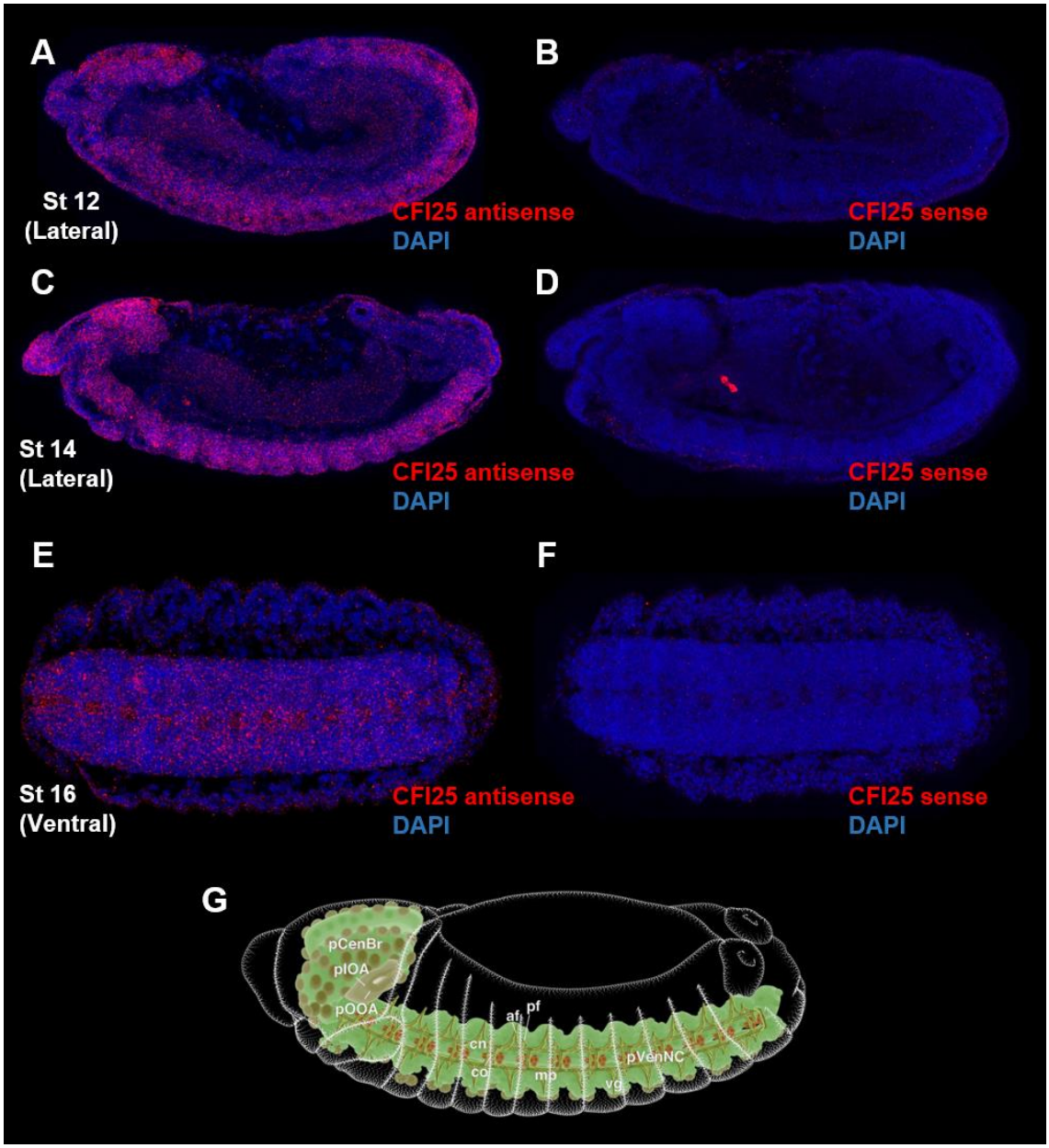


Figure 3.10 CFI25 shows neural expression pattern during late embryogenesis

(Legend on the following page)

Figure 3.10 CFI25 shows neural expression pattern during late embryogenesis

(A-F) CFI25 mRNA staining by *in situ* hybridization in embryonic stages 12 **(A)**, stage 14 **(C)** and stage 16 **(E)**. Control sense probes were used for the corresponding stages **(B, D and F)**. At stage 12 we see a ubiquitous expression pattern, which is also seen for earlier embryonic stages (Data not shown). However, at stage 14 **(C)** the embryos show an enrichment of signal in their ventral nerve cord and brain. During stage 16 **(E)**, the embryos show strong signal in the ventral nerve cord. DAPI was used to label the nuclei and therefore the contour of the embryos. **(F)** Diagram of a lateral view of a stage 13 embryo depicting the location and anatomy of the CNS (Modified from Hartenstein, 1993). In all pictures and diagram, anterior is to the left and dorsal to the top.

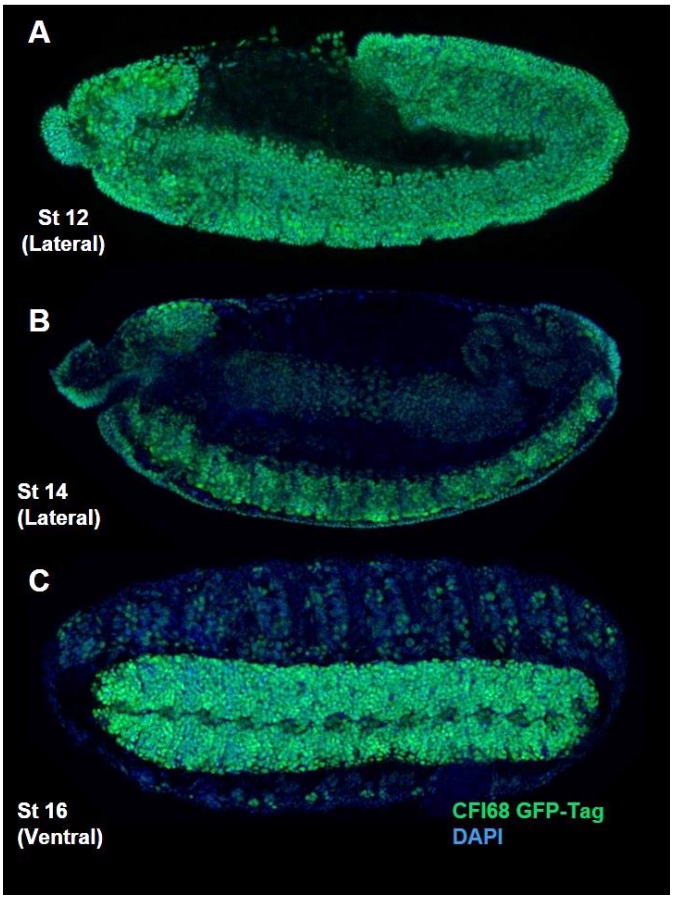


Figure 3.11 CFI68 shows neural expression pattern during late embryogenesis

(A-C) GFP stainings in *Drosophila* embryos for a CFI68-GFP reporter line. **(A)** Lateral view of a stage 12 embryo (Anterior is to the left), during this stage, the expression pattern of this GFP reporter is ubiquitous; this is also the case for earlier stages (Data not shown). However, during embryonic stage 14 **(B)** the GFP signal shows an enrichment in the central nervous system (Brain and ventral nerve cord). During stage 16 **(C)** the ventral nerve cord shows strong GFP signal (Ventral view). DAPI was used to label the nuclei and therefore the contour of the embryos. Anterior is to the left and dorsal to the top.

genes (Thomsen *et al*. 2010; Hilgers *et al*. 2011; Smibert *et al*. 2012). Thus, I address the question on what are the biological roles played by CFI factors in *Drosophila*, as well as other CPA factors.

3.2.5 Biological roles of *Drosophila* cleavage and polyadenylation factors

Drosophila offers a great advantage to assess the biological roles of CPA factors, not only because of its homology with mammalian CPA factors and its lower redundancy (Figure 3.1 and 3.2), but also because of the accessibility to collections of mutant stocks which are publicly available and can be used to study gene function. Although more than 85% of *Drosophila* CPA factors have available mutations as transgenic stocks, only four of them have been studied in this context to uncover their biological roles: Pcf11 (Milchanowski *et al*. 2004), PABPN1 (Benoit *et al*. 2005; Kwon *et al*. 2013), PAP (Murata *et al*. 2001; Cruz *et al*. 2009; Milchanowski *et al*. 2004) and CSTF77 (Audibert and Simonelig 1999; Fitch *et al*. 1992; Simonelig *et al*. 1996; Perrimon *et al*. 1989; Schalet and Lefevre 1973; Peter *et al*. 2002). Unexpectedly, the study of the majority of CPA factors have been largely ignored in *Drosophila.* A representation of this discrepancy is shown in Figure 3.12 and a detailed analysis is shown in Table 3.1. From the four studied CPA factors, the variety of phenotypes observed is diverse, from defects in wing morphology in the case of PAP, where this gene plays a role in the Notch cascade (Murata and Ogura 1996) to developmental defects as in the case of PABPN1, where this gene controls poly(A) tail length during embryogenesis (Benoit *et al*. 2005). These observations suggest that the CPA machinery is key for animal formation and function at the organismal level beyond its molecular function. As a result of the analysis of CFI25 and CFI68 (Fig 3.9, 3.10 and 3.11),

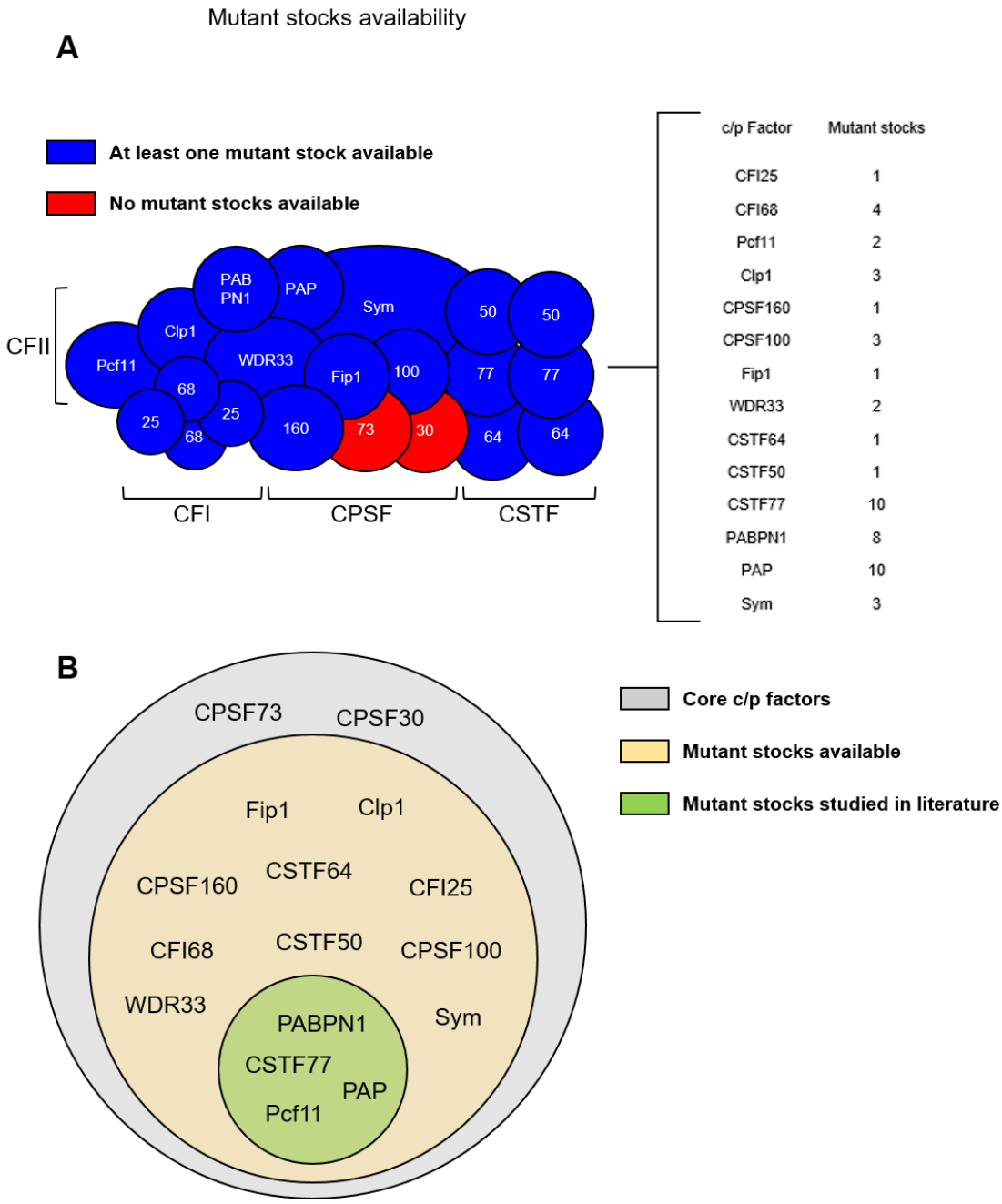


Figure 3.12 Availability of *Drosophila* CPA factor mutant stocks v/s biological roles uncovered by mutational analysis

(Legend on the following page)

Figure 3.12 Availability of *Drosophila* CPA factor mutant stocks v/s biological roles uncovered by mutational analysis

(A) Diagram of the CPA machinery, as used in figure 3.1, representing the CPA factors that have publicly available mutant stocks versus the factors with no available stocks. An account of the number of stocks available is shown in the brackets, with a detailed analysis in table 3.1 **(B)** Diagram showing the CPA factors with available mutant stocks and the factors that have been studied by mutational analysis to uncover their biological functions.

Table 3.1 Availability of *Drosophila* CPA factor mutant stocks v/s biological roles uncovered by mutational analysis

Complex	Gene	*Drosophila* name	Mutant stocks available	Genotype (Simplified)	Nature of mutation	Phenotype	Reference
CFI	CFI25	CG3689	1	P{EP}CG3689^{G19200}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
	CFI68	CG7185	4	P{wHy}CG7185^{DG14101}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				P{PTTGC}CG7185^{CC00645}	Transposable element insertion	No information available	(Buszczak *et al*. 2007)
				P{RS3}CG7185$^{CB-6365-3}$	Transposable element insertion	No information available	(Ryder *et al*. 2004)
				PBac{RB}CG7185^{e04468}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
CFII	Pcf11	Pcf11	2	P{lacW}Pcf11^{k08015}	Transposable element insertion	Reduced number of crystal cells in late embryos	(Milchanowski *et al*. 2004)
				P{lacW}Pcf11^{k08023}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
	Clp1	cbc	3	PBac{RB}cbc^{e00083}	Transposable element insertion	No information available	(Bellen *et al*. 2004)

				cbc^{T7-1}	Point mutation by ethyl methanesulfonate	No information available	Personal communication to FlyBase
				cbc^{T13-5}	Point mutation by ethyl methanesulfonate	No information available	Personal communication to FlyBase
CPSF	CPSF160	CPSF160	1	$P\{EP\}Cpsf160^{G8231}$	Transposable element insertion	No information available	(Bellen *et al*. 2004)
	CPSF100	CPSF100	3	$PBac\{WH\}Cpsf100^{f00376}$	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				$PBac\{RB\}Cpsf100^{e01814}$	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				$PBac\{WH\}Cpsf100^{f00691}$	Transposable element insertion	No information available	(Bellen *et al*. 2004)
	CPSF73	CPSF73	0	-	-	-	-
	CPSF30	Clp	0	-	-	-	-
	Fip1	Fip1	1	$P\{EPgy2\}Fip1^{EY20218}$	Transposable element insertion	No information available	(Bellen *et al*. 2004)
	WDR33	CG1109	2	$P\{wHy\}CG1109^{DG23504}$	Transposable element insertion	No information available	(Bellen *et al*. 2004)

				PBac{RB}CG1109^{e00264}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
CSTF	CSTF77	Su(f)	10	su(f)^{8}	ethyl methanesulfonate	Adult females with reduced number of follicle cells in the ovaries.	(Audibert and Simonelig 1999)
				su(f)^{1}	X ray	enhancement of w^{a} phenotype at 29°C and almost complete suppression of f^{1} phenotype	(Fitch *et al*. 1992)
				su(f)^{5}	Internal deletion by X ray	No information available	(Simonelig *et al*. 1996)
				su(f)$^{16\text{-}185}$	Breakpoint within su(f)	No information available	(Simonelig *et al*. 1996)
				su(f)^{6}	Internal deletion by X ray	Lethality during larval or pupal stages	(Perrimon *et al*.1989)
				su(f)^{3}	diethyl sulfate, ethyl methanesulfonate	su(f)^{4}/su(f)^{3} females show the "pale bristles" phenotype in some of the bristles and hairs	(Schalet and Lefevre 1973)
				P{lacW}su(f)^{G0393}	Transposable element insertion	lethality before end of larval stage	(Peter *et al*. 2002)
				Mi{ET1}su(f)^{MB05229}	Transposable element insertion	No information available	(Bellen *et al*. 2004)

				Mi{MIC}su(f)^{MI07412}	Transposable element insertion	No information available	(Venken *et al*. 2011)
				Mi{MIC}su(f)^{MI08254}	Transposable element insertion	No information available	(Venken *et al*. 2011)
	CSTF64	CSTF64	1	P{EP}CstF-64^{G16431}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
	CSTF50	CSTF50	1	P{ID.GAL4AD}CstF50$^{G10.1}$	Transposable element insertion	No information available	(Gohl *et al*. 2011)
Non-Complex	Symplekin	Symplekin	3	P{EPgy2}Sym^{EY20504}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				Sym^{30m1}	P-element activity	No information available	Personal communication to FlyBase
				Sym^{16m2}	P-element activity	No information available	Personal communication to FlyBase
	PAP	hrg	10	P{RS5}hrg$^{5-SZ-3571}$	Transposable element insertion	No information available	(Ryder *et al*. 2004)
				PBac{WH}hrg^{f06255}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				P{EPgy2}hrg^{EY10340}	Transposable element insertion	No information available	(Bellen *et al*. 2004)

				P{GSV1}hrg$^{s\text{-}222.2}$	Transposable element insertion	Dpp pathway-like phenotype in wing	(Cruz *et al*. 2009)
				P{RS5}hrg$^{5\text{-}HA\text{-}2037}$	Transposable element insertion	No information available	(Ryder *et al*. 2004)
				P{SUPor-P}hrg^{KG01510}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				hrg^{10}	P-element activity	Notched wing margin	(Murata *et al*. 2001)
				P{UAS-RBZ.ftz-F1}hrg^{P1}	Transposable element insertion	Mutants have a notched wing phenotype	(Murata *et al*. 2001)
				hrg^{1}	P-element activity	Reduced number of crystal cells in late embryos	(Milchanowski *et al*. 2004)
				P{lacW}hrg^{k07619}	Transposable element insertion	Reduced number of crystal cells in late embryos	(Milchanowski *et al*. 2004)
	PABPN1	Pabp2	8	P{PTT-GA}Pabp2^{ZCL3178}	Transposable element insertion	No information available	(Benoit *et al*. 2005)
				P{XP}Pabp2^{d09497}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				PBac{RB}Pabp2^{e02513}	Transposable element insertion	No information available	(Bellen *et al*. 2004)

				P{UAST-YFP.RabX1.T19N}Pabp2^{01}	Transposable element insertion	Enhanced wing nicking phenotype caused by dom knockdown in wings	(Kwon *et al*. 2013)
				P{SUPor-P}Pabp2^{KG02359}	Transposable element insertion	No information available	(Bellen *et al*. 2004)
				Pabp2^{55}	P-element activity	Lethality during late embryogenesis, different distribution of polyA tail lengths	(Benoit *et al*. 2005)
				Pabp2^{6}	P-element activity	Development arrested at stage 8, about half have polarity defects or a thin chorion	(Benoit *et al*. 2005)
				P{GawB}Pabp2^{NP5913}	Transposable element insertion	No information available	(Hayashi *et al*. 2002)

I addressed the question of what is their biological relevance for *Drosophila* development. In order to do this, I first analysed mutant lines for CFI25 and CFI68 (alleles P{EP}CG3689^{G19200} and P{PTTGC}CG7185^{CC00645}) (See table 2.1 and 3.1) for viability from embryos to adulthood. By doing this I could obtain an indication of their roles for development, which will be then studied in subsequent chapters. As shown in Figure 3.13, mutants for both CFI25 and CFI68 show lethality and are inviable as adults. The alleles tested are transposable element insertions, meaning that a complete removal of the gene product, as in the case of a null mutation, is not necessarily expected. Nevertheless, a penetrance of 100% is still observed for this phenotype. Furthermore, an unexpected observation in mutants for CFI25 led me to investigate this line further, as we will discuss in Chapter 4.

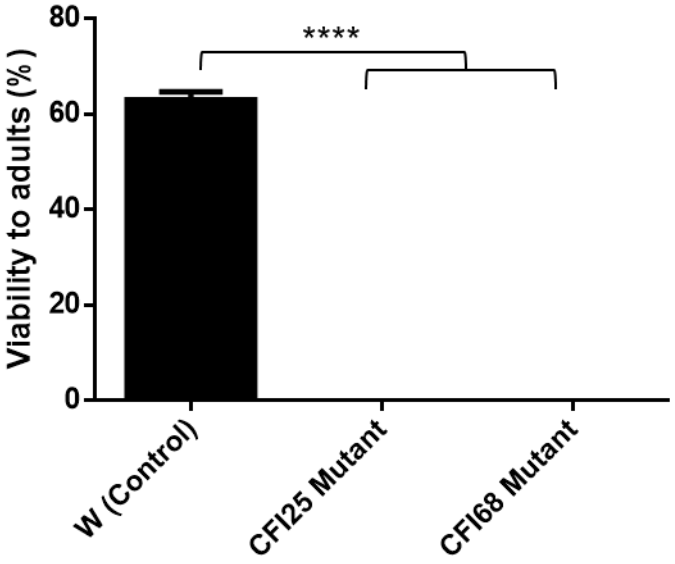


Figure 3.13 CFI mutants are inviable and do not develop into adulthood

Viability essay with wild types (W) and homozygous mutants for CFI25 and CFI68 (P{EP}CG3689^{G19200} and P{PTTGC}CG7185$^{CC00645)}$, the values represent the mean of the percentage of adults (N = 30, biological triplicate). Note that both mutations make the individual inviable as adults. Error bars represent the SEM, statistical significance was assessed with unpaired two tailed t-test (* $p < 0.05$).

3.3 Discussion

The work presented in this chapter first shows that the *Drosophila* CPA machinery is as complex as its human counterpart. The advantages of *Drosophila* as a powerful genetically tractable system make it an excellent system to study the main question of this study, which is to understand how APA is controlled during neural development. Although humans and *Drosophila* diverged more than 700 million years ago (Nei *et al*. 2001), I show that all CPA factors are not only present in the *Drosophila* genome, but show a level of similarity to the human version of the proteins up to 79% for the case of CFI members. These observations, though somehow expected given previous comparisons of gene similarity between *Drosophila* and humans (Bier 2005; Aoyagi and Wassarman 2000), are reassuring for the use of this model system to address the mechanisms controlling APA in the nervous system. Moreover, they suggest that the CPA machinery has been under high evolutionary pressure over time. Accordingly, most of the key protein domains for CPA factor function have been conserved between *Drosophila* and humans (Darmon and Lutz 2012). Exceptions to this fact are, for instance, the absence of the “pro/gly-rich” region in the *Drosophila* CSTF64 protein, also absent in yeast, which is present in the human counterpart (Takagaki *et al*. 1992), this region is adjacent to the hinge, which interacts with CSTF77 and symplekin in a mutually exclusive manner (Ruepp *et al.* 2011). The molecular function of this “pro-gly rich” region is yet unknown. Likewise, a pentapeptide repeat motif known as “MEARA/G” which is next to the “pro/gly rich” domain in the human CSTF64 protein, towards its C-terminal, is also absent in *Drosophila*. This region has also an unknown function, although it has been shown *in vitro* to form a helical structure in solution (Richardson *et al.* 1999).

Although the former domains are not present in the *Drosophila* CSTF64 protein, the key domains for the function of this protein are, such as the RNA Recognition Motif (RRM), which allows CSTF64 to physically bind to pre-mRNAs on their Downstream Sequence Elements (DSE) (Takagaki and Manley 1997), as well as the hinge region and its C-terminal domain, which was shown to form a three-α-helix bundle structure. This structure is required for polyadenylation and for interaction with Pcf11 (Qu *et al*. 2007). Thus, the presence of these domains in *Drosophila* allow us to infer that its function has been conserved during evolution. The results of this chapter also describe that the expression levels of CPA factors during *Drosophila* embryogenesis show a comparable trend. In which high levels of expression are observed during early stages and low levels are observed during late stages of embryogenesis. A question that emerges from this observation is: What is the cause of this general reduction in expression levels? One way of investigating this would be to observe whether common regulatory elements in the form of promoters and enhancers, control CPA factor genes locally in the genome. However, these genes are dispersed in the genome, being located in all 3 main chromosomes and distant from each other (*Drosophila* has four chromosome pairs, but chromosome 4 is much smaller and contains fewer genes (Sun *et al*. 2000)). Thus, gene expression control through common regulatory elements seems unlikely. However, the expression levels mined from databases show that all CPA factors are maternally delivered into embryos. In adults, the high expression levels in both testis and ovaries suggest that CPA factor transcript levels decrease as a function of use during embryogenesis. From this information it could be speculated that during the maternal to zygotic transition and towards late embryogenesis, CPA factors start being expressed

and enriched in the nervous system. Nonetheless, the point at which this process happens for CPA factors is not known, neither are all promoters and regulatory elements that control their expression. I also show that the larval nervous system shows an enrichment in expression of nearly 70% of all CPA factors, with the members of CFI being the ones with the highest expression values not only in larval brains, but also in adult brains. Despite the fact that adult brains do not show an enrichment in CPA factor expression in comparison with other tissues, such as the ovaries and the testis, where an enrichment is definitely observed.

From *in situ* hybridisation, I have clarified this picture by showing that the members of CFI: CFI25 and CFI68 in *Drosophila* embryos show neural expression patterns during late stages of embryogenesis. These results show an interesting correlation with the observed phenomenon of longer 3'UTR isoforms in hundreds of genes observed in *Drosophila* during late embryonic, larval, pupal and adult life (Hilgers *et al*. 2011; Thomsen *et al*. 2010; Smibert *et al*. 2012). This observation also raises the question: How is APA controlled in non-neural tissues given their apparent lack of expression of CFI members? A plausible answer is that because of the detection limits of the techniques employed, non-neural tissues still express CFI and potentially other CPA factors, although in much lower levels than the ones observed in neural tissues. Furthermore, at least in larval and adult tissues we can observe that all other tissues show expression of CPA factors even if the nervous system of larvae or the reproductive organs in adults show considerably higher expression levels (Figures 3.6, 3.7 and 3.8). Fascinatingly, this biological property of the nervous system to express long 3'UTRs has been shown to be even more widespread, concerning also the complex mammalian brain (Miura *et al*. 2013). These observations point towards

the need for mechanisms that can explain this pervasive trend. Experiments addressing this observation have started to be explored in recent years. For example, The RNA-binding protein ELAV was shown by our laboratory to control Hox RNA-processing. More specifically, removal of ELAV leads to a shortening in the 3'UTR of the Hox gene *Ultrabithorax* (Ubx). As a matter of fact, ectopic expression of ELAV during germ band extension stage, during which the short isoform of Ubx is expressed, leads to an increase in the long isoform (Rogulja-Ortmann *et al*. 2014). These experiments show that ELAV is sufficient to control Hox RNA-processing *in vivo*. Yet, the control of APA during neural development by core CPA factors is not well understood. Other experiments have shown that the knockdown of CPA factors can alter APA decisions in cells. Interestingly, CFI25 has been shown to be primarily mediating this process in HeLa cells (Kubo *et al*. 2006; Masamha *et al*. 2014). This has also been shown to be the case for more CPA factors by using siRNA knockdowns in C2C12 cells, where some principles were discovered. For instance, Pcf11 and Fip1 enhance the use of proximal PAS, while CFI25, CFI68 and PABPN1 enhance the use of distal PAS (Li *et al*. 2015).

Finally, I show that although more than 85% of all *Drosophila* CPA factors offer the option of uncovering their biological roles by studying what occurs in mutant conditions given the availability of mutant stocks, only a surprising 25% of them have been studied in this context in *Drosophila*. Hence, the main aims of this thesis are to (i) understand how APA is controlled during neural development and (ii) address the biological roles of APA and core CPA factors. In light of the preliminary and promising results presented in this chapter, the further questions are addressed for the case of CFI in the next chapters.

Chapter 4

Normal expression of the cleavage and polyadenylation factor CFI25 is required for feeding behaviour in *Drosophila*

4.1 Chapter overview

In the previous Chapter I showed that 14 out of the 16 core CPA factors in *Drosophila* have publicly available mutant stocks that can be used to study CPA factor function and biological relevance *in vivo.* So far, only four *Drosophila* CPA factors have been studied in the literature (Figure 3.12): Clp1, PABPN1, PAP and CSTF77. Interestingly, mutations in the studied CPA factors show a variety of effects on different tissues. PAP mutations affects wing morphology (Murata and Ogura 1996) while CSTF77 mutations affects bristle formation in the adult (Schalet and Lefevre 1973) (Table 3.1). These results point towards CPA factors having varied roles at the tissue and organismic level beyond their established molecular functions.

In the following Chapter, I analyse the uncharacterised *Drosophila* gene *CG3689*, which encodes the orthologue of mammalian CFI25. I show that a mutation in CFI25 (by means of an insertion of a transposable element) leads to a developmental arrest at the first instar larval stage. Moreover, the cause of this arrest is a defect in larval feeding, in which CFI25 mutant larvae show a collapsed mouth hook phenotype as well as defects in passage of food from the pharynx to the midgut. I hypothesised that one or more downstream target genes must be affected in CFI25 mutants so that their abnormal APA isoforms or total mRNA expression levels trigger the observed phenotypes. A bioinformatic pipeline provided nine candidate target genes, from which four have affected APA patterns in CFI25 mutants (*S6k*, *klu*, *RanBPM* and *lov*) and three have affected mRNA expression levels (*S6k*, *klu* and *lov*). These results suggest that CFI25 has important roles for *Drosophila* neural development and behaviour.

4.2 Results

4.2.1 CFI25 mutants are developmentally arrested during first instar larvae stage

In Chapter 3 I analysed the *Drosophila* CPA orthologues and showed that the members of CFI are (i) the most conserved factors at the protein level between humans and *Drosophila* (Figure 3.1), (ii) they are among the most abundantly expressed genes in both larval and adult brains (Figure 3.6), (iii) they show neural expression patterns during late embryogenesis (Figure 3.10 and 3.11) and (iv) mutations in CFI25 and CFI68 lead to unviability in adulthood (Figure 3.13). In fact, I observe that CFI68 mutants died during embryogenesis, which was not the case for CFI25 mutants. The reason for this difference can be due to both the nature of the alleles tested and also because CFI25 and CFI68 do not play exactly the same molecular function, as discussed in Chapter 1.

The first question to answer was to understand at what point during the life cycle of *Drosophila* CFI25 mutants die, if at all, to see whether CFI25 is required for survival in the same way as CFI68. To address this, I did a viability experiment with three independent populations of synchronised embryos (N=50), using Oregon R wild type embryos as a control, and let them develop into adulthood, scoring for each individual that reached every life stage from embryo to L1, L2, L3, pupae and adults (Different larval stages were distinguished according to previously described anatomical features (Ashburner, 1989) (For a diagram of the *Drosophila* life cycle, see Figure 3.5 panel A). From this experiment I observed that the majority of CFI25 mutants did not reach the L2 stage (Figure 4.1). This obviously led to the next question, regarding the reasons behind the

lack of viability of second instar larva in CFI25 mutants. To address this, I again did a survival experiment with three independent populations of synchronised embryos (N=50) using Oregon R wild type embryos as a control, but this time, only the numbers of first instar larvae were measured over time (in days). By doing this, we can know at what point the lethality occurred in the mutants, i.e. if it occurs during the first instar larvae stage (L1), or just after entering the second larval instar stage (L2). What I found was quite surprising, as it could be observed that CFI25 mutants stay as L1 larvae for up to 5 days after hatching from the egg chambers, whereas wild type larvae stay as L1 larvae only for 24 hours and then moult into the second larval instar (L2), following the normal developmental program. This means that CFI25 mutants do not reach the second larval instar because they remain developmentally arrested at the first instar larval stage (Figure 4.2).

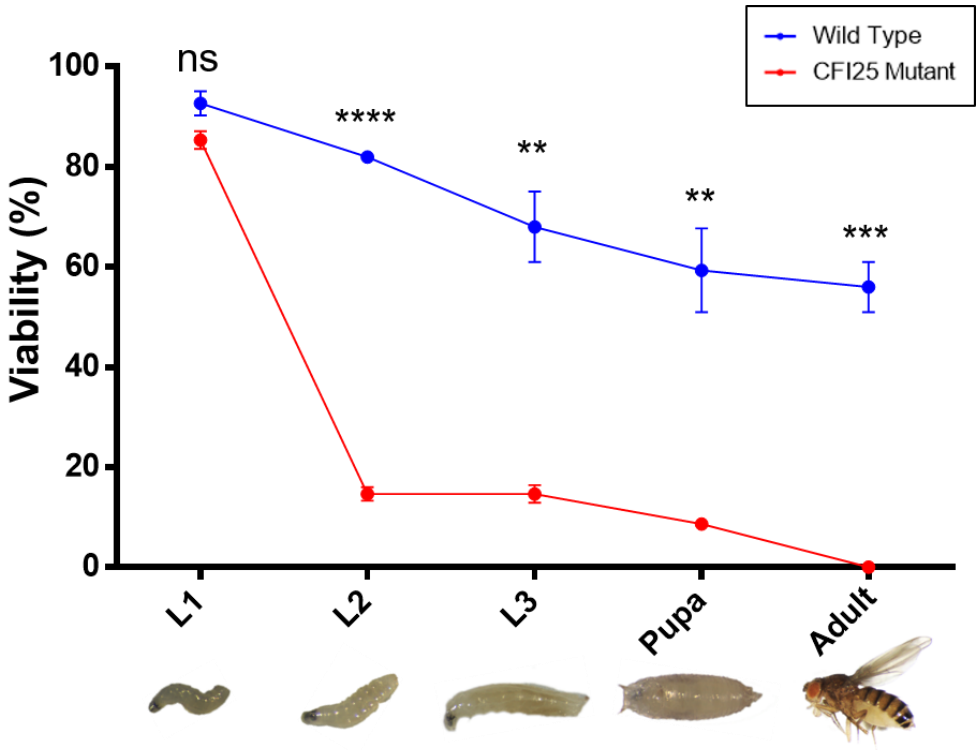


Figure 4.1 CFI25 is essential for larval viability

Survival experiment with wild types and CFI25 homozygous mutants, the values represent the mean of the percentage of individuals that reach each life stage (N = 50, biological triplicate). Note that less than 20% of the mutants reach the L2 stage. For each stage within the life cycle, a representative image is shown (From figure 3.5). Error bars represent the SEM, statistical significance was assessed with unpaired two tailed t-test per life stage (* $p < 0.05$).

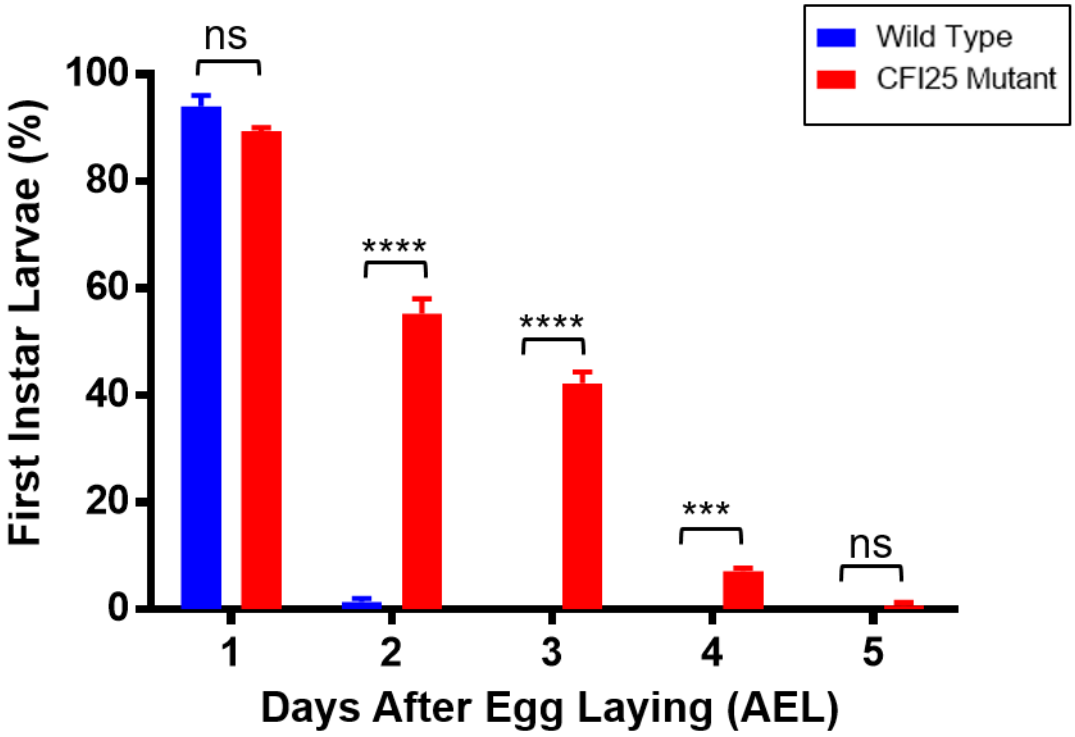


Figure 4.2 CFI25 mutants are developmentally arrested during first instar larval stage

Survival experiment with wild types and CFI25 homozygous mutants, the values represent the mean of the percentage of first instar larvae during each day after egg laying (N = 50, biological triplicate). Note that from day 2, a significant number of L1s are present for CFI25 mutants, while the wild types moult into L2s. The number of L1s for the mutant gradually decrease over time because they die at this stage. Error bars represent the SEM, statistical significance was assessed with unpaired two tailed t-test per life stage. All comparison were above sampling error. Note: n.s = Non significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

4.2.2 CFI25 mutants show a feeding phenotype

After observing that CFI25 mutants stay developmentally arrested as first instar larvae, I explored the reasons behind the observed phenotype. One possibility is that defects in hormonal signalling involved in larval transitions, for example Ecdysone, which normally pulses throughout the life-cycle to induce developmental transitions (Yamanaka *et al*. 2013), has some kind of defect, therefore impairing the ability of CFI25 mutants to moult into the L2 stage. Another possibility is that the defects in CNS formation from the mutation of CFI25 make these mutants unable to move or function at all, therefore only allowing them to hatch from an embryo as an L1 that will not develop further. What was found, unexpectedly, was that CFI25 mutants show a feeding phenotype, by which most hatched L1 individuals do not feed and therefore do not grow, which results in the delayed or arrested development of the larvae.

I decided to quantify this phenotype, and to score for feeding, I exposed freshly hatched L1s to blue yeast paste for one hour at 25ºC (See Chapter 2). This allows the food to be seen through the translucent skin of the larvae. I scored for blue food intake in CFI25 mutants and I observed that approximately 80% of the freshly hatched L1s do not show any food intake (Figure 4.3). CFI25 mutants (Allele G19200) have an insertion of a 7987 bp P{EP} element in its 5'UTR, which was generated by the Gene Disruption Project (Bellen *et al.* 2011; Bellen *et al*. 2004). I measured feeding in both heterozygous (G19200/+) and homozygous (G19200/G19200) mutants and I observed that the phenotype is significant only in the latter case, showing that this a recessive allele. As controls, I generated a

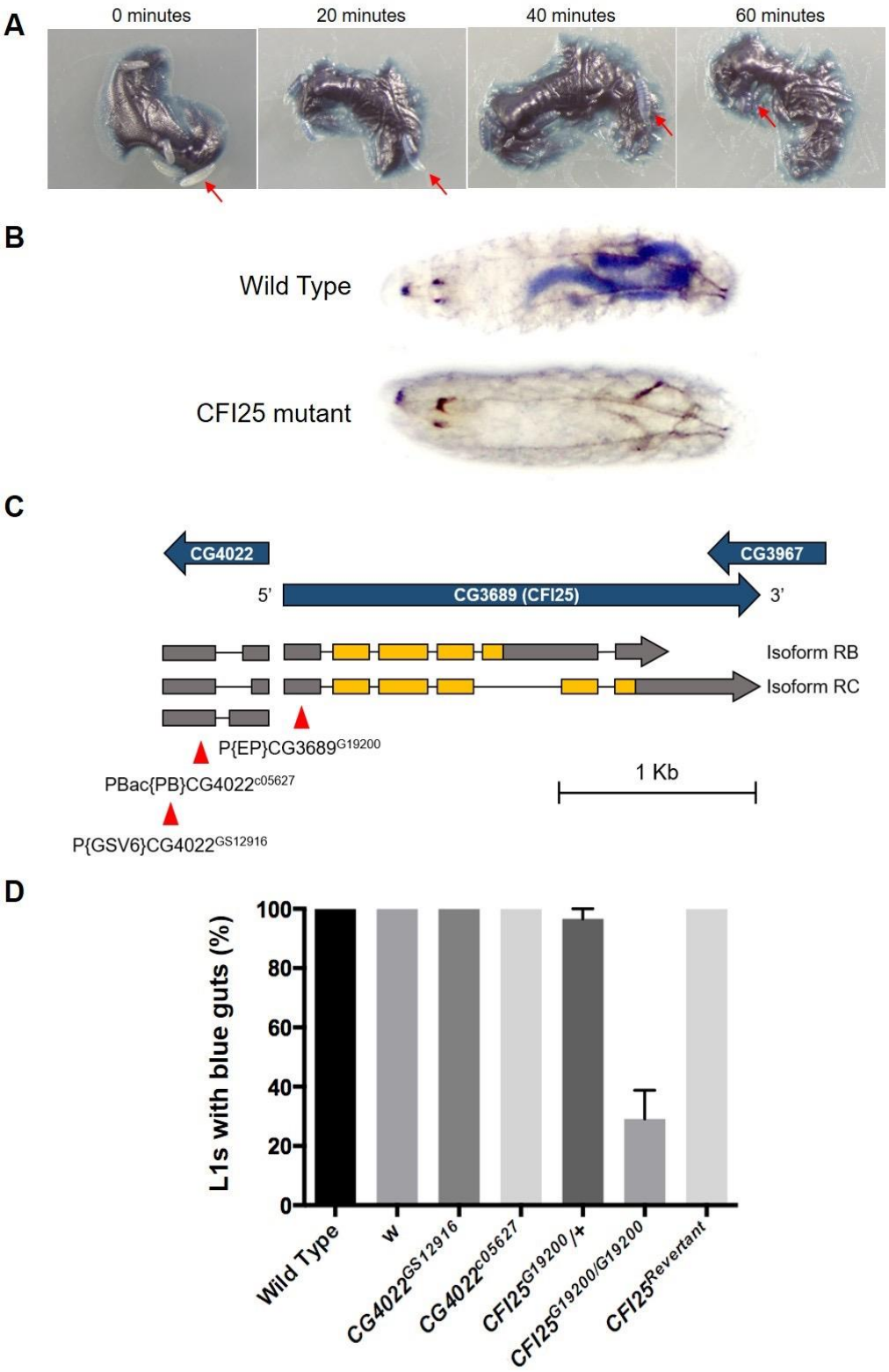


Figure 4.3 CFI25 mutants show a feeding phenotype

(Legend on the following page)

Figure 4.3 CFI25 mutants show a feeding phenotype

(A) Representative pictures of CFI25 mutant L1s during a feeding test on plates with blue yeast paste at 0, 20, 40 and 60 minutes. The red arrow in each panel indicate a larva during the feeding test. Note that the animals stay attracted to the blue yeast during the time that the experiment takes, similarly to wild types **(B)** Representative pictures of wild type and mutant first instar larva after one hour of exposure to blue yeast paste, note that CFI25 mutants do not show signs of food intake. **(C)** Gene map diagram of the location of CFI25 in the *Drosophila* genome, the grey boxes represent UTRs, the yellow boxes represent exons and the horizontal lines connecting boxes represent introns, the arrowheads represent the orientation of transcription. The location of the insertion of transposable elements used for this experiment is indicated by red triangles **(D)** Quantification of food intake in Wild Types (OR, see table 2.1), w (Genetic background of CFI25 mutation), CG4022 mutants (PBac{PB}CG4022^{c05627} and P{GSV6}CG4022^{GS12916}), CFI25 heterozygous (G19200/+) and homozygous mutants (G19200/G19200) and CFI25 revertants (N = 30, biological triplicate, error bars represent the SEM).

revertant by excision of this P{EP} element from the original CFI25 mutants by using a Δ2-3 transposase transgenic line (See Chapter 2), these revertants lost the observed feeding phenotype and behave as the wild type. In addition, the transcriptional start site of CFI25 is located 200 bp away from the transcriptional start site of CG4022, a gene with no reported molecular function (See figure 4.3). This led to the question of whether the phenotype observed is caused by transcriptional disruption of this gene instead of CFI25, given that the promoters of both genes are not characterized. To address this consideration, I used two insertional mutants affecting the 5' UTR of CG4022. The first mutant is a 7257 bp insertion from the Gene Disruption Project (Allele PBac{PB}CG4022^{c05627}) (Bellen *et al*. 2004) and the second is another 6836 bp insertion from the Gene Search System (P{GSV6}CG4022^{GS12916}) (See Chapter 2) (Toba *et al*. 1999). A feeding phenotype was not observed in these mutant stocks. These experiments show that the insertion of this P{EP} element in the 5' UTR of CFI25 mutants (Allele G19200) is specific to the cause of the observed phenotype. As mentioned earlier, the nature of this mutation in CFI25 is an insertion of a transposable element in its 5'UTR (Figure 4.3, panel C). Therefore, a complete removal of CFI25 is not necessarily expected, but requires verification. To test this, the levels of CFI25 in both wild types and mutants were quantified in different ways. First, I looked at total mRNA levels by RT-PCR during late embryogenesis, where a reduction of approximately 80% of CFI25 could be observed in the mutants when compared with the control. Secondly, by using an antibody for the human version of CFI25 (See Chapter 2), whose immunogenic region is 86% similar in *Drosophila*, I was able to detect *Drosophila* CFI25 and observe a similar trend in Western Blot experiments. Lastly, by using the same *in situ* RNA probes used to

study the expression pattern of CFI25 in embryos (See Chapter 2 and Figure 3.10), I observe that the mutants show a substantial reduction of signal in the ventral nerve cord in late embryos (Figure 4.4). I had previously tried to use this same antibody for immunofluorescence experiments against CFI25 in embryos, given the lack of antibodies for the *Drosophila* CFI members, but unfortunately the antibody did not show any signal (Data not shown). This is likely to be caused by this antibody recognising only the denatured protein, as used in WB experiments (Chapter 2).

To address why CFI25 mutants are not able to feed at L1 stage, we first need to understand whether this is a morphological or behavioural defect. I first analysed the phenotype in more detail by examining the larval cephalopharyngeal skeleton (CPS). These structures are a group of connected sclerites of the anterior embryonic/larval digestive system and are used for feeding. The main CPS components are the mouth hooks, the H-piece and the cephalopharyngeal plates. The mouth hooks are two movable structures situated at the tip of the skeleton and used to shovel food into the pharynx, which are connected to the rest of the CPS by bilaterally symmetric processes. The H-piece is situated between the mouth hooks and the cephalopharyngeal plates on the ventral side and together with smaller components including the median tooth, act as a bridge to give rigidity to the CPS. The cephalopharyngeal plates are simpler in structure, showing both a ventral and dorsal division, which attach the CPS to the rest of the structures in the larval head. Their formation during embryogenesis is achieved during late stages in the gnathal segments, more specifically, the mandibular segment, where its structures are shaped by

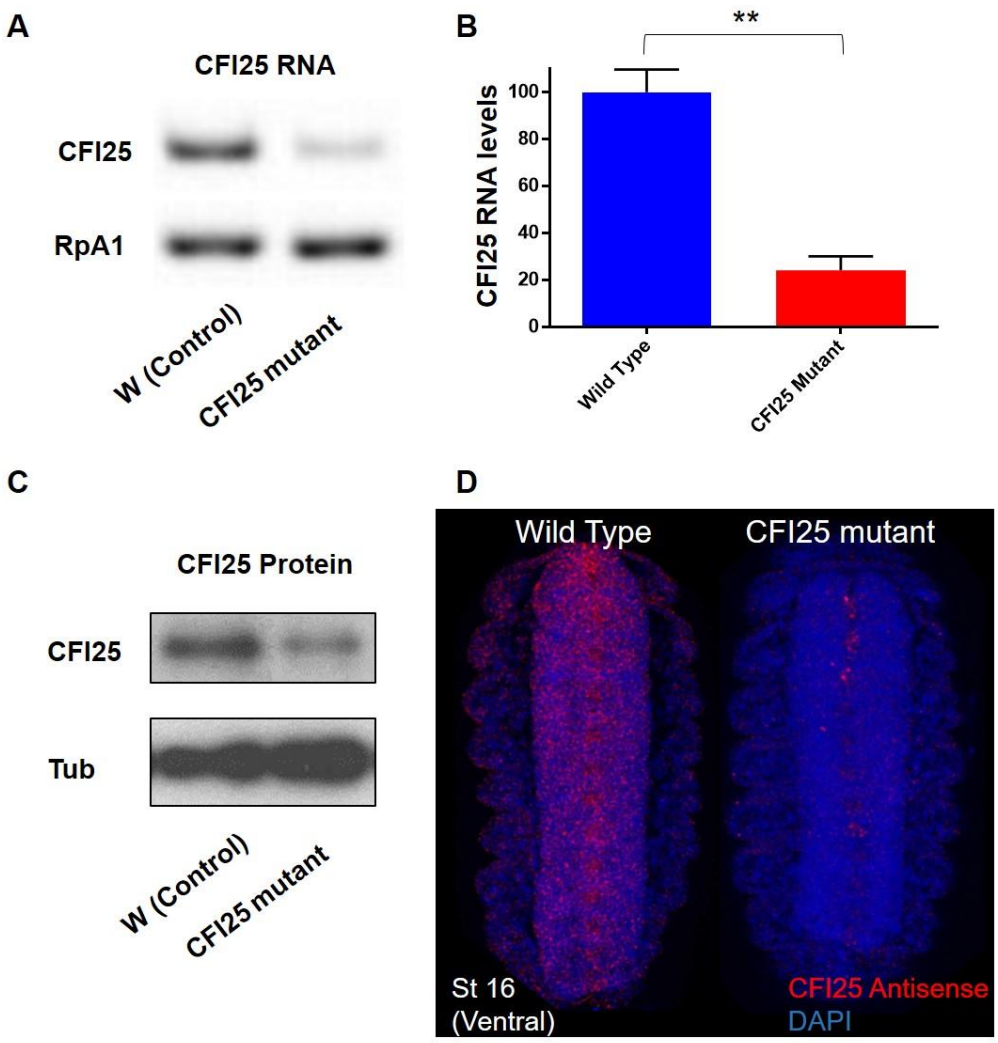


Figure 4.4 CFI25 mutants show reduced levels of CFI25 RNA, as well as protein and a reduction of expression from the ventral nerve cord

(Legend on the following page)

Figure 4.4 CFI25 mutants show reduced levels of CFI25 RNA, as well as protein and a reduction of expression from the ventral nerve cord

(A-B) Representative bands from an RT-PCR experiment to measure CFI25 mRNA levels in stage 16 embryos using RpA1 as a loading control (See Chapter 2). **(B)** Quantification of **(A)** (Biological triplicate), mRNA levels are shown as percentage from the control. Note that CFI25 mutants show a significant reduction in mRNA levels. An unpaired two-tailed t-test was used to compare genotypes, * $p < 0.05$. **(C)** Western Blot experiment to measure CFI25 protein levels in stage 16 embryos using tubulin as a loading control (See Chapter 2). The same trend as in **(A)** is observed, this experiment was also performed in three independent biological replicates. **(D)** *In situ* hybridisation experiment for CFI25 in wild types and mutants, note that the mutants have a reduction in the signal coming from the ventral nerve cord in stage 16 embryos. DAPI was used to label the nuclei. The confirmation of the presence, location and orientation of the P{EP} element in CFI25 mutants was done by PCR (Data not shown).

secretion of cuticle by the cells lining the atrium, pharynx and frontal sac (Campos-Ortega and Hartenstein, 2013). The CPS is connected to specialized muscles involved in feeding, among them the cibarial dilator muscles that lift the roof of the pharynx and thereby widen its lumen, allowing food to be ingested (Schoofs *et al*. 2010; Jürgens *et al*. 1986).

I observed that CFI25 mutants have anatomical defects in their mouth hooks, more specifically, their CPS shows a “collapsed” phenotype in which the dorsal and ventral processes of the cephalopharyngeal plates point inwards. As a result of this, the angle between the median tooth and the caudal processes is significantly larger in CFI25 mutants (Figure 4.5). This phenotype shows a similar penetrance as that of feeding (Nearly 80%, N = 30) and can explain in part the feeding defects observed in CFI25 mutants. Further, there are cases in the literature where mutated genes that control the fate of feeding structures from the mandibular and maxillary segments can also have an impact on feeding. For example, the Hox gene *deformed* (*Dfd*) precisely controls patterning in these segments. Its name comes from the deformed cephalic phenotype observed in mutants, reflecting failures during the process of head involution (Merrill and Turner 1987). Interestingly, a recent study showed that *Dfd* is required for the establishment and maintenance of the neuromuscular unit required for feeding, by specification of synapses and stage-specific sets of target genes during larval stages (Friedrich *et al*. 2016). Thus, the phenotype observed in CFI25 mutants could also be a reflection on effects on target genes with neural roles.

I examined at what point during the larval feeding cycle CFI25 mutants were not able to ingest food. To address this, I exposed them for one hour with yeast paste

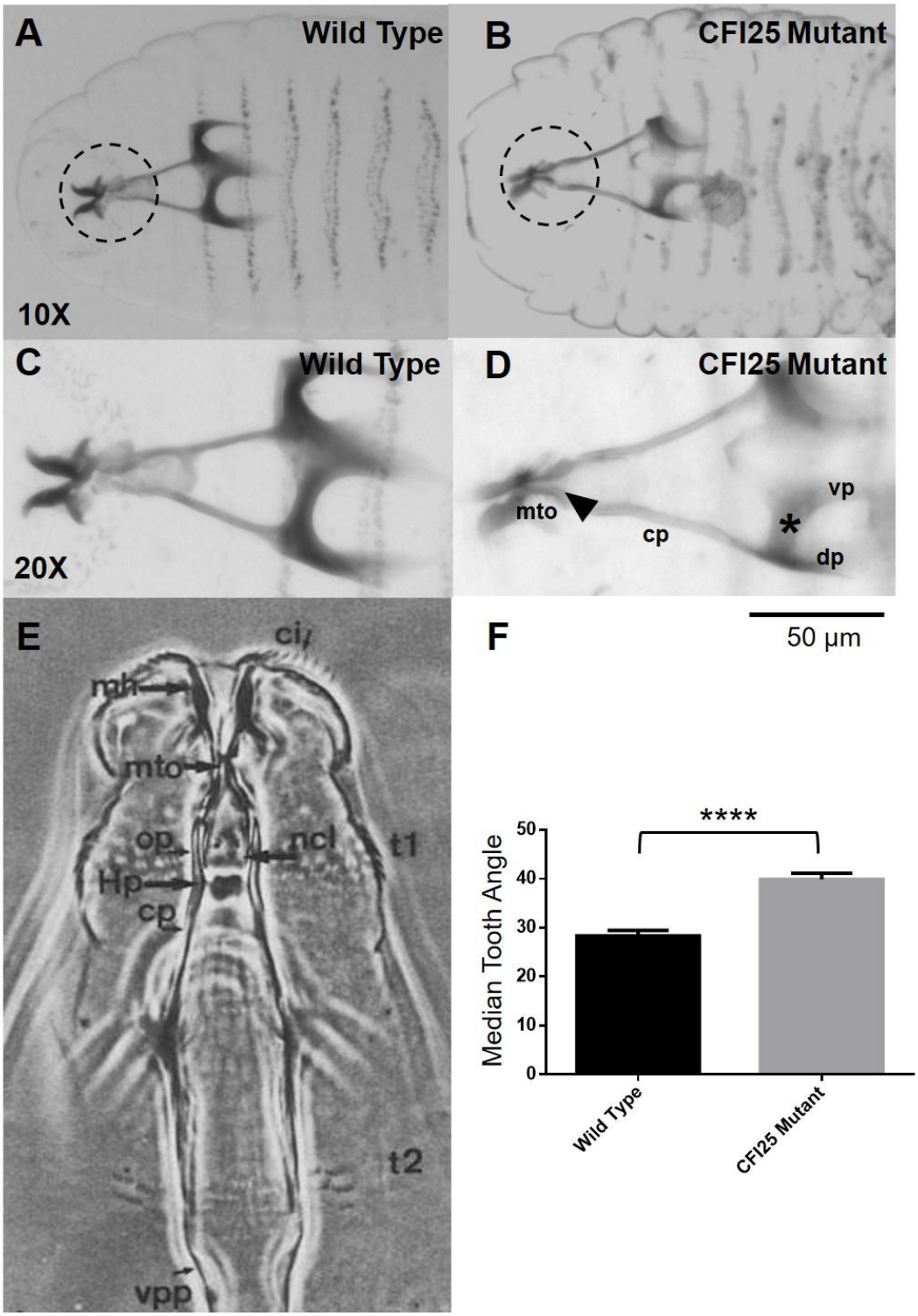


Figure 4.5 CFI25 mutants show anatomical defects in their cephalopharyngeal skeleton

(Legend on the following page)

Figure 4.5 CFI25 mutants show anatomical defects in their cephalopharyngeal skeleton

(A-B) Ventral view of anterior structures of first instar larvae in Wild Type **(A and C)** and CFI25 mutant **(B and D)**, anterior is to the left. **(C)** Magnification of the CPS from the wild type **(A)** to show its normal anatomy during this stage. The dotted circle represents the angle between the median tooth and the caudal processes of the cephalopharyngeal plates. **(D)** Magnification of the CPS from CFI25 mutants **(B)**, note that the dorsal and ventral processes of the cephalopharyngeal plates point inwards (marked with an asterisk) and also the angle between the median tooth and the caudal processes is larger, denoted by an arrowhead. **(E)** Structures of the CPS in the larval head from (Campos-Ortega and Hartenstein, 2013) for reference. **(F)** Quantification of the angle between the median tooth and the caudal processes in wild types and mutants (N = 20). An unpaired two-tailed t-test was used to compare genotypes, * $p < 0.05$. The scale bar represents 50 µm. mto = Median tooth, Hp = H-Piece, cp = Caudal process of the cephalopharyngeal plates, vp = Ventral process of the cephalopharyngeal plates, ci = Maxillary cirri, mh = Mouth Hook, ncl = neck clasps, op = oral process of the H-piece

mixed with DAPI (See Chapter 2), which has been shown to label the nuclei of not only fixed, but also living cells (Martin *et al*. 2005). The rationale of this experiment is that as the DAPI-stained yeast moves though the digestive system, the cells that come in contact with it will become labelled, as well as yeast cells themselves (Krol *et al*. 2003). Then, after the hour of exposure, I look at the extent of DAPI staining. It would be expected in the wild type controls, to have labelled the digestive system at least until the midgut. Compared to the wild type larvae, it was observed that CFI25 mutants are not able to pump food into their midgut, and the yeast paste stays trapped in their pharynx (Figure 4.6).

In summary, these results show that CFI25 mutants are developmentally arrested at the L1 stage, the reasons behind which is a feeding phenotype, which prevents the onset of the larval transition into L2 stage. These defective CFI25 mutants have also been shown to have lower levels of CFI25 at the mRNA and protein level, which can now be considered the reason behind the phenotype observed. Lastly, CFI25 mutants also display defects in the anatomy of their CPS, as well as problems in the movement of food from the pharynx to the midgut, relating the phenotype to a morphological defect in these larvae. In the next section we will analyse other important aspects of feeding in CFI25 mutants, such as the nervous system and behaviour.

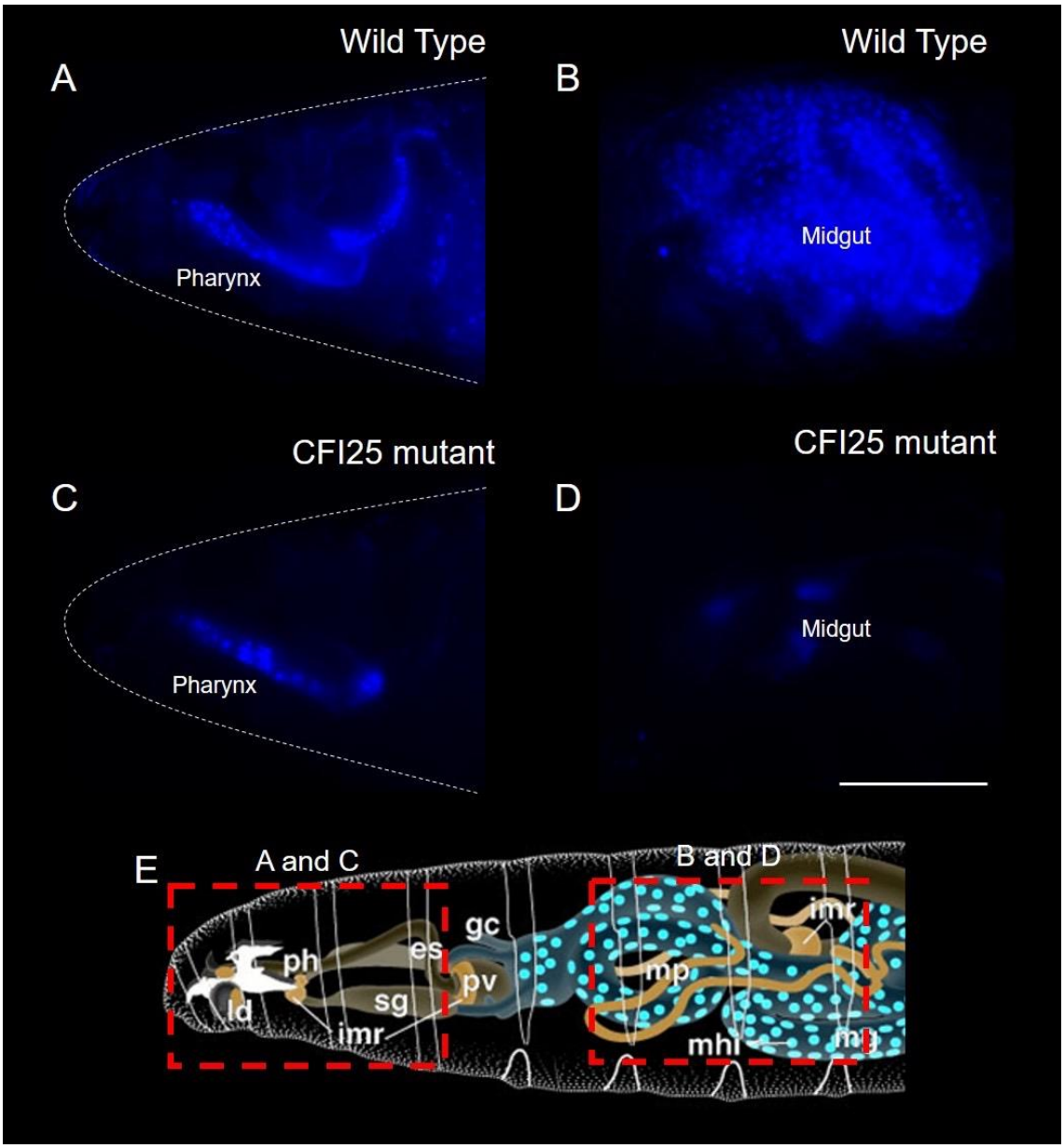


Figure 4.6 CFI25 mutants do not pump food into their midgut

(Legend on the following page)

Figure 4.6 CFI25 mutants do not pump food into their midgut

(A-D) Lateral view of anterior structures of first instar larvae in Wild Type **(A and C)** and CFI25 mutants **(B and D)**, anterior is to the left. **(A)** After one hour of exposure to DAPI-stained yeast paste, the wild types show their anterior portion of the digestive system, known as the pharynx, completely labelled. **(B)** The same is observed for midgut in the wild types. A white dotted line is used to show the outline of the larvae. **(C)** Similarly, CFI25 mutants also show stained cells in their pharynx. However, **(D)** their midgut is not labelled and shows only background signal also observed in Non-DAPI control wild types used for auto fluorescence (Data not shown). These results show that CFI25 mutants do not pump food into their midgut and instead it gets trapped in the pharynx. Panels **(A-B)**, as well as panels **(C-D)**, were taken from the same individual. **(E)** Diagram of the larval digestive system (modified from Hartenstein, 1993), the two red frames represent the regions of interest shown in panels **(A-D)**. The scale bar represents 50 µm. Relevant structures are ph = Pharynx, es = Esophagus, pv = Proventriculus, mg = Midgut.

4.2.3 CFI25 mutants do not show locomotory differences or major defects in the anatomy of the nervous system when compared with wild types

In Chapter 3 the expression pattern of CFI25 in the late embryo was analysed and it was observed that it shows an enrichment in the nervous system (Fig 3.10). In the case of CFI25 mutants, a reduction of CFI25 expression can be observed in this tissue (Figure 4.6). In order to test whether the phenotype is related to the function of the nervous system, I performed behavioural experiments by measuring three different parameters related with locomotory behaviour and feeding. First, a common locomotory parameter was analysed, which is called the frequency of forward peristaltic waves per minute. By determining this, I can identify if CFI25 mutants move less and generally do not explore the substrate on which they are growing. The neural circuitry controlling this behaviour has been studied (Berni *et al.* 2012; Fushiki *et al*. 2016) and can be related to other neural genes that also show a feeding phenotype when mutated, such as *lov*, which has been described as presenting this phenotype as a consequence of locomotor defects (Bjorum *et al*. 2013).The results of this analysis showed that there is no significant difference in the frequency of forward peristaltic waves per minute between CFI25 mutants and wild type larvae (Fig 4.7).

The remaining behavioural parameters I measured can be related between them. First I measured the fraction of the time of each recording (2 minutes) that the individual spent performing mouth hook movements. Then, given that the larvae display this activity in bouts, I measured how many bouts of mouth hook movements they did in each recording. These parameters allow us to detect how often the larvae decide to engage in the food intake motor program. Recurrent motor programs such as locomotor or feeding programs are controlled by central

pattern generators in the central nervous system. It has been shown that neuronal activation of only 20 neurons in the *Drosophila* larval brain, known as the “hugin” neurons, can suppress the motor program for food intake while induce the program for locomotion (Schoofs *et al*. 2014). Hugin neurons have that name because they express this neuropeptide which is a homolog of mammalian neuromedin U.

After measuring the fraction of time spent eating and the amount of eating bouts in 2 minutes, I also did not see any difference between CFI25 mutants and wildtypes (Figure 4.7). These results suggest that it is unlikely that CFI25 mutants show this feeding phenotype because of defects in general movements involved in this behaviour. The central pattern generator controlling *Drosophila* feeding has been shown to be located in the subesophageal zone, and through lesion experiments, there is evidence that all the feeding rhythmic activity does not require the brain hemispheres or the ventral nerve cord for its function (Huckesfeld *et al*. 2015).

Because the expression pattern of CFI25 during late embryogenesis shows an enrichment in the nervous system, I asked whether CFI25 mutants show abnormalities in the overall anatomy of their nervous system. To analyse and quantify this, I stained both the central and the peripheral nervous system in late embryos by labelling elav and futsch, respectively. (See Chapter 2).

Elav is an RNA binding protein that is widely used a pan-neuronal marker. Also, mutations in elav can affect the anatomy of the nervous system

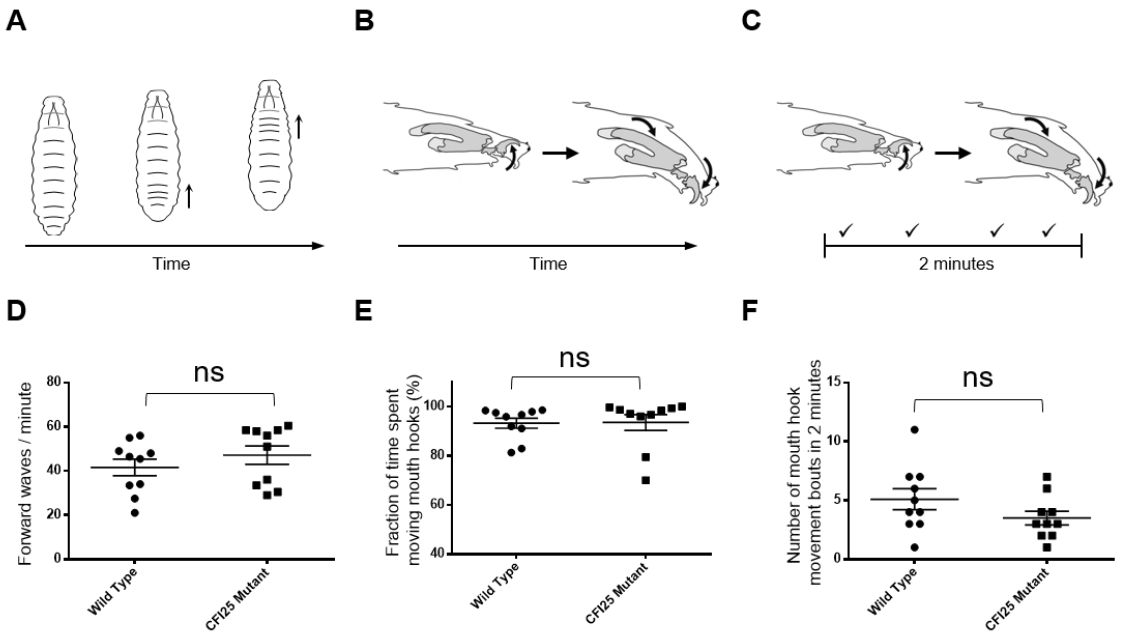


Figure 4.7 CFI25 mutants do not show behavioural differences related to feeding when compared with wild types

(A-C) Diagrams representing the parameters measured in this experiment, **(A)** diagram of the forward peristaltic waves per minute, **(B)** Diagram of the fraction of time spent performing mouth hook movements. **(C)** Diagram of the number of mouth hook movement bouts in 2 minutes. **(D-F)** quantification of the parameters shown in **(A-C)** in CFI25 mutants and wild types. None of the parameters measured shows a significant difference between CFI25 mutants and wild types (N = 10). A non-parametric two-sided Mann-Whitney test was used to test for statistical significance, * $P < 0.05$. Larvae diagrams from panel A were kindly provided by Dr. Joao Osorio and were modified from (Picao-Osorio *et al*. 2015). Larvae diagrams from panels B and C were modified from (Schoofs *et al*. 2009).

(Zaharieva *et al*. 2015). Elav also has important roles for RNA processing and APA for several genes, including the Hox genes (Rogulja-Ortmann *et al*. 2014). Futsh is a protein that associates with microtubules and is necessary for dendritic and axonal growth (Hummel *et al*. 2000), this protein is detected by the monoclonal antibody 22C10 (See Chapter 2). This antibody is widely used to visualize the anatomy of the peripheral nervous system. After staining late *Drosophila* embryos with these markers and comparing wild types with CFI25 mutants, I did not see any major differences between these genotypes (Figure 4.8). When considering the results on locomotory behaviour, these anatomical analysis suggests that although the expression pattern of CFI25 shows an enrichment in the nervous system, neither locomotory behaviour nor major anatomical defects can be tied to the nervous system in these mutants.

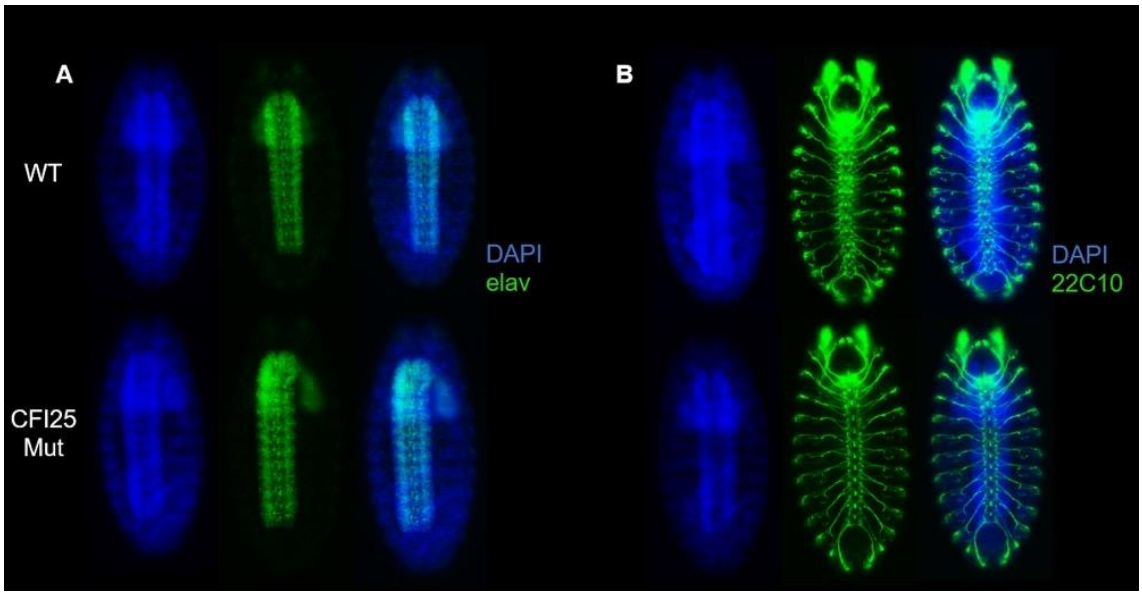


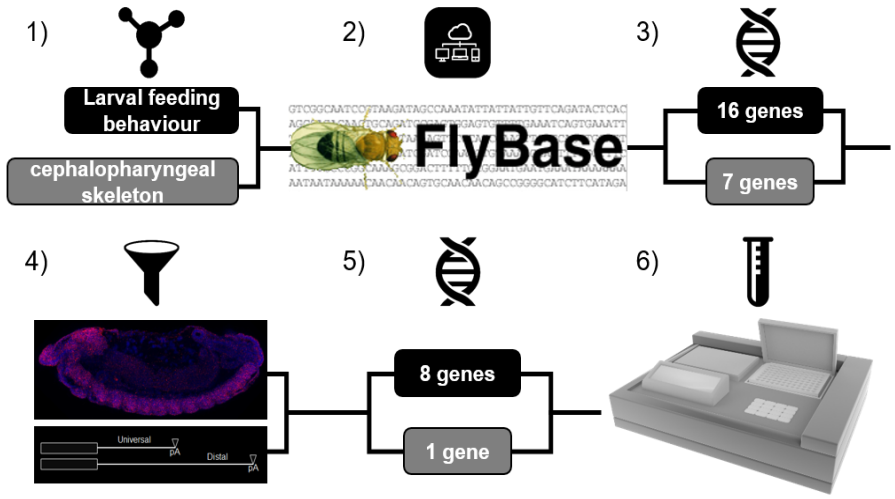
Figure 4.8 CFI25 mutants do not show major defects in the anatomy of the central and peripheral nervous system when compared with wild types

(A) Representative embryos from wild types (top row) and CFI25 mutants (Bottom row) stained for elav to show their CNS, as well as DAPI, to label the nuclei and show the contour of the embryos. **(B)** Representative embryos from wild types (top row) and CFI25 mutants (Bottom row) stained for Futch (22C10) to show their PNS and also DAPI. Note that there are no major visible defects in the anatomy of these structures between genotypes.

4.2.4 Identification of targets downstream of CFI25 and the feeding machinery

CFI25 is a CPA factor that has been shown to control APA *in vitro* (Masamha *et al*. 2014; Kim *et al*. 2010; Masamha *et al*. 2016; Li *et al*. 2015). As previously discussed, the crystal structure of CFI25 suggests that when it binds to the UGUA motifs in an antiparallel fashion, the CFI complex is able to control APA by looping the pre-mRNA molecule (Yang *et al*. 2011). From these data, it can be hypothesised that the phenotype observed in CFI25 mutants is caused by alterations in the cleavage and polyadenylation patterns of one or more target genes at the molecular level. To test this hypothesis, I looked for potential target genes using a simple bioinformatic pipeline (Figure 4.9). Firstly, two "Gene Ontology" (GO) terms (Consortium 2000; Blake *et al.* 2015) were used to scan for target genes, "Larval Feeding Behaviour" and "Cephalopharyngeal skeleton". These were selected because of their connection with the phenotype observed. Secondly, these GO terms were used as input in the *Drosophila* database "Flybase" (Attrill *et al.* 2016) and as a result, 16 genes were detected labelled by the GO term "Larval Feeding Behaviour" and seven genes with "Cephalopharyngeal skeleton". This list of genes was then filtered by the requirement of two conditions, the first was if the genes showed neural expression, given the expression pattern of CFI25 (Figure 3.10) and the second was if the genes showed more than one APA isoform, because of the known molecular function of CFI25. As a result, eight genes remained labelled with "Larval Feeding Behaviour" and one gene with "Cephalopharyngeal skeleton".

A



B

Name	Symbol	APA	Molecular Function	Biological Function	Neural Expression
foraging	for	Yes	cGMP-dependent protein kinase	larval locomotory behavior	Yes
Inositol 1,4,5,-tris-phosphate receptor	Itp-r83A	Yes	Calcium channel	Fatty acid homeostasis, response to starvation	Yes
klumpfuss	klu	Yes	Zinc finger transcription factor	Neuroblast development, sensory organ precursor cell fate determination	Yes
jim lovell	lov	Yes	DNA binding	Gravitaxis, larval walking behavior	Yes
Neuropeptide F receptor	NPFR	Yes	G protein-coupled receptor	Neuropeptide signalling pathway	Yes
Ran-binding protein M	RanBPM	Yes	Ran GTPase binding	Cytoskeleton organization, ovarian follicle cell-cell adhesion	Yes
RPS6-p70-protein kinase	S6k	Yes	Ribosomal protein S6 kinase	Axon guidance, dendrite morphogenesis	Yes
shibire	shi	Yes	Microtubule motor	Regulation of synapse structure and activity	Yes
kurtz	krz	Yes	Protein binding	Locomotory exploration behaviour, cuticle development	Yes

Figure 4.9 Bioinformatics pipeline to find candidate target genes downstream of CFI25

(Legend on the following page)

Figure 4.9 Bioinformatic pipeline to find candidate target genes downstream of CFI25

(A) Diagram of the bioinformatics pipeline used to identify candidate target genes to be studied in CFI25 mutants. (1) Two “Gene Ontology” (GO) terms (Consortium 2000; Blake *et al.* 2015) were used to find target genes, “Larval Feeding Behaviour” and “Cephalopharyngeal skeleton”. These were selected because of their connection with the phenotype observed. (2) The selected GO terms were used as input in the *Drosophila* database “Flybase” (Attrill *et al*. 2016), (3) 16 genes were detected labelled by the GO term “Larval Feeding Behaviour” and seven genes with “Cephalopharyngeal skeleton”. (4) This list of genes was filtered by two conditions, the first condition was if the genes showed neural expression because of the expression pattern of CFI25 and the second condition was if the genes showed more than one APA isoform, because of the known molecular function of CFI25. (5) After filtering, eight genes remained labelled with “Larval Feeding Behaviour” and one gene with “Cephalopharyngeal skeleton”. (6) These genes were selected for molecular analysis in CFI25 mutants. **(B)** Table of genes obtained by this pipeline, the first seven genes are the ones labelled by “Larval Feeding Behaviour” and the last one (*kurtz*) is the one labelled by “Cephalopharyngeal skeleton”. A brief summary of their molecular and biological function is shown.

Lastly, these genes were selected for molecular analysis in CFI25 mutants by assessing their patterns of APA by RT-PCR experiments.

“Foraging” (*for*) encodes a cGMP-dependent protein kinase. As its name indicates, this gene was associated with larval foraging behaviour. Two natural alleles from the wild were described as “sitters” and “rovers”. “Sitters” cover a small area while engaged in feeding, whereas “rovers” traverse a large area while feeding. Interestingly, the frequency at which these alleles were observed in the wild was constant (Sokolowski 1980; Pereira and Sokolowski 1993). Later the expression pattern of *for* was described as showing neural expression, as well as expression in the larval proventriculus. In this same study, it was shown that overexpression of *for* in the adult nervous system increases sucrose responsiveness in sitters, which was known to be naturally higher in rovers, highlighting the roles of *for* in food-related behaviour (Belay *et al*. 2007). The broad expression pattern of *for* in the nervous system makes it an interesting candidate to be explored in CFI25 mutants. Nonetheless, CFI25 mutants do not show locomotory differences with wild types in terms of hyper o hypo activity, as assessed by their frequency of forward waves per minute, time spent moving the mouth hooks or activity bouts (Figure 4.7). Making them potentially less likely to show this phenotype because of behavioural genes such as *for*.

Another candidate, “*Inositol 1,4,5,-tris-phosphate receptor*” (*Itp-r83A*), belongs to the InsP3 receptor family; it is a calcium channel that releases calcium from intracellular stores in response to extracellular signals. It has been shown that mutants for *Itp-r83A* show feeding defects and are smaller than the wildtypes. Interestingly, expression of *Itp-r83A* in the insulin producing cells (IPCs) in the larval brain of mutants is sufficient to restore the feeding deficit. However, despite

the fact that ubiquitous RNAi knockdown of *Itp-r83A* can mimic the feeding defect, specific knockdown in IPCs does not trigger the phenotype. There is a model proposed in which *Itp-r83A* activity in non-overlapping neuronal domains independently rescues larval phenotypes by non-cell autonomous mechanisms (Agrawal *et al*. 2009). The expression pattern of *Itp-r83A* has been mapped to the adult brain, peripheral sense organs, the mesoderm and muscle precursors in the embryo (Raghu and Hasan 1995). A few years later it was also described to show expression in the larval ring gland, an endocrine organ in the larva that secretes ecdysone, the hormone that triggers moulting and metamorphosis (Yamanaka *et al*. 2013). Accordingly, various alleles of *Itp-r83A* showed a delayed larval moulting phenotype (Venkatesh and Hasan 1997). Although broader than just confined to the nervous system, *Itp-r83A* is also a good candidate to be studied in CFI25 mutants.

“Klumpfuss” (*klu*) is a zinc finger transcription factor whose relation with larval feeding was discovered in screen for larvae defective in this behaviour. Microarray analysis indicated that *klu* mutants showed altered expression of the neuropeptide *Hugin (hug)* (Melcher and Pankratz 2005). *Hugin* is expressed in a cluster of 20 neurons which can suppress the motor program for food intake while inducing the program for locomotion in larvae (Schoofs *et al*. 2014). *Hug* expression is also regulated by food signals and neuroanatomy of the *hug*-expressing neurons shows that they project axons to several components of the feeding machinery, such as the pharyngeal muscles, as well as higher brain centers and neuroendocrine organs related to feeding. In addition, the dendrites of hug neurons are innervated by gustatory receptor-expressing neurons and also by internal pharyngeal chemosensory organs (Melcher and Pankratz 2005).

Klu has also been shown to have an important role in larval neuroblast (NB) self-renewal, as it not only functions as a NB-specific transcription factor, but loss of *klu* causes premature differentiation of NBs. Overexpression of *klu* generates the formation of transplantable brain tumours (Berger *et al.* 2012). These observations reveal that *Klu* plays key roles in the specification of the nervous system by controlling gene expression and also acts as a transcription factor. These characteristics make *Klu* a good candidate that can connect CFI25 function with the observed phenotype.

“Jim Lovell” (*lov*) is another transcription factor of the BTB/POZ family, identified in a forward genetic screen for animals with deficient responses to gravitaxis (Armstrong *et al*. 2006). Its name is derived from that of an astronaut who made pioneering work in microgravity. In another study, one *lov* allele was found to show locomotor defects as well as feeding defects in larvae. The authors showed that *lov* mutants are sluggish when compared to wildtypes and also show a higher frequency of spontaneous backward movements, which in wild types are only commonly seen when the individuals have contact with obstacles. As a result of problems in motor control, *lov* mutants show a feeding phenotype because of a slower rate in mouth hook movement, which results in impairment of the shovelling movements used to get food into their digestive system. As expected, *lov* starts being expressed during embryogenesis in nuclei of the developing nervous system. During late stages of embryogenesis, *lov* shows expression in nuclei of both the central and the peripheral nervous systems (Bjorum *et al*. 2013). Although its expression pattern resembles that of CFI25 (Figure 3.10), the locomotory phenotypes observed in *lov* mutants are not observed in CFI25

mutants (Figure 4.7), suggesting that if *lov* is affected molecularly in CFI25 mutants, it may not necessarily reflect the effects reported in lov mutations.

"*Neuropeptide F Receptor*" (*NPFR*), as its name indicates, is the receptor of the *"short neuropeptide F"* (*sNPF*), which is structurally related to vertebrate regulatory peptides of the neuropeptide Y family. The expression levels of *sNPF* in the larval brain are correlated with the larvae's attraction to food and engagement in feeding behaviour. Its knockdown in the nervous system can lead to a decrease in food intake in both larvae and adults (Lee *et al*. 2004; Brown *et al*. 1999; Wu *et al*. 2003). *NPFR* is a G-protein coupled receptor whose relation with larval feeding comes from a study which showed that when *NPFR* is overexpressed in its own expression domain, both fed and fasted larvae show similar intake rates of liquid food. When *NPFR*-overexpressed larvae are forced to feed on solid food, which is less preferred under normal conditions, fed experimental larvae exhibited significant intake of the solid food when compared with fed controls (Wu *et al*. 2005). Also, the knockdown of *NPFR* by RNAi in its expression domain, or in the whole nervous system, leads to larvae that are deficient in motivated feeding of solid but not liquid food. These results show that *NPFR* is involved in food selection driven by hunger in *Drosophila* (Wu *et al*. 2005). Thus, expression changes in *NPFR* caused by aberrant APA patterns in CFI25 mutants could in part explain why these mutants do not feed, even though they perform all the required movements (Figure 4.7). Experiments on food source selection will be required to assess if CFI25 mutants show differences in food preference and thus allowing us to relate it with *NPFR* if affected.

"Ran-binding protein M" (*RanBPM*) is the *Drosophila* orthologue of the vertebrate "Ran Binding Protein in the Microtubule organizing center" genes. Its similarity

with the human *RANBP9* gene suggests that *RanBPM* binds to *RAN GTPase*. In *Drosophila*, its biological function has been implicated in the regulation of germ line stem cell niche organization in the ovary as well as microtubule dynamics (Dansereau and Lasko 2008). The relation of this gene with larval feeding comes from a study in which an insertion of a P-element in the second exon of *RanBPM* makes the larvae show a feeding defect and impairs locomotion and response to light in larvae. These phenotypes are rescued after excision of this P-element, as in the case of CFI25 mutants. The expression pattern of *RanBPM* show that is expressed in the larval mushroom bodies, showing high expression in Kenyon cells (Scantlebury *et al*. 2010). The mushroom bodies are centers associated with olfaction, learning and memory in the central nervous system which are originated from only four neuroblasts per brain hemisphere during embryogenesis (Kunz *et al*. 2012; Kurusu *et al*. 2002). In addition, the mushroom bodies have an important role for feeding via the insulin signalling pathway (Zhao and Campos 2012). Although the molecular mechanism that links *RanBPM* with larval feeding has not been established, it is proposed that *sNPF* may be a target of *RanBPM* in neurons of the mushroom bodies, among others (Scantlebury *et al*. 2010). Thus, *RanBPM* could also be affected in CFI25 mutants and therefore affect feeding by gene expression changes through APA in the mushroom bodies.

"Ribosomal protein S6 kinase" (*S6k*), as its names indicates, is a serine/threonine kinase whose mammalian orthologue is $p70^{S6k}$. *S6k* was first described in 1996 as being 78% similar at the protein level in the catalytic domain with its human counterpart. *Drosophila* S6k transfected in mammalian cells is able to phosphorylate the mammalian ribosomal protein S6 (RP6), which is a component of the 40S ribosomal subunit in eukaryotes (Watson *et al*. 1996; Erikson 1991).

The relation of *S6k* with larval feeding comes from the same study that covered the relation of *NPFR* with feeding (Wu *et al*. 2005), in this case, the authors show that manipulating *S6k* expression in larval insulin producing cells (IPC) can have an effect on larval feeding. More specifically, overexpression of S6k in IPCs leads to attenuated hunger response in fasted larvae and opposingly, downregulation of *S6k* by RNAi in the same cells triggers fed larvae to display motivated foraging and feeding. These results suggest that while *NPFR* can mediate food source selection, *S6k* within IPCs can mediate hunger regulation and response to food. The relation of *S6k* with the insulin pathway has also been studied in cultured cells and it has been shown that *S6k* is directly activated and phosphorylated by insulin through phosphoinositide 3-kinase (PI3K) and protein kinase B (PKB), as well as *Drosophila* target of rapamycin (dTOR) (Lizcano *et al*. 2003). These observations suggest that *S6k* could also connect the observed phenotype in CFI25 through changes in expression in IPCs. Although the exact expression pattern of S6k is not reported, meaning that it could be expressed in other tissues.

"Shibire" (*shi*) is the *Drosophila* orthologue of dynamin, a microtubule-activated GTPase involved in endocytosis and synaptic vesicle recycling. Its expression pattern is broad, though it shows an enrichment in both the central and peripheral nervous system during different life stages. In *Drosophila* there are several alleles of this gene that are temperature-sensitive, which means that at a restrictive temperature, there is an arrest in the recycling of synaptic vesicles at the neuromuscular junction, leading the individuals to be paralysed. In fact, the word "shibire" means "paralysed" in Japanese (Chen *et al*. 1992; Siddiqi and Benzer 1976). This interesting phenomenon of disrupting synaptic transmission at a specific temperature, led to the development of genetic tools, by which

temperature-sensitive alleles of *shi* can be expressed using the Gal4/UAS system in a targeted fashion to induce expression in specific cells and tissues at specific times (Brand and Perrimon 1993). For example, by using this approach, synaptic transmission was blocked in neurons from the mushroom bodies in *Drosophila* to show that synaptic transmission is required for memory retrieval but not during acquisition or storage (Dubnau *et al*. 2001).

The relation of *shi* with larval feeding behaviour comes from the same study that points to both *S6k* and *NPFR* as being involved in feeding (Wu *et al*. 2005). In this case the temperature-sensitive alleles were expressed in the *NPFR* expression domain. When the restrictive temperature was applied, fasted experimental larvae showed an attenuated feeding response to solid but not liquid food. While at the permissive temperature, both fasted control and experimental larvae displayed normal feeding responses to both liquid and solid foods. These results suggest that the *sNPF*/*NPFR* neuronal pathway can influence the intensity and the duration of the feeding response in *Drosophila* larvae and thus, so can *shi*. What is more, because of its relationship with *NPFR, shi* could also explain in part the phenotype observed in CFI25 mutants. Although its expression pattern is broad and this will make difficult to point what specific tissues are responsible for the phenotype observed in CFI25 mutants.

"*kurtz*" (*krz*) is a β-arrestin (non-visual arrestin); these proteins are involved in the desensitization and endocytic internalization of G protein coupled receptors (Lefkowitz 2005). During late embryogenesis, *krz* is expressed in the central nervous system, maxillary cirri, the pharynx and antennal sensory organs (Roman and Davis 2000). These anterior structures are related to the cephalopharyngeal skeleton, which is why this gene appeared in our pipeline as

it can be found under this GO term. An allele of *krz* generated by insertion of a P-element (*krz*^{1}) was shown to generate dark melanotic tumours primarily in the fat body of third instar larvae. This gene is named kurtz after the character in Joseph Conrad's "The Heart of Darkness" alluding to the dark melanotic tumours seen in the larvae by the authors (Roman *et al*. 2000). Although this gene has not been involved directly in larval feeding, its relevance for the development of the CPS made me take it into account to be tested in CFI25 mutants (See Figure 4.5).

A diagram of the anterior structures of the *Drosophila* larva involved in feeding, as well as the neural network and the tissues where the genes described above are expressed are shown in Figure 4.10. In the next section, I analyse the expression of these genes in CFI25 mutants, to test the hypothesis on whether their expression lies downstream of CFI25 and can therefore explain the feeding phenotype observed in CFI25 mutants.

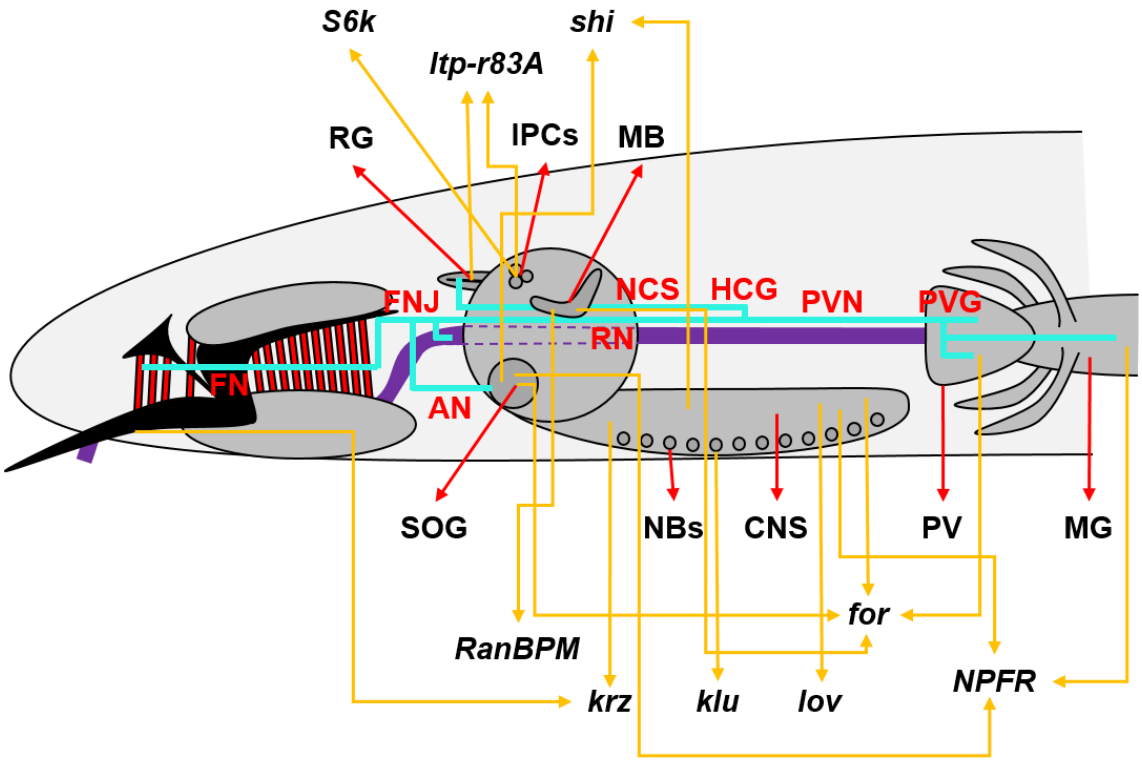


Figure 4.10 The larval feeding machinery and expression of target feeding genes

(Legend on the following page)

Figure 4.10 The larval feeding machinery and expression of target feeding genes

Diagram of anterior feeding structures in the *Drosophila* third instar larva: In black are represented the mouth hooks, in red are represented the muscles that control mouth hook movements, in purple is represented the esophagus, in turquoise are represented nerves of the ENS. The yellow arrows indicate relevant tissues of expression of feeding target genes, the red arrows indicate larval structures and the red labels indicate each nerve and ganglion of the ENS. Anterior is to the left. AN = Antennal nerve, FNJ = Frontal Nerve Junction, FN = Frontal Nerve, RN = Recurrent Nerve, HCG = Hypocerebral Ganglion, NCS = Nervi Cardiostomatogastrici, PVN = Proventricular Nerve, PVG = Proventricular Ganglion, SOG = Subesophageal Ganglion, NBs = Neuroblasts, CNS = Central Nervous System, PV = Proventriculus, MG = Midgut, RG = Ring Gland, IPCs = Insulin Producing Cells, MB = Mushroom Body. Based on (Schoofs *et al*. 2014).

,

4.2.5 *RanBPM*, *S6k*, *klu* and *lov* have affected APA patterns in CFI25 mutants

The previous section highlighted the results of applying a simple bioinformatics pipeline with which I found nine candidate genes that have been involved in feeding and show both APA isoforms and neural expression (Figure 4.9). The molecular function of these genes is diverse and there are kinases, arrestins and transcription factors present in the sample pool. Though all these genes have a component of neural expression during *Drosophila* development, the tissues where they are expressed are also diverse, going from specific cells as the case of *S6k*, where changes in expression in IPCs is sufficient to affect feeding (Wu *et al*. 2005), to broad expression patterns, such as the entire nervous system in the case of *lov* (Bjorum *et al*. 2013).

Given that a reduction in CFI25 levels can affect APA patterns in cells in culture (Li *et al*. 2015; Kubo *et al*. 2006; Masamha *et al*. 2014) and CFI25 mutants show a reduction of CFI25 levels at the RNA and protein levels, as well as a decrease in expression in the VNC of late embryos at the RNA level (Figure 4.6), I addressed the question of whether these target genes have affected APA isoforms in CFI25 mutant first instar larvae. The role that APA plays for the biological function of these genes is not described in the literature. Nonetheless, I detected the existence of these different 3' UTR isoforms from the databases, which are validated by RNA-seq (Attrill *et al*. 2016). In order to assess APA changes in these selected target genes related to feeding, I designed primers for semi-quantitative RT-PCR experiments to get amplicons from the 3'UTRs of these genes (See Chapter 2 and Table 2.5). More specifically, I aimed to detect an amplicon labelled as "Universal" that amplifies the short 3'UTR of the gene,

therefore, amplifying all or most of the RNA isoforms of it, and an amplicon labelled as “Distal” that amplifies only the long or longest 3’UTR of the gene. After obtaining these amplicons by RT-PCR with cDNA from whole wild type and CFI25 mutant first instar larvae (And using RpA1 as a loading control), I divided the Distal signal by the Universal signal, obtaining a “Distal/Universal” (Dis/Uni) ratio. This ratio is a dimensionless value that reflects changes in the length of the 3’UTR of a given gene. Thus, a higher “Dis/Uni” ratio reflects a lengthening of the 3’UTR, whereas a lower “Dis/Uni” ratio reflects a shortening of the 3’UTR. This experimental approach has been widely use to study the effects on APA in different biological conditions both *in vitro* and *in vivo* (Miura *et al*. 2014; Rogulja-Ortmann *et al*. 2014). The results of this analysis in the feeding target genes is shown in Figure 4.11. It can be observed that *RanBPM*, *S6k*, *klu* and *lov* show significantly lower Dis/Uni ratios in CFI25 mutants when compared with wild types. As shown in the representative RT-PCR amplicons (shown in the middle), *RanBPM* shows lower signal for the distal amplicon, which can be interpreted as a retraction of its 3’UTR in the CFI25 mutant condition. *S6k*, *klu* and *lov* show stronger signals for the universal amplicon, which can be interpreted as more production of the short 3’UTR isoforms or a general overexpression of these genes in the mutant condition. All these effects, when measured as Dis/Uni ratios, result in a lower value for CFI25 mutants when compared with wild types. To obtain an indication of the total mRNA expression levels of these genes in CFI25 mutants, I used the signal from the Universal amplicons and normalized them by RpA1. In the case of *RanBPM*, we used primers that amplify the gene at the sixth

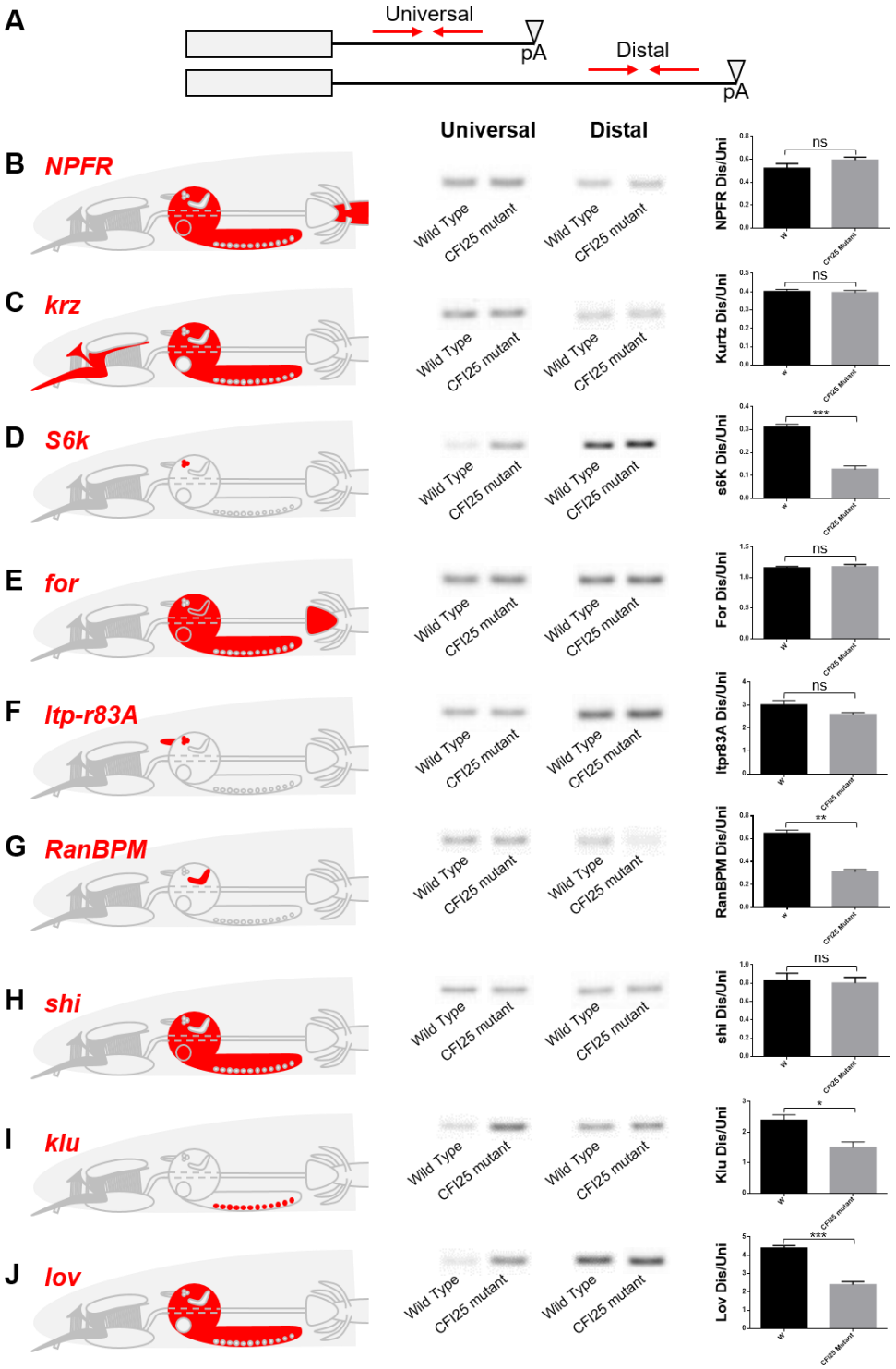


Figure 4.11 *RanBPM*, *S6k*, *klu* and *lov* have affected APA patterns in CFI25 mutants

(Legend on the following page)

Figure 4.11 *RanBPM*, *S6k*, *klu* and *lov* have affected APA patterns in CFI25 mutants

(A) Diagram of the strategy to assess APA in feeding target genes, the grey boxes represent the coding sequence and the black horizontal lines represent both the universal (top) and the distal (bottom) 3'UTRs. The red arrows indicate RT-PCR primer amplicons. **(B-J)** Semi-quantitative RT-PCR experiments to measure APA in feeding genes. To the left, a diagram of the feeding machinery as shown in fig 4.10, this time indicating the relevant tissues of expression of each target gene in red. In the middle, representative amplicons for the universal and distal 3'UTRs in both wild types and CFI25 mutants. To the right, quantification of RT-PCR experiments, for each gene a biological triplicate was performed, values are shown as Dis/Uni ratios (See Chapter 2 for information on semi-quantitative RT-PCR experiments design). Note that *S6k* **(D)**, *RanBPM* **(G)**, *Klu* **(I)** and *Lov* **(J)** have different Dis/Uni ratios when compared with the control. These experiments were performed as biological triplicates. Error bars represent the SEM. Unpaired two-tailed t-tests were used to compare genotypes, * $p < 0.05$.

exon-exon junction, amplifying all mRNA isoforms (See table 2.5). The results of this analysis are shown in Figure 4.12. Here it could be seen that *S6k*, *klu* and *lov* showed higher expression levels in CFI25 mutants when compared with wild types. However, *RanBPM* expression levels were not different between genotypes. These results show that the effects on the 3'UTRs of these genes by CFI25 are different, and while for *RanBPM* only the long isoform was affected, for the remaining *S6k*, *klu* and *lov* there is more production of the short portion of their 3'UTRs. Thus, *RanBPM* is the only case were we can clearly distinguish APA as retraction of the long 3' UTR from a general overexpression observed in *S6k*, *klu* and *lov*. A summary of the effects in feeding genes in CFI25 mutants is shown in figure 4.13

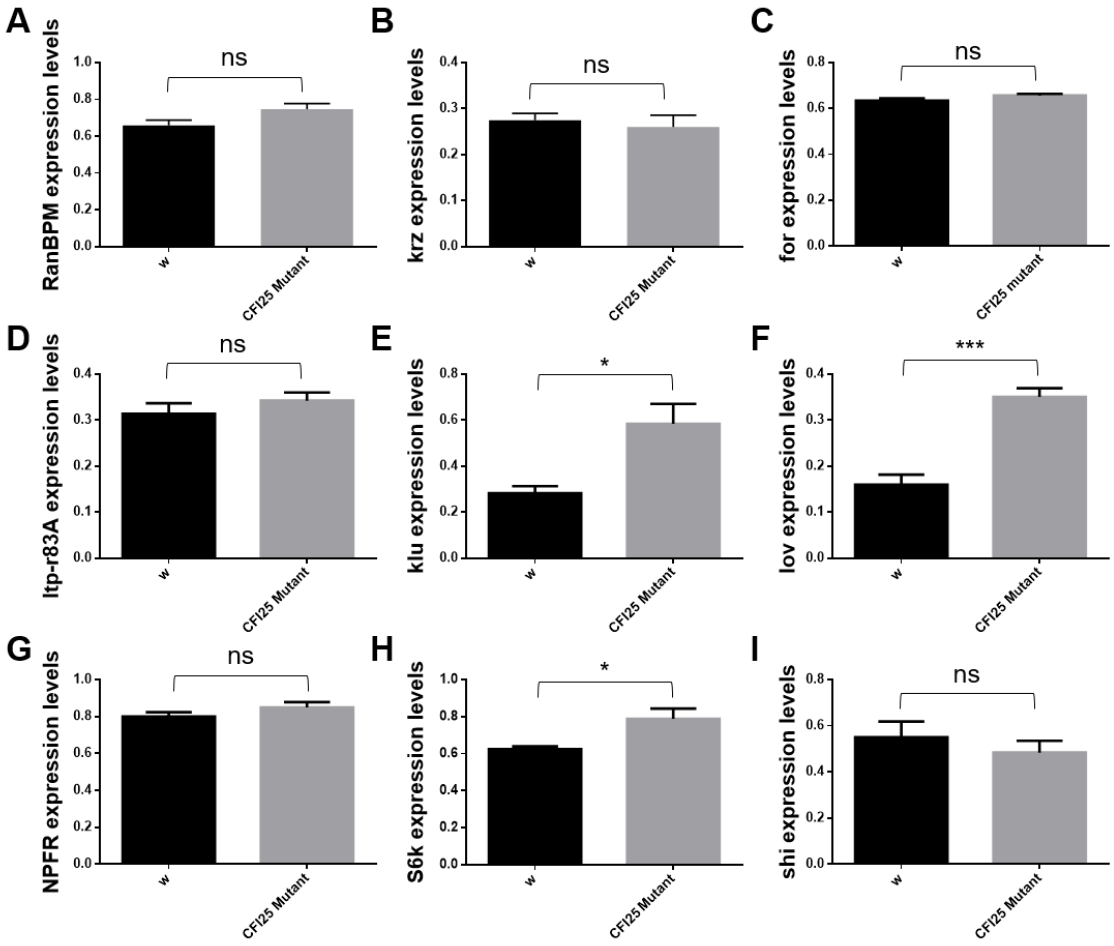


Figure 4.12 *S6k*, *klu* and *lov* have higher mRNA levels in CFI25 mutants

(Legend on the following page)

Figure 4.12 *S6k*, *klu* and *lov* have higher mRNA levels in CFI25 mutants

(A-I) Quantifications of mRNA target feeding genes levels by semi-quantitative RT-PCR, the values are normalized by RpA1 and shown as biological triplicates (See Chapter 2). Note that *klu* **(E)**, *lov* **(F)** and *S6k* **(H)** have higher mRNA levels when compared with the controls. The universal primers for *lov* also bind to the last exon of isoform RE, which is 3'UTR in isoform RF (lov has 5 RNA isoforms). Isoform RE has a different transcription start site than isoforms RC and RF, thus, the increase in *lov* mRNA levels might be a specific increase in isoform RE. *shi* has 15 RNA isoforms, the "Uni" amplicon was set to cover as many 3'UTRs as possible, which are 8. In the other 7 isoforms these primers target intron sequences. Given that polyadenylated mRNA was used for these RT-PCR experiments, these isoforms were not amplified by PCR and thus, the expression levels of *shi* are possibly an underestimation of the absolute mRNA levels. Nonetheless, the "Dis/Uni" ratios for *shi* in fig 4.11 were not affected. These experiments were performed as a biological triplicates. Error bars represent the SEM. Unpaired two-tailed t-tests were used to compare genotypes, * $p < 0.05$.

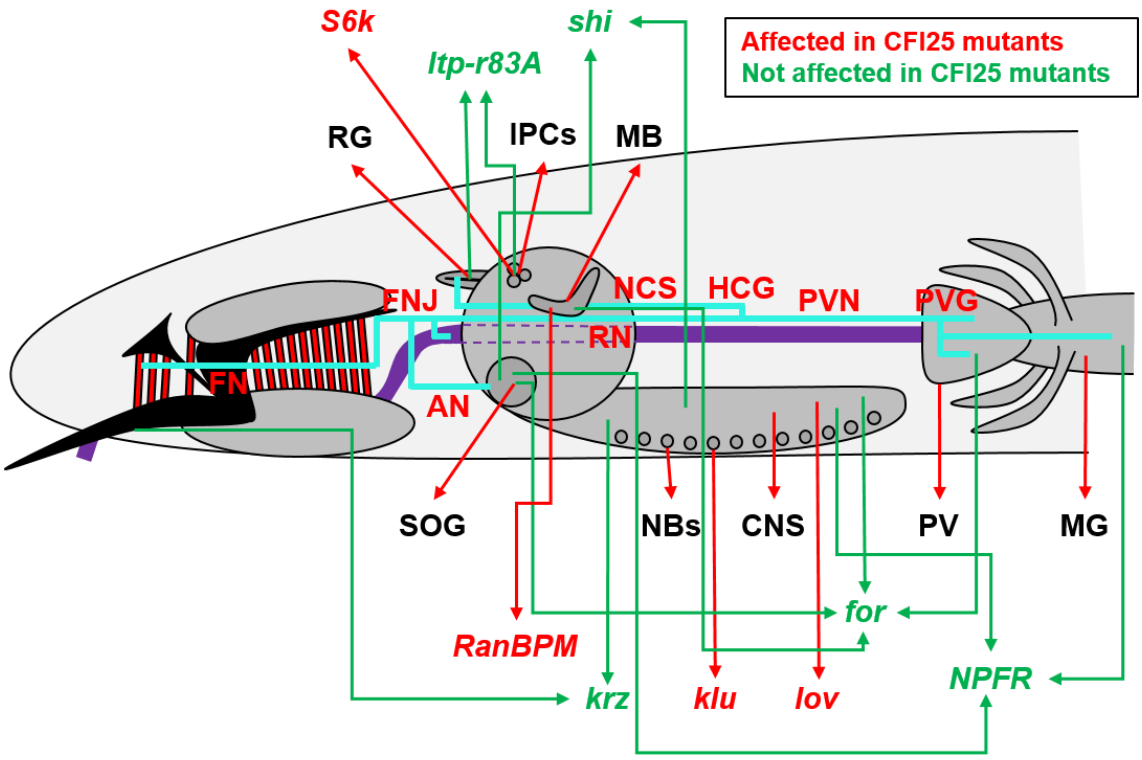


Figure 4.13 Summary of feeding genes affected in CFI25 mutants

Diagram from figure 4.10 showing in red the genes with affected APA patterns in CFI25 mutants (Figure 4.11) and in green the genes that do not show significant differences. From these 4 genes, *klu*, *lov* and *S6k* also show significantly higher expression of mRNA levels (Figure 4.12).

4.3 Discussion

The work presented in this Chapter shows that a mutation by means of an insertion of a transposable element in the *Drosophila* orthologue of the CPA factor CFI25 leads to a developmental arrest during the first instar larvae stage (Figure 4.2) and this phenotype is due to a defect in larval feeding behaviour (Figure 4.3). Excision of this P-element by using Δ2-3 transposase is sufficient to rescue the mutants to a wild type phenotype. Additionally, CFI25 mutants at the phenotypic level show a “collapsed” mouth hook phenotype (Figure 4.4) and defects in food movement from the pharynx to the midgut (Figure 4.5). Regarding the former, analysis of *krz* RNA in CFI25 mutants, a β-arresting gene involved in the formation of the CPS, among other biological roles (Roman *et al*. 2000), did not show differences in both APA and mRNA expression levels, making this gene unlikely to be a link with the phenotype observed (Figures 4.11 and 4.12). Nonetheless, several alleles of *klu* have been reported to show mouth hook defects, but the authors neither give a detailed account nor show images for the phenotypes observed (Klein and Campos-Ortega 1997) to be able to compare this with the phenotype observed in CFI25 mutants. Besides, CFI25 mutants do show differences in *klu* APA, as well as mRNA levels, making this gene an interesting candidate to study further (Figures 4.11 and 4.12). Regarding the latter phenotype observed in CFI25 mutants, showing trapped food in their pharynx (Figure 4.5), none of the candidate genes that were studied showed a similar phenotype when genetically manipulated. Nevertheless, this phenotype has been observed before for genes related with feeding. For example, mutants for *ppl*, a protein with homology to a vertebrate enzyme involved in glycine catabolism, show exactly the same phenotype as CFI25 mutants (Zinke *et al*. 1999). Yet, *ppl*

is exclusively expressed in the embryonic and larval fat body, a tissue where we do not see expression of CFI25 at the mRNA level (Figure 3.10). Also, *ppl* mutants show a wandering phenotype, a behaviour we do not see in CFI25 mutants (Figure 4.7). Taken together, these observations indicate that *ppl* is also unlikely to be a linking gene for the phenotype observed in CFI25 mutants. Another candidate target gene affected in CFI25 mutants is *lov*, a transcription factor of the BTB/POZ family. *lov* shows a broad expression pattern in the nervous system of late embryos (Bjorum *et al*. 2013), which is comparable to the expression pattern observed for CFI25 at similar developmental stages (Figure 3.10). In spite of this, *lov* mutants feeding phenotype is attributed to locomotor defects in regards to forward locomotion and mouth hook movements (Bjorum *et al*. 2013). None of these effects were observed in CFI25 mutants (Figure 4.7), also making this gene unlikely to be the link with the feeding phenotype observed in CFI25 mutants.

Although changes in the 3'UTR of genes do not change the encoded protein sequence and therefore, their structure. They can still affect important aspects of protein function, such as protein amounts and localization. A shortening in 3'UTRs can lead to an increase in protein production, because a shorter 3'UTR will now bind fewer trans-acting factors that can negatively impact mRNA such as microRNAs and RBPs. For example, it has been shown that the overexpression of proto-oncogenes, which leads to tumorigenesis, is due to a shift of APA patterns from long to short 3'UTRs (Mayr and Bartel 2009). On the contrary, an elongation of 3'UTRs is observed during organogenesis and cell differentiation, where mRNAs are exposed to more intricate post-transcriptional regulation by these same trans-acting factors. For example, differentiation of

C2C12 cells into myotubes is associated with 3'UTR lengthening (Ji *et al*. 2009), as well as neural development (Thomsen *et al*. 2010; Hilgers *et al.* 2011). Changes in 3'UTR length can affect protein localization, for example, the long 3'UTR of the human CD47 gene enables expression of this protein in the cell surface, whereas the short 3'UTR RNA isoform enables localization to the endoplasmic reticulum (Berkovits and Mayr 2015).

A shift could be observed in 3'UTR APA for *s6K*, *lov*, *RanBPM* and *klu* in CFI25 mutants, which was reflected in higher production of the short isoforms for *s6K*, *lov* and *klu,* while *RanBPM* shows a decrease in the long isoform without noticeable changes in the production of the short isoform (Figure 4.11). Since we are unable to rule out effects on the other genes that were not detected because of limitations in the sensitivity of the technique used (Semi-quantitative RT-PCR) it can only be partially confirmed. Still, a similar trend has been observed after CFI25 depletion in human cells for *TIMP-2*, *syndecan2*, *ERCC6* and *DHFR*, where the short RNA isoforms where increased after CFI25 knockdown, when assessed by northern blot experiments (Kubo *et al*. 2006). The feeding phenotypes observed for *RanBPM*, *lov,* and *klu* in the literature, were caused by a reduction of these genes products by means of mutation by insertion of transposable elements (Scantlebury *et al*. 2010; Bjorum *et al*. 2013; Melcher and Pankratz 2005), as in the case of our CFI25 mutants. In CFI25 mutants, there is an upregulation of these genes at the mRNA level (Figure 4.12), which potentially also leads to an upregulation at the protein level, although I could not measure protein levels of these genes due to lack of antibodies. These results make it difficult to predict a molecular mechanism that explain the feeding defect observed in CFI25 mutants that involves these genes. Nevertheless, *S6k* gives

an interesting case. This because it was shown that overexpression of *S6k* in IPCs leads to an attenuation of feeding response of both solid and liquid food, even though both control and experimental larvae were fasted for 120 minutes before the essay. Furthermore, their frequency of mouth hook contraction decreases in comparison with the control, leading to a defect in feeding (Wu *et al*. 2005). Similarly, CFI25 mutants show an increase in *S6k* at the mRNA level (Fig 4.12), making *S6k* an interesting gene to be studied in the future. Its detailed expression pattern is not known, although it has been shown by northern blot experiments that is expressed across all embryonic and larval stages, as well as in adults (Watson *et al*. 1996).

Modification of CFI25 expression in IPCs would be an interesting experiment to assess if the changes provoked by CFI25 levels in only these cells are sufficient to induce a feeding phenotype. Likewise, a reduction of *S6k* levels by RNAi in IPCs in a CFI25 mutant background would also be informative, the prediction to see a partial rescue of feeding in these animals because of a correction in *S6k* levels. A more informative experiment to show that changes in *S6k* APA are sufficient to cause a feeding phenotype in CFI25 mutants would be to mimic the changes in the 3'UTR of *S6k* observed in CFI25 mutants by means of targeted mutation. For example, by elimination of the distal PAS using CRISPR (Hsu *et al*. 2014; Bassett and Liu 2014). Consequently, I propose a model in which depletion of CFI25 in neural tissues leads to an upregulation of S6k in IPCs by a switch to the use of its proximal PAS. Thus, triggering the phenotype observed. Future experiments such as the ones proposed previously can help us in the understanding of this phenotype at the molecular level.

Finally, though this allele of CFI25 has a very interesting phenotype, which is bound to be informative to show the biological roles of CPA factors *in vivo* when studied further, it is the only available allele of CFI25 so far which is limiting because the phenotypes observed in a particular allele, such as this one, can at times only show an incomplete view of gene function because of pleiotropy. For example, different polymorphisms of the gene *Catsup*, a negative regulator of tyrosine hydroxylase, are independently associated with diverse phenotypes, such as variation in longevity, locomotor behaviour and sensory bristle number (Carbone *et al*. 2006). What is more, phenotypes observed by P-element insertional mutants can be highly diverse and different from a true null mutant even if different transposable elements are inserted within the same gene only a few base pairs apart. For example, different P-element insertions in the gene *neutralized* have different effects on olfactory behaviour, aggression and mechanosensory stimulation, these behavioural phenotypes are correlated with distinct structural changes in integrative centers in the brain, the mushroom bodies, and the ellipsoid body of the central complex (Rollmann *et al*. 2008). The P{EP} insertion in CFI25 mutants is localized in the 5'UTR, meaning that although I assessed for feeding phenotypes in insertions upstream of this region (Figure 4.3 panel D), I cannot rule out effects on regulatory regions affecting CFI25 transcription, whose promoters and enhancers are not well characterised. This could potentially lead to quantification of the effects of an allele that does not behave as a true amorph. There is only one chromosomal deficiency that covers this region, which is more than 70 kbps long and disrupts 29 genes (Df(3L)BSC113), making it cumbersome to carry out informative complementation tests using this deficiency line. What is more, this CFI25 allele

over this large deficiency is not able to trigger a feeding phenotype. An analysis of the genes covered by this deficiency indicates that more than half of them have unknown molecular functions and are not characterized, making it necessary in the future to generate new alleles of CFI25 by independent methods, such as CRISPR, as well as more genetic tools in this chromosomic region, to understand how CFI25 is involved in *Drosophila* larval behaviour fully.

In summary, despite the limitations of this study, I uncover new biological and neural roles for a CPA factor that is part of the most conserved complex between humans and *Drosophila* (Figures 3.1 and 3.3), which has also been implicated as a central actor in the control of APA in mammalian cells (Kim *et al*. 2010; Kubo *et al*. 2006; Li *et al*. 2015) as well as in relevant medical conditions, such as glioblastoma tumorigenesis (Masamha *et al*. 2014). In the next Chapter I will study how the levels of the members of the CFI complex can control APA during *Drosophila* neural development.

Chapter 5

CFI levels control alternative polyadenylation within the developing nervous system

5.1 Chapter overview

In the previous chapter I showed that a mutation in the *Drosophila* orthologue of CFI25 had effects in larval feeding behaviour, as well as effects in APA patterns in the target feeding genes *S6k*, *klu*, *RanBPM* and *lov* (Figure 4.11). In Chapter 3 I also showed that the expression pattern of both CFI25 and CFI68 is enriched in the nervous system during late stages of embryogenesis (Figures 3.10 and 3.11). Given that the nervous system expresses long 3' UTR isoforms in a tissue-specific manner, in this chapter I address the question of whether the expression levels of CFI factors within this tissue are used as a cue to express such isoforms.

I expressed RNA-interference (RNAi) constructs made to target both CFI25 and CFI68 within the physiological context of the developing nervous system in late embryos, where I then assessed effects in APA on genes with reported 3' UTR neural extensions by categorising them according to the length of these extensions, given their potentially different biological roles. First, I observed that the reported 3' UTR lengths of 29 out of 30 neural-extended genes do not match the values given by the current databases, as their transcripts are shorter or longer than originally published. Some of these have 3' UTR lengths similar to those observed in mammalian transcripts. Furthermore, 50% of the tested neural extended genes have altered APA patterns after the depletion of CFI factors within the nervous system after RNAi. These results suggest that the observed neural enrichment in expression of CFI factors within the nervous system are used as a cue for 3' UTR extensions during *Drosophila* neural development.

5.2 Results

5.2.1 There is an extension of 3'UTRs within the developing nervous system

In *Drosophila*, as in all insects, there are 8 *Hox* genes that are clustered into two complexes. The Antennapedia complex consists of five genes that are expressed towards the anterior end of the embryo, within the anteriorposterior axis: *labial* (*lab*), *proboscipedia* (*pb*), *deformed* (*Dfd*), sex combs reduced (*Scr*), and *Antennapedia* (*Antp*). The *Bithorax* complex consists of the remaining three genes that are expressed towards the posterior end of the embryo within the same axis: *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*) (Mallo and Alonso 2013; Lewis 1978). Four of the *Hox* genes: *Antp*, *Ubx*, *abd-A* and *Abd-B,* present APA isoforms where there is evidence of short and long 3'UTRs being produced (Thomsen *et al.* 2010). This process helps to diversify the way in which miRNAs and RBPs can regulate these key developmental genes to ensure precise expression levels in different tissues as well as the precise patterning of the animal along the anteroposterior axis. As mentioned earlier, it is interesting that during early embryogenesis in *Drosophila*, the short 3'UTR is predominantly expressed in the epidermis, however, during late embryogenesis the long 3'UTR starts being produced and is restricted to the nervous system (Thomsen *et al.* 2010). A detailed example of this phenomenon is shown in Figure 5.1 for *abd-A*, where it can be seen that from stage 13 onwards, the long 3' UTRs starts being expressed and this expression becomes restricted to the nervous system. This trend of neural 3' UTRs has been shown to be shared among the

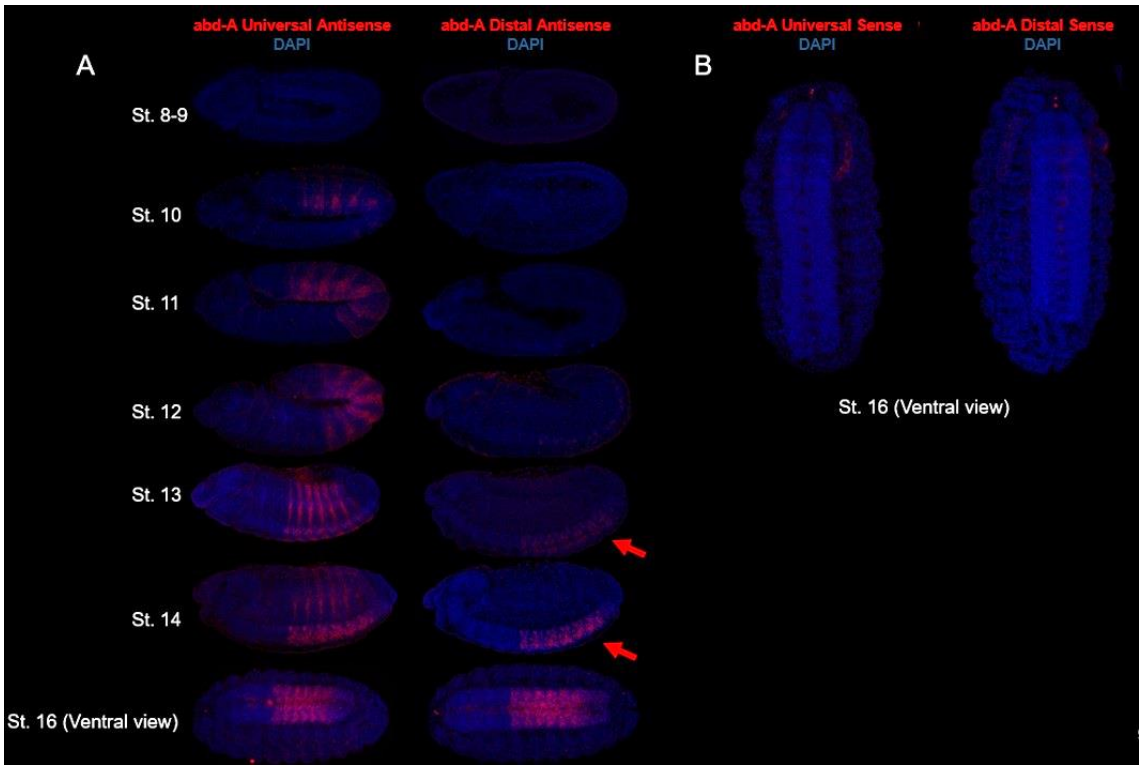


Figure 5.1 *Hox* genes show a lengthening of their 3'UTRs during neural development

(A) *Drosophila* embryos across different stages from top to down showing the mRNA localization of *abdominal-A* during embryogenesis by FISH using RNA probes that target the short (universal) or long (distal) APA isoforms (See chapter 2). Note that the extended RNA isoform starts being expressed during stage 13 (red arrow) and is restricted to the nervous system. **(B)** Sense controls for the probes used in stage 16 embryos. Note that there is no signal from the ventral nerve cord.

four *Hox* genes that display APA isoforms (Thomsen *et al.* 2010). This phenomenon was later shown to be more global and involves many more genes than the *Hox* gene clusters (Hilgers *et al*. 2011; Smibert *et al.* 2012). Curiously, while the extended 3'UTR (only extension beyond short 3' UTR considered) observed in *Hox* genes can be of up to 1.4 kb long, as in the case of *AntP* and *Ubx,* other genes have reported extensions of up to 11 kb (only extension beyond short 3' UTR considered as well), as in the case of *mei-P26* (Hilgers *et al.* 2011). While it has been shown that the biological function of long 3'UTRs is linked to the targeting of trans-acting factors, such as miRNAs and RNA binding proteins (Patraquim *et al*. 2011; Gupta *et al*. 2014), extensions of lengths such as the reported ones in the literature may not be explainable only by this biological function in the cell. Thus, the reasons why 3' UTRs can be extended beyond the requirements for differential miRNA/RBP regulation remain unknown.

I questioned whether the abundance of CFI factors within neurons acts as a cue to control PAS selection in extended genes during neural development. For this purpose, given that the expression patterns of both CFI25 and CFI68 show an enrichment in the nervous system during late embryogenesis (Figure 3.10 and 3.11), they are the most abundant CFI factors in neural tissues (Figure 3.6 and 3.9) and the most conserved complex between humans and *Drosophila* (Figure 3.3), it appears that CFI is also good candidate to address this question in *Drosophila.* Furthermore, the molecular structure and function of CFI in mammals have been well described in the literature (Yang *et al.* 2011).

This hypothesis of neural APA control by CFI factor abundance will also put to test the model proposed by James Manley and colleagues in the physiological context of neural development. This was described in the case of CSTF64 with

the IgM heavy chain during B-cell differentiation, in which the abundance of CSTF64 acts as a switch for the selection of two different PAS within this gene, which determines whether there is either production of a membrane-bound or secreted protein (Takagaki *et al.* 1996). While analysing the reported genes with neural-3' UTR extensions to test this hypothesis, I found that the 29 out of the 30 "Neural-extended genes" reported in 2011 by Valérie Hilgers and colleagues have reported 3' UTR lengths that do not match the current databases. This was found by comparing the current FlyBase data (Flybase release 6.12, July 2016) with their results. The length of the 3' UTRs of the 30 extended genes as published in the Hilgers study in comparison with the FlyBase current values are shown in Table 5.1. The authors only give approximate values for the length of 3' UTRs in that study, and because of this, if the differences are of less than 50 bp, this has been considered as 'no difference' between the old and new data, especially given that 30% of the 3'UTR lengths discrepancies fell below this range (Difference ≈ 0 in table). The difference calculated between the new and old data is shown in red for negative values (current 3' UTR shorter than previously reported) and in green for positive values (current 3' UTR longer than previously reported). "Short" is the shortest 3' UTR sequence available and "Extended" is the longest 3' UTR sequence available without taking in account the "Short" sequence included in it when applies.

The most dramatic difference in FlyBase versus experimental data is the longest gene reported in the Hilgers study – *mei-P26* – for which the extended

Table 5.1 3' UTR length comparison between Hilgers *et al* and current data

Gene	Short 3'UTR length (Hilgers et al) (kb)	Current short 3'UTR length (kb)	Difference (kb)	Extended 3'UTR length (Hilgers et al) (kb)	Current extended 3'UTR length (kb)	Difference (kb)
pum	1.200	1.188	-0.012 ≈ 0	3.400	3.414	0.014 ≈ 0
nrg	0.100	0.128	0.028 ≈ 0	3.000	3.051	**0.051**
ago1	1.400	0.435	**-0.965**	4.600	4.150	**-0.45**
gB13F	0.100	0.050	**-0.05**	3.400	3.329	**-0.071**
step	0.300	0.279	-0.021 ≈ 0	3.800	1.263	**-2.537**
wdb	0.500	0.527	0.027 ≈ 0	4.000	3.367	**-0.633**
nmo	0.500	0.410	**-0.09**	4.200	3.814	**-0.386**
fne	0.300	1.202	**0.902**	4.200	3.725	**-0.475**
nej	0.500	0.192	**-0.308**	4.500	4.073	**-0.427**
ADAR	0.200	0.247	0.047 ≈ 0	4.300	4.064	**-0.236**
shep	0.900	0.927	0.027 ≈ 0	5.500	4.865	**-0.635**
hrb27C	0.700	0.591	**-0.109**	6.400	6.011	**-0.389**
elav	0.900	0.556	**-0.344**	7.200	8.001	**0.801**
brat	1.300	0.593	**-0.707**	8.500	7.906	**-0.594**
imp	1.100	0.856	**-0.244**	8.400	8.344	**-0.056**
mei-P26	0.900	0.844	**-0.056**	11.900	17.650	**5.75**
dpld	1.500	0.238	**-1.262**	4.500	4.303	**-0.197**
heph	0.900	0.371	**-0.529**	4.700	4.497	**-0.203**
rbp6	1.500	0.057	**-1.443**	6.900	6.836	**-0.064**
bol	1.100	0.603	**-0.497**	5.000	4.454	**-0.546**
fas1	0.900	0.857	-0.043 ≈ 0	2.600	1.723	**-0.877**
ga49B	1.200	0.120	**-1.08**	3.600	3.931	**0.331**
cam	0.800	0.116	**-0.684**	3.400	4.899	**1.499**
msi	1.700	0.389	**-1.311**	5.000	7.287	**2.287**
pdp1	2.000	1.989	-0.011 ≈ 0	4.000	2.879	**-1.121**
tyf	0.800	0.262	**-0.538**	3.000	3.760	**0.76**

cip4	0.500	0.527	0.027 ≈ 0	1.500	1.039	**-0.461**
mub	1.400	0.926	**-0.474**	7.100	8.588	**1.488**
CG34360	1.100	0.461	**-0.639**	7.600	7.115	**-0.485**
rbp9	1.500	0.730	**-0.77**	5.000	4.318	**-0.682**

3’ UTR was reported to be of 11.9 kb but it actually is 17.650 kb long. Although this discrepancy in 3’ UTR length may not be of high relevance to address the question on the mechanisms by which neural APA is controlled, it suggests that 3’ UTR length control is relevant for the nervous system and perhaps more genes undergo this process that have not been yet discovered. When considering the case of *mei-P26*, which has an actual length of the distal 3’ UTR of almost 150% the originally reported value, it highlights what could be the biological function of these 3’ UTRs (discussed later).

The reason why the length of 3’ UTRs in the Hilgers study are different to those originally reported has probably to do with the fact that the RNA-seq database used for reference: The National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) database (Pruitt *et al.* 2014) is continually curated and updated. Therefore, over the course of five years there was more sequencing depth which revealed the 3’ UTR lengths in more detail.

In this chapter I address the question of the biological mechanism by which the 3’ UTR extensions are achieved. In order to do this, I performed RNAi knockdowns of CFI25 and CFI68 within the developing nervous system by using the Gal4/UAS system (Brand & Perrimon 1993) to target expression of the dsRNA with an *elav* driver, which drives expression in neural tissues (See Table 2.1) in stage 16 embryos (Campos-Ortega and Hartenstein, 2013). The embryos were then assessed for the effects on APA in reported 3’ UTR neural-extended genes. These experiments are discussed in the next section.

5.2.2 The knockdown of CFI25 and CFI68 within the developing nervous system affects APA in *Hox* genes and 3' UTR-extended genes

Natural variations of CPA factor expression during biological processes have been reported previously. For example, during differentiation of C2C12 cells into myotubes, there is a downregulation in the expression levels of the three CSTF members: CSTF50, CSTF64 and CSTF77 (Ji *et al*. 2009). In this same study the authors found that the levels of τCSTF64 were also upregulated during the same process. τCSTF64 is a paralog of CSTF64 which is absent in *Drosophila* (Figure 3.2). τCSTF64 has been shown to mediate testis-specific PAS selection in mammals, whereas the normal CSTF64 gene is inactive during male meiosis in the mouse. Similarly, as shown in Chapter 3, there was a correlation between CPA factor mRNA levels and the progression of *Drosophila* embryogenesis. More specifically, factors that are part of the same complex showed a decrease in mRNA levels as embryonic development progresses (Figure 3.4). As shown for CFI (Figure 3.10 and 3.11), this is caused by restricted expression patterns in the nervous system, hence showing an increase in expression in this tissue when compared with others (Figure 3.6). These observations suggest that variation in CFI factors abundance during neural development is what controls 3' UTR length.

To test this hypothesis, we need to manipulate CFI factor levels in order to evaluate whether this is what controls 3' UTR length in this tissue. Since both subunits of CFI, CFI25 and CFI68, are essential for the functioning and coherent assembly of the CFI complex, I address this question by focusing on both of them. Similar experiments have been widely performed *in vitro,* for example, human cells have been extensively used to downregulate CFI factors to then assess effects on 3' UTRs (Kubo *et al*. 2006; Masamha *et al*. 2014). However, HeLa cells

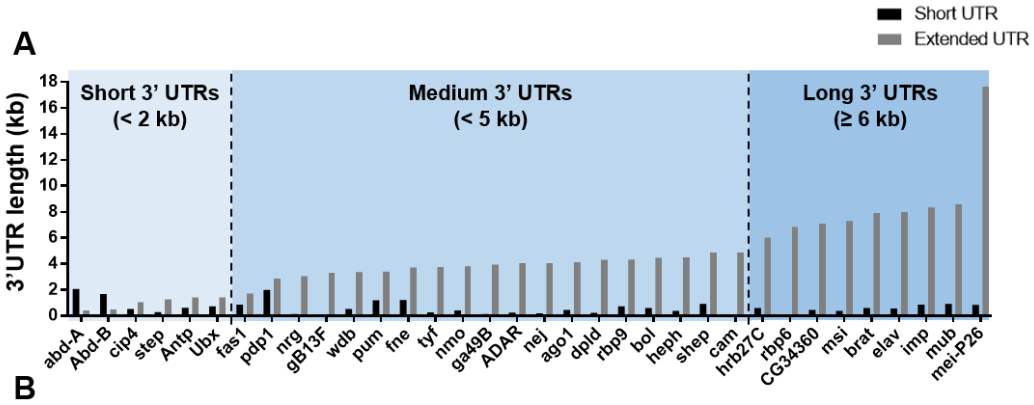
and other cell lines have themselves a cancerous origin, and it has been shown that 3' UTR lengths and APA are misregulated during cancer (He *et al*. 2016; Mayr and Bartel 2009; Erson-Bensan and Can 2016). Also, it has been shown that alterations in Fip1 abundance can alter stem cell self-renewal and somatic cell reprogramming by 3' UTR shortening (Lackford *et al.* 2014). Yet, the read-out obtained by these experiments may not reflect the changes that would occur in a natural context *in vivo*. These experiments still provide evidence in favour of the idea of CPA factor abundance behind the control of APA and thus influencing biological processes.

In order to test whether CFI factors abundance acts as the mechanism controlling APA within the developing nervous system, I selected the genes described in the Hilgers study with their curated 3' UTR lengths, as well as the *Hox* genes. Further, different 3' UTR lengths may have different biological functions. While short extensions are coherent with the concept of differential miRNA/RBP regulation (Patraquim *et al.* 2011; Gupta *et al.* 2014), extremely long 3' UTRs may play other biological functions beyond the aforementioned ones, which if existing, are largely unexplored. Hence, I separated the extended genes in three categories according to the length of their extended 3' UTRs. By doing this, I can then relate if effects observed after CFI factor depletion can be related with this property. Thus, the three different categories are labelled as "Short 3' UTRs" for the extensions of less than 2 kb. "Medium 3' UTRs" for the extensions of less than 5 kb and "Long 3' UTRs" for the extensions of 6 kb or more. This classification is based on the distribution observed for distal 3' UTR length, which is non-linear and reflects these three categories. This classification resulted in seven genes within the "Short 3' UTR" category, with all four *Hox* genes falling within this

category. 18 genes within the “Medium 3’ UTR” category and nine genes within the “Long 3’ UTR” category. I then selected representative genes from each category in a proportional way to how many genes are within each one. Thus, I selected 5 genes from the “Short” and “Long” category and 10 genes from the “Medium” category to be tested for effects on APA after knockdown of CFI factors within the developing nervous system. The ranking of extended genes plus the selection of genes for testing, together with their biological function is shown in Figure 5.2.

To assess for the efficiency of the RNAi treatments, for CFI25 the levels of protein expression were quantified using the same antibody previously used for Western Blot experiments (Figure 4.6). For CFI68 I tried an antibody against the human version of CFI68 whose immunogenic region is 80% similar in *Drosophila* but unfortunately it did not work (Data not shown). The reasons of this likely to be for not having the exact required matches for antibody binding in spite of the high overall similarity. Nonetheless, by semi-quantitative RT-PCR, a reduction in CFI68 levels could be observed although this may be an underestimation of the real magnitude of the knockdown at the peptide level. After quantification, it was concluded that RNAi led to a reduction of nearly 40% in CFI25 protein levels and of nearly 30% in CFI68 mRNA levels (Figure 5.3).

A



B

Category	Gene	Curated length of Short 3' UTR (kb)	Curated length of Distal 3' UTR (kb)	Biological Function	Reference
Short 3' UTRs	abd-A	2.060	0.413	Hox gene	(Thomsen et al. 2010)
	Abd-B	1.674	0.503	Hox gene	(Thomsen et al. 2010)
	Step	0.279	1.263	Regulation of endocytosis	(Lee & Harris 2013)
	Antp	0.623	1.412	Hox gene	(Thomsen et al. 2010)
	Ubx	0.729	1.420	Hox gene	(Thomsen et al. 2010)
Medium 3' UTRs	nrg	0.128	3.051	Axon extension	(Yamamoto et al. 2006)
	GB13F	0.050	3.329	Asymmetric neuroblast division	(Yu et al. 2003)
	wdb	0.527	3.367	Sensory perception of pain	(Neely et al. 2010)
	pum	1.188	3.414	Dendrite morphogenesis	(Ye et al. 2004)
	fne	1.202	3.725	Regulation of mRNA processing	(Zaharieva et al. 2015)
	nmo	0.410	3.814	Gravitaxis	(Toma et al. 2002)
	ADAR	0.247	4.064	Adenosine to inosine editing	(Keegan et al. 2011)
	nej	0.192	4.073	Synapse assembly	(Marek et al. 2000)
	ago1	0.435	4.150	Gene silencing by miRNA	(Rehwinkel et al. 2005)
	shep	0.927	4.865	Neuron remodelling	(Chen et al. 2014)
Long 3' UTRs	hrb27C	0.591	6.011	Axon guidance	(Berger et al. 2008)
	Brat	0.593	7.906	Ventral cord development	(Bivik et al. 2015)
	elav	0.556	8.001	CNS development	(Hummel et al. 1999)
	Imp	0.856	8.344	Neuron remodelling	(Medioni et al. 2014)
	Mei-P26	0.844	17.650	Germ cell development	(Page et al. 2000)

Figure 5.2 Categories for 3' UTR length of *Hox* genes and reported neural-extended genes and biological function of selected genes

(Legend on the following page)

Figure 5.2 Categories for 3' UTR length of *Hox* genes and reported neural-extended genes and biological function of selected genes

(A) Ranking of *Hox* and neural-extended genes from (Hilgers *et al.* 2011) according to the curated length of their extended 3'UTR. The length of their short 3' UTR is also shown for reference. **(B)** Genes selected from each category for APA testing experiments after neural knockdown of CFI factors. The precise curated length of their 3' UTR, as well as an example of their biological function is also shown.

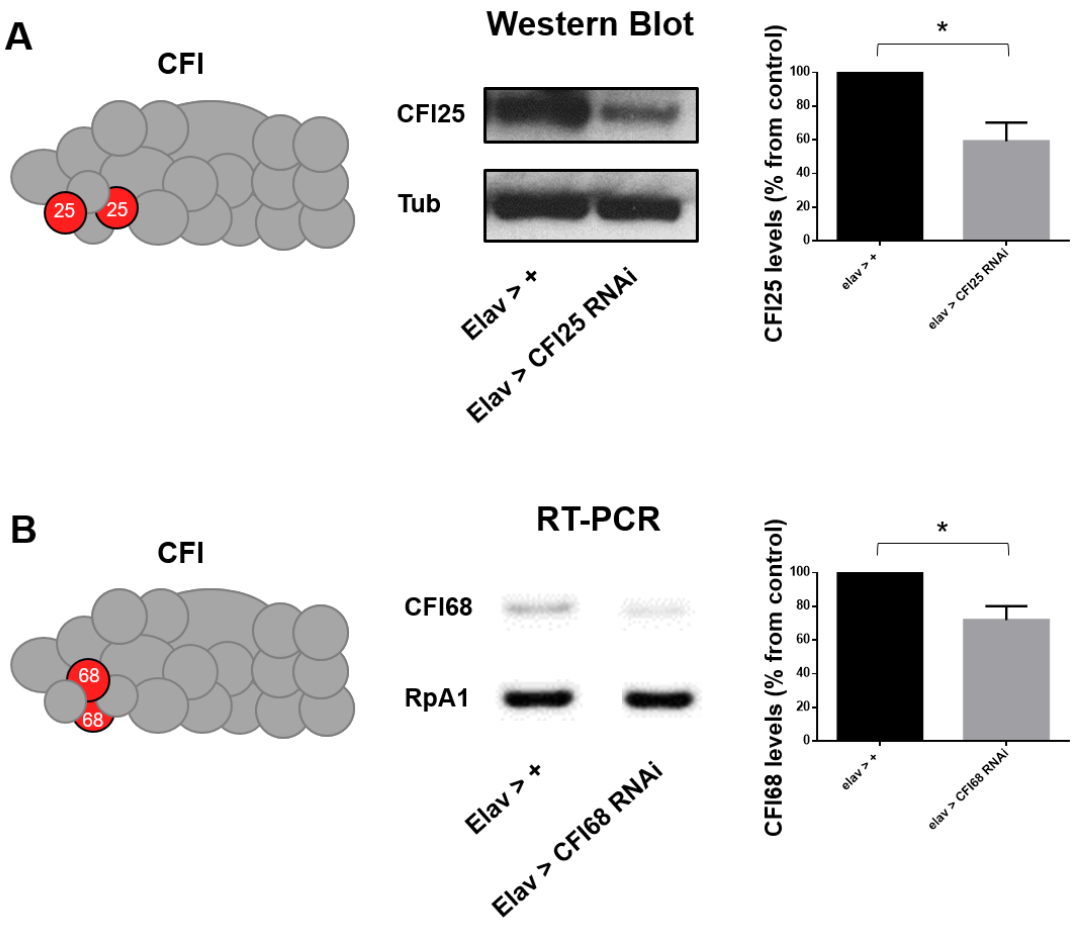


Figure 5.3 RNAi efficiencies of neural knockdown of CFI factors during late embryogenesis

(Legend on the following page)

Figure 5.3 RNAi efficiencies of neural knockdown of CFI factors during late embryogenesis

(A) Diagram representing the location of CFI25 within the CPA machinery, CFI25 levels in stage 16 embryos for the control (elav > +) and the knockdown of CFI25 within the nervous system (elav > CFI25 RNAi), assessed by Western Blot in three independent biological replicates. In each experiment, after normalizing by tubulin, each control was used as 100% and the value for the knockdown was transformed according to its control in each separate experiment. **(B)** Diagram representing the location of CFI68 within the CPA machinery, CFI68 levels in stage 16 embryos for the control (elav > +) and the knockdown of CFI68 within the nervous system (elav > CFI68 RNAi), assessed by semi-quantitative RT-PCR in three independent biological replicates. In each experiment, after normalizing by RpA1, each control was also used as 100% and the value for the knockdown was transformed according to its control in each separate experiment. An unpaired two-tailed t-test was used to compare genotypes, * $p < 0.05$.

To assess the effects on APA in the selected genes after neural depletion of CFI factors, semi-quantitative RT-PCR was used (See Chapter 2). This was conducted similarly to the experiments done in Chapter 4, to assess for effects on APA in feeding genes. Primers were designed to conduct amplification of the "Universal" and the "Distal" 3'UTRs. Depending on the number of mRNA isoforms of each gene, and their specific patterns of APA, the "Universal" amplicons were designed to amplify the shortest 3' UTR in all or as many isoforms as possible for the cases where, given their sequence length and composition, it was feasible to amplify those regions by PCR. When this was not possible, the next short, shared 3'UTR sequence was selected as "Universal". Conversely, the "Distal" primers were designed to amplify the longest 3'UTRs for each of the target genes, which also depending on how many and the nature of the mRNA isoforms for each gene, would amplify only one or more than one isoform (For primer design see Chapter 2). Please note that if the "Universal" primers are less efficient than the "Distal" ones for PCR amplification, or if the "Universal" primers only target a subset of transcripts for a given gene, such as the case of *nrg*, it is possible to have lower "Universal" than "Distal" signal, although this may seem counterintuitive.

After amplifying the above-mentioned regions by PCR from late embryos from the control and the knockdown of either CFI25 or CFI68 within the nervous system, I calculated the "Distal/Universal ratio" for the control and each one of the treatments. For each sample, RpA1 was amplified as a loading control. The results of this analysis are shown in Figures 5.4, 5.5, 5.6 and 5.7. From these results it can be observed that within the "Short 3' UTR" category of genes, two

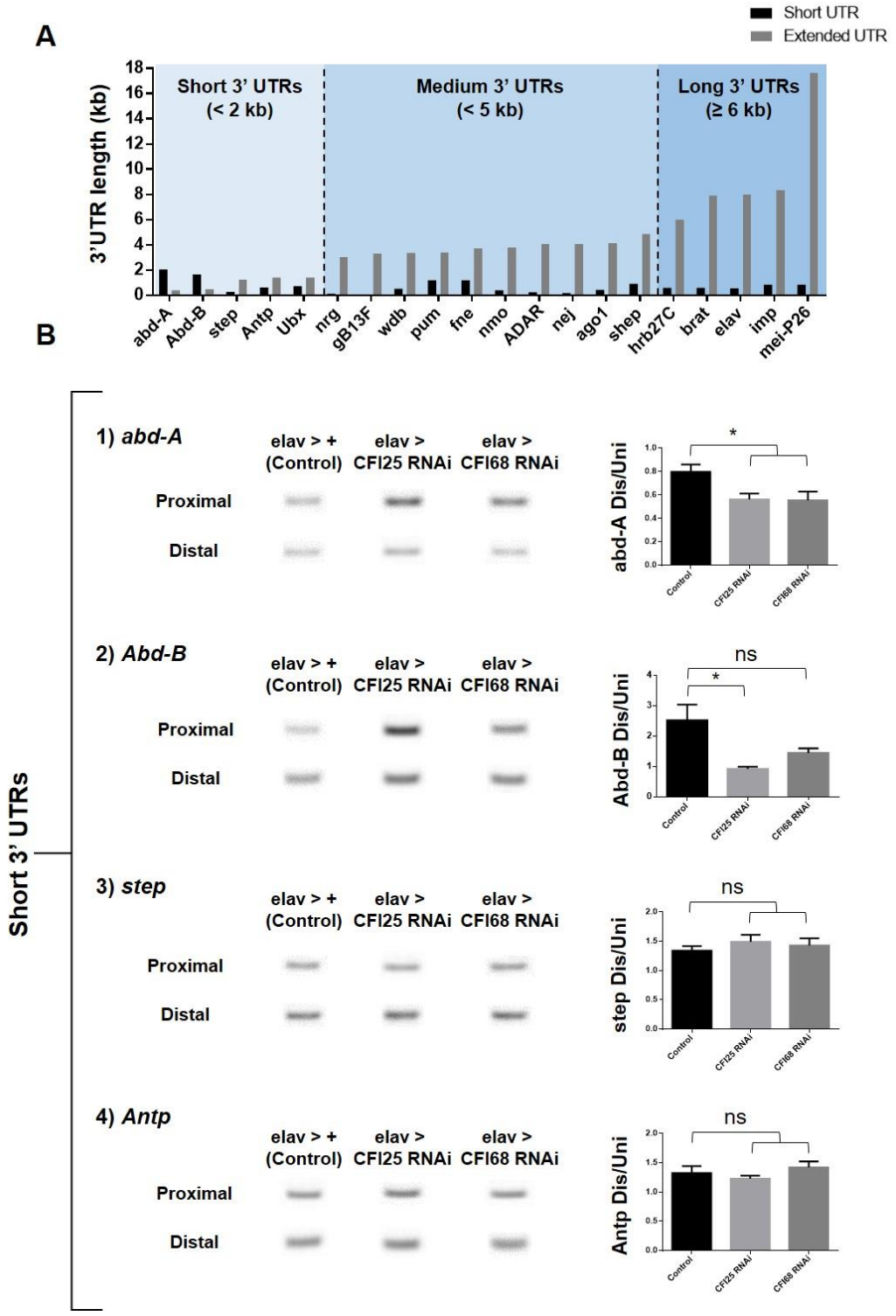


Figure 5.4 The knockdown of CFI25 and CFI68 within the developing nervous system affects APA in *abd-A* and *Abd-B*

(Legend on the following page)

Figure 5.4 The knockdown of CFI25 and CFI68 within the developing nervous system affects APA in *abd-A* and *Abd-B*

(A) Ranking and categories of *Hox* and selected neural-extended genes for knockdown experiments from (Hilgers *et al.* 2011) according to the curated length of their extended 3' UTRs. The curated length of their short 3' UTR is also shown for reference **(B)** "Dis/Uni" ratios of target genes in the knockdown of each one of the members of CFI within the nervous system in stage 16 embryos. Each gene was analysed in three independent biological replicates together with the control in which no RNAi was used (elav > +). For each graph, representative RT-PCR bands of the Universal and Distal amplicons are shown for each one of the genotypes. Note that the "Universal" amplicons *of abd-A* **(1)** and *Abd-B* **(2)** in the knockdown of both CFI25 and CFI68 are more intense than the control, which generates lower "Dis/Uni" ratios. Unpaired two-tailed t-tests were used to compare each RNAi against CFI25 or CFI68 with the control, * $p < 0.05$.

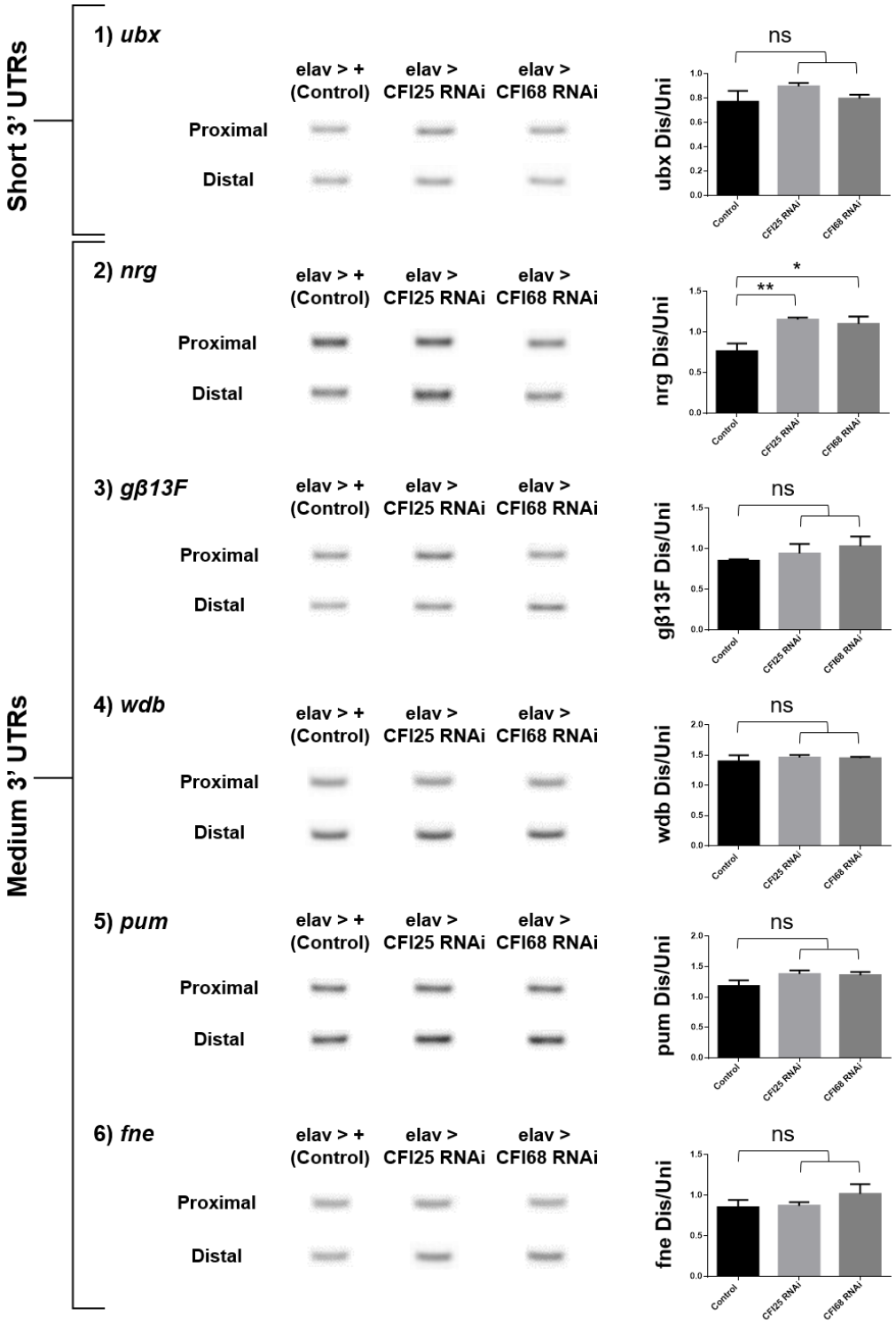


Figure 5.5 The knockdown of CFI25 and CFI68 within the developing nervous system affects APA in *nrg*

(Legend on the following page)

Figure 5.5 The knockdown of CFI25 and CFI68 within the developing nervous system affects APA in *nrg*

"Dis/Uni" ratios of target genes in the knockdown of each one of the members of CFI within the nervous system in stage 16 embryos. Each gene was analysed as in Figure 5.4. Note that the "Distal" amplicon of *nrg* **(2)** in the knockdown of CFI25 is more intense than the control. On the contrary, the "Universal" amplicon of this gene in the knockdown of CFI68 is less intense than the control. These changes generate significantly higher "Dis/Uni" ratios in the knockdowns when compared with the control. Unpaired two-tailed t-tests were used to compare each RNAi against CFI25 or CFI68 with the control, * $p < 0.05$.

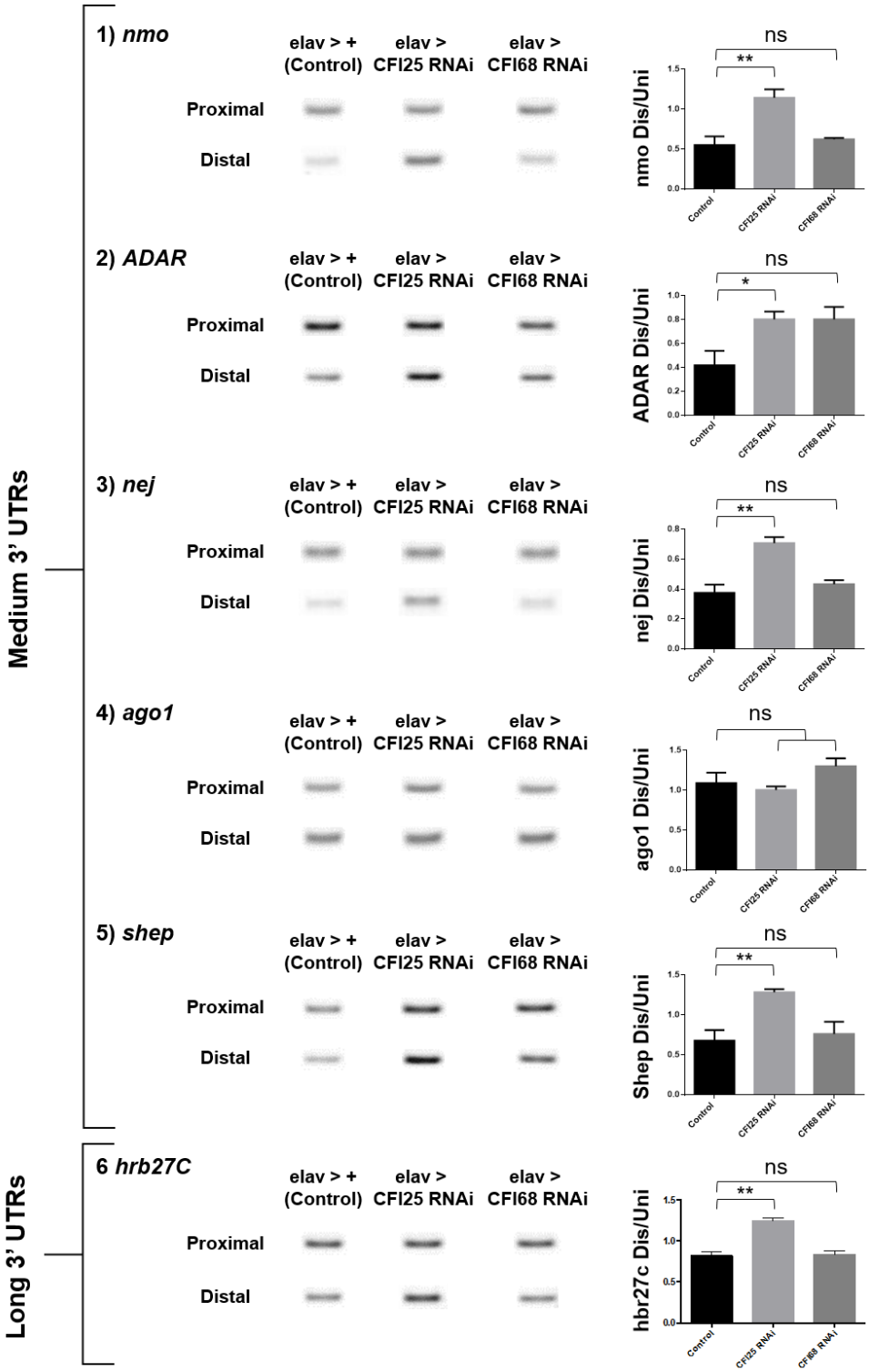


Figure 5.6 The knockdown of CFI25 within the developing nervous system affects APA in *nmo*, *ADAR*, *nej*, *shep*, and *hrb27C*

(Legend on the following page)

Figure 5.6 The knockdown of CFI25 within the developing nervous system affects APA in *nmo*, *ADAR*, *nej*, *shep*, and *hrb27C*

"Dis/Uni" ratios of target genes in the knockdown of each one of the members of CFI within the nervous system in stage 16 embryos. Each gene was analysed as in Figure 5.4. Note that the "Distal" amplicons of *nmo* **(1)**, *ADAR* **(2)**, *nej* **(3)**, *shep* **(5)** and *hrb27C* **(6)** in the knockdown of CFI25 are more intense than the control, which generates significantly higher "Dis/Uni" ratios in the knockdown when compared with the control. However, none of the knockdowns for CFI68 generates significant changes in the "Dis/Uni" ratio when compared with the control. Unpaired two-tailed t-tests were used to compare each RNAi against CFI25 or CFI68 with the control, * $p < 0.05$.

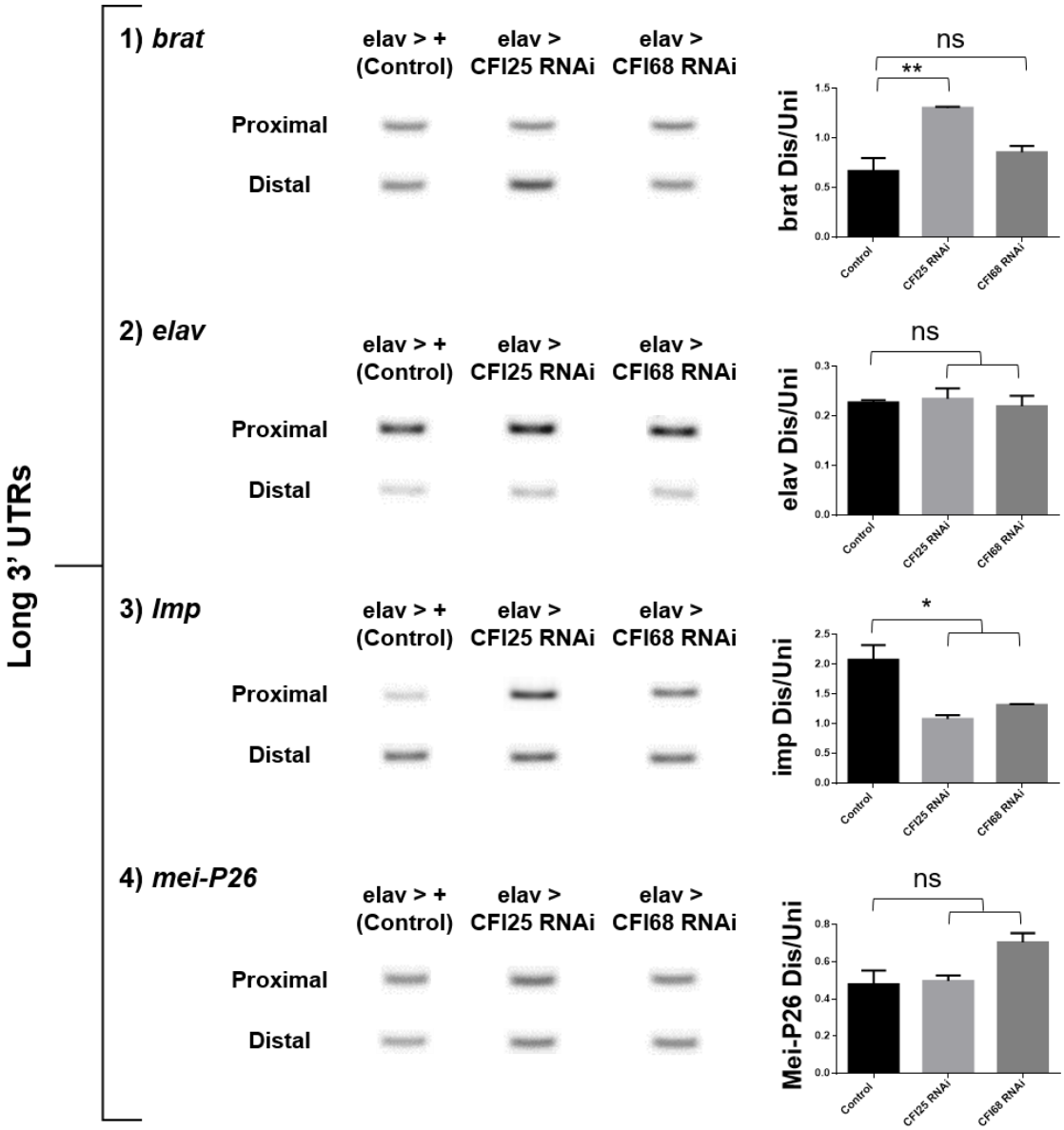


Figure 5.7 The knockdown of CFI25 within the developing nervous system affects APA in *brat*, while the knockdown of CFI25 and CFI68 affects APA in *imp*

(Legend on the following page)

Figure 5.7 The knockdown of CFI25 within the developing nervous system affects APA in *brat*, while the knockdown of CFI25 and CFI68 affects APA in *imp*

"Dis/Uni" ratios of target genes in the knockdown of each one of the members of CFI within the nervous system in stage 16 embryos. Each gene was analysed as in Figure 5.4. Note that the "Distal" amplicon of *brat* **(1)** in the knockdown of CFI25 is more intense than the control, which generates significantly higher "Dis/Uni" ratios. On the contrary, the "Universal" amplicon of *imp* **(3)** in the knockdown of both CFI25 and CFI68 within the nervous system is more intense than the control, which generates a lower "Dis/Uni" ratio when compared with the control. Unpaired two-tailed t-tests were used to compare each RNAi against CFI25 or CFI68 with the control, * $p < 0.05$.

of the *Hox* genes show effects in APA: *abd-A* and *Abd-B*. After the knockdown of both CFI25 and CFI68 within the developing nervous system, these posterior *Hox* genes undergo an increase in the production of the "Universal" 3' UTR amplicon (Figure 5.4), which is reflected in a significant decrease in their "Distal/Universal" ratios when compared with the control. Interestingly, these genes express the shortest extended 3' UTRs within the ranking, with extensions of only 413 and 503 bp for *abd-A* and *Abd-B*, respectively. These observations could mean that there is an increase in the production of short *abd-A* and *Abd-B* after CFI depletion. Within the "Medium 3' UTR" category, I observe that *neuroglian* (*nrg*), *nemo* (*nmo*), *ADAR*, *nejire* (*nej*) and *alan shepard* (*shep*) show an increase in the production of the "Distal" 3' UTR amplicon after the knockdown of CFI25 (Figures 5.5 and 5.6), while the knockdown of CFI68 generates a decrease in the production of the "Universal" 3' UTR amplicon in *nrg* (Figure 5.5). Both changes generate an increase in the "Distal/Universal" ratios. *nrg* is a cell-surface protein member of the Ig superfamily which is involved in axonal sprouting and dendrite branching within the *Drosophila* central and peripheral nervous (Bieber *et al*. 1989; Yamamoto *et al.* 2006), it expresses nine different mRNA isoforms and uses five different PAS to generate diverse APA isoforms, its shortest 3' UTR is only 128 bp long and only shared with two other of the mRNA isoforms, therefore not working as a true Universal 3' UTR. To design primers for the "Universal", I selected a 3' UTR of approximately 1 kb long shared among five mRNA isoforms. The primers for the "Distal" targeted the longest 3' UTR which is approximately 3 kb long and only present in one of the nine mRNA isoforms (for primer design see Chapter 2, Table 2.5). The effects observed for CFI68 therefore could be due to a decrease in the production of the selected short 3' UTR for PCR, leading to

an increase in the “Dis/Uni” ratio. Within the “Long 3’ UTR” category, *heterogeneous nuclear ribonucleoprotein at 27C* (*Hrb27C*) and *brain tumour* (*brat*) also show a significant increase in the production of the “Distal” 3’ UTR amplicon after the knockdown of CFI25 (Figure 5.6 and 5.7), the knockdown of CFI68 does not generate significant changes. On the other hand, *IGF-II mRNA-binding protein* (*imp*) shows a significant increase in the “Proximal” 3’ UTR amplicon after the knockdown of both CFI25 and CFI68, leading to a reduction in their “Distal/Universal” ratios, similarly to the case observed in the *Hox* genes (Figure 5.7). *Imp* is within the opposite range of 3’UTR lengths in the ranking, being the second longest extended 3’ UTR, with an extension of more than 8 kb (Figure 5.2). These results suggest that CFI depletion enhances the use of the distal PAS of *nrg*, *nmo*, *ADAR*, *nej* and *shep*, while this depletion in CFI enhances the use of the proximal PAS of *imp*, generating an overexpression of the short isoform.

In summary, I have observed and shown that 10 out of the 20 genes tested show changes in their APA patterns after CFI factor depletion in the nervous system. Furthermore, from these 10 affected genes, two belong to the ‘short’ 3’ UTR category (*abd-A* and *Abd-B*), five belonged to the ‘medium’ 3’ UTR category (*nrg*, *nmo*, *ADAR*, *nej* and *shep*) and three belonged to the ‘long’ 3’ UTR category (*hrb27C*, *brat* and *imp*). Naturally, this led to the question of why only a subset of genes showed significant effects in APA after neural CFI depletion as interpreted by the increase or decrease in their “Distal/Universal” ratios. In order to address this question, the next step was to look for a potential relationship between the length of their extended 3’ UTR and sensitivity to CFI factor depletion.

5.2.3 There is no clear relationship between 3' UTR length of target genes and effects in APA by knockdown of CFI factors within the developing nervous system

In the previous section I showed that there were a range of different responses in APA of 50% of the genes tested after the knockdown of CFI factors within the developing nervous system in stage 16 embryos. The two *Hox* genes with the shortest 3' UTR, *abd-A* and *Abd-B*, as well as the second longest 3'UTR gene, *imp*, showed an increase in the production of short 3' UTR isoforms after the knockdown of both CFI25 and CFI68 (Figure 5.4). This trend has been previously observed in other genes after inducing knockdown of CFI25 in human cells (Kubo *et al.* 2006). Seven genes within the "Medium" and "Long" 3' UTRs categories presented an increase in the production of the extended 3' UTR isoforms after the knockdown of CFI25. Effects of this nature are rare, yet still reported for several human genes in C2C12 cells after the knockdown of CFI members including *Casp8*, *Map3k3*, *Rab13*, *Fbxo8* and *Sox11*, among several other (Li *et al*. 2015)

The results observed after CFI depletion generate two immediate questions: (i) What are the reasons for the different directions in Dis/Uni ratios observed after CFI factor depletion in the nervous system? And (ii) Why only a subset of genes are sensitive to CFI depletion while others are not? To address the first question, I hypothesised that this is due to the length of the extensions of each of the 3' UTRs tested. Thus, an increase in sensitivity is expected as the length of the extension increases. To test this, I assigned a parameter named as "Δ APA", which is the quotient between the "Distal/Universal" ratio for each one of the RNAis and their respective controls. Thus, values close to 1 mean 'no difference',

values <1 mean ‘shortening', and >1 are ‘lengthening’ of 3’ UTRs. For all genes tested, Δ APA values against the length of their extended 3’ UTRs are shown in Figure 5.8. There is no statistically robust correlation between 3’ UTR length and the observed effects on APA. This observation suggests that it is not the extension of 3’ UTRs *per se* what makes these genes more or less sensitive to CFI depletion. Instead, sequence composition or other properties such as transcript stability or transcriptional input may be the cause. To investigate whether composition rather than overall 3’ UTR length can explain why only a subset of the tested genes are affected by CFI factor depletion, I scanned the 3’ UTR sequences of the genes affected by CFI depletion for enrichment of motifs in an unbiased way in comparison with the genes that were not affected by using MEME (Multiple Em for Motif Elicitation) (Bailey and Elkan 1994). Thus, the hypothesis to be tested is whether genes affected by CFI depletion show an enrichment in binding sites for CFI. The results of these experiments are discussed in the next section.

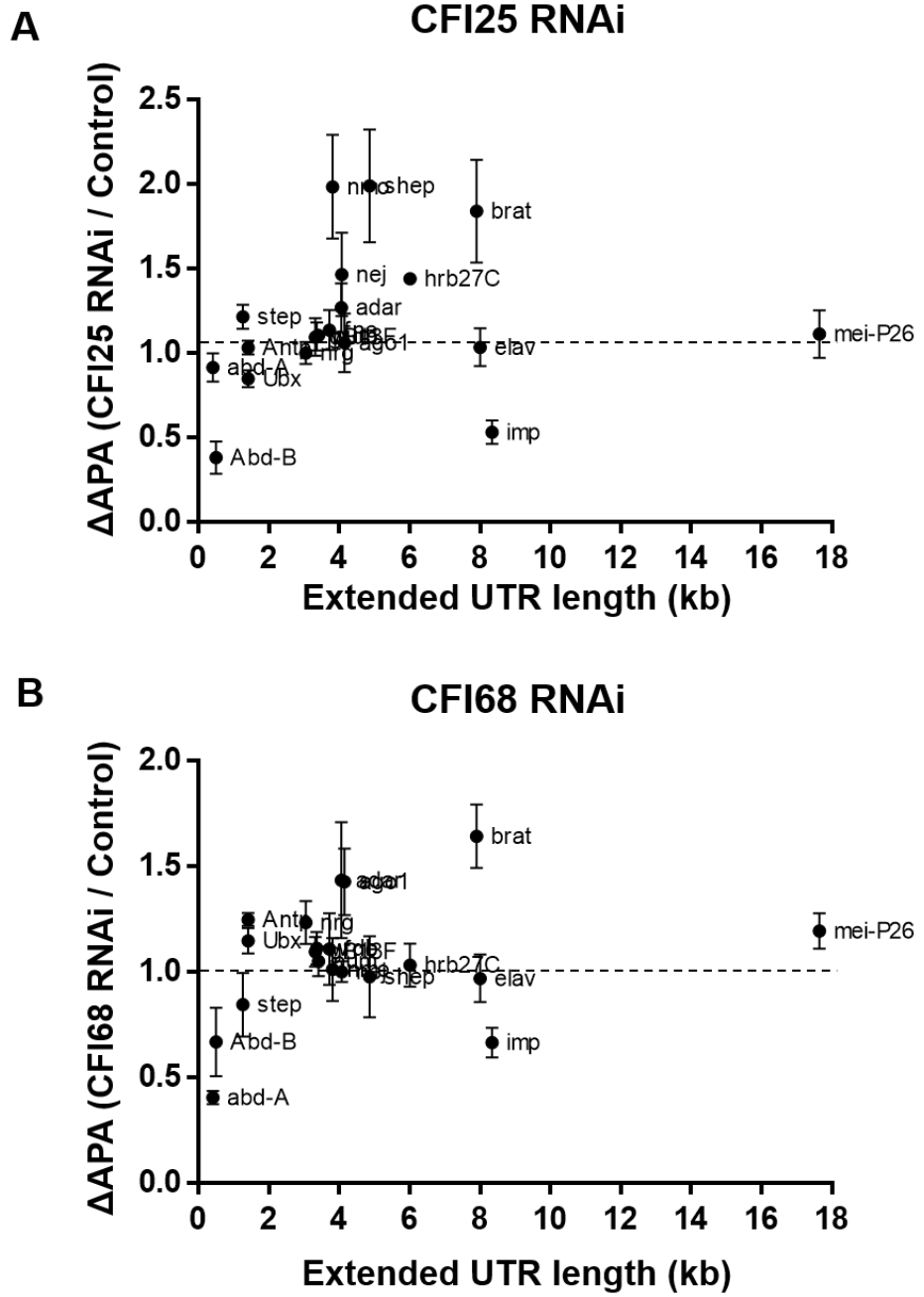


Figure 5.8 There is no clear relationship between 3' UTR length of target genes and effects in APA by knockdown of CFI factors within the developing nervous system

(Legend on the following page)

Figure 5.8 There is no clear relationship between 3' UTR length of target genes and effects in APA by knockdown of CFI factors within the developing nervous system

(A and B) Average of Δ APA values in 3 biological replicate experiments for the knockdown of CFI25 **(A)** and CFI68 **(B)**. Linear, logarithmic and polynomial curves were fitted through the data points with and without *mei-P26*, which given its extremely long 3' UTR can be considered and outlier within the axis. The best fit only gives a low multiple R-squared coefficient in the case of CFI25 RNAi (0.2, data not shown), which drops to half when is adjusted by sample size. A dotted line is drawn in the ratio with value 1 to represent the inflexion point in Δ APA. Error bars represent the SEM.

5.2.4 Bioinformatic analysis of motif enrichment in 3' UTRs of neural-extended genes

In order to verify whether it is sequence composition and not only 3' UTR length behind the reason why only 50% of the tested extended genes were sensitive to CFI depletion, I scanned their 3' UTR sequences for motif enrichment in an unbiased way by using discriminative analysis with MEME (Multiple Em for Motif Elicitation) (Bailey and Elkan 1994). This bioinformatic tool discovers ungapped motifs that are enriched in a given sequence by using statistical modelling techniques to automatically choose the best width, number of occurrences, and description for each motif. The hypothesis that drives these experiments is that this platform will be able to detect an enrichment in binding sites for CFI in the genes that were sensitive for CFI depletion. Furthermore, experiments in mammalian cells have shown that genes sensitive to CFI depletion have an enrichment in CFI binding motifs in their 3' UTRs (Li *et al*. 2015). However, contrary to the canonical "UGUA" element known in mammals, which was accurately described by SELEX experiments in 2003 (Brown and Gilmartin 2003), evidence suggests that insect pre-mRNAs use binding sites for CFI with a more degenerate sequence (Hutchins *et al*. 2008). In order to scan for motif enrichment in the 3' UTR of genes that were sensitive to CFI depletion, I extracted the sequences of the long 3' UTRs (Complete 3' UTR sequence) of the 20 genes tested from Flybase (Release from July 28, 2016) (Attrill *et al.* 2016). Then, MEME was used to scan for significantly (E-value < 0.05. The E-value is an estimate of the expected number of motifs with the same width and site count, that one would find in a similarly sized set of random sequences) enriched motifs in the long 3' UTRs of genes sensitive to CFI depletion compared with the long 3'

UTRs of genes that were not sensitive to CFI depletion. As a negative control, I performed the same comparison but with the shuffled sequences of the 3' UTRs. This last comparison, as expected, did not detect any significantly enriched motif, showing that the discovered enriched motifs are not the product of random combinations of nucleotides. The results of this analysis are shown in Figure 5.9. Two motifs were significantly enriched in the long 3' UTRs of the genes affected by neural CFI depletion when compared with the 3' UTR of the genes that were not affected. Interestingly, Motif 2 shows similarity with the mammalian sequences in its first nucleotides (Figure 5.9 Panel C) (Brown and Gilmartin 2003) and also an enrichment of 'T's that is consistent with the broad description of the insect CFI binding site, which in the mosquito *Anopheles gambiae* has been described as "AAAN(A/T)TTT" (Hutchins *et al.* 2008). Hence, these results show that genes that were sensitive to CFI depletion have an enrichment in this motif, which is proposed as the binding site for CFI. However, further evidence is needed to interpret these results as a true binding site for CFI. To address this, I mapped the position of motif 2 within the 3' UTRs of each one of the seven genes in which it was found by MEME: *nrg*, *abd-A*, *ADAR*, *hrb27C*, *imp*, *brat* and *Abd-B* and compared its location with respect to the location of the different polyadenylation signals (PAS). The binding site for CFI in mammals has been described as located 30-40 nucleotides upstream from the PAS (Brown and Gilmartin 2003; Georges Martin *et al*. 2012; Venkataraman *et al*. 2005). Although binding of CFI has also been reported within 100 nucleotides upstream of the cleavage site in mammals (Martin *et al.* 2012). In insects, the position

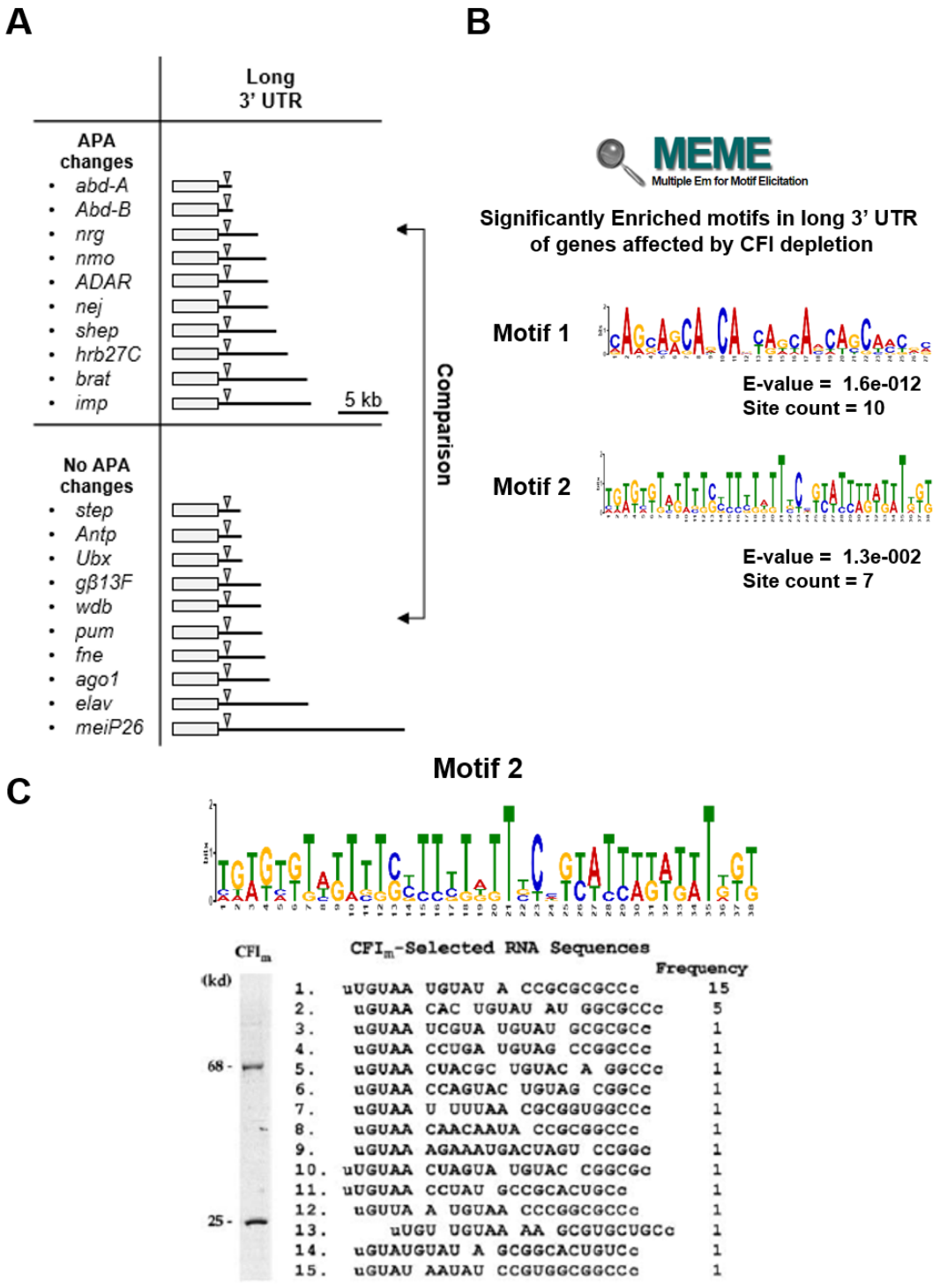


Figure 5.9 Bioinformatic analysis of motif enrichment in 3' UTRs of neural-extended genes affected by CFI depletion

(Legend on the following page)

Figure 5.9 Bioinformatic analysis of motif enrichment in 3' UTRs of neural extended genes affected by CFI depletion

(A) Table showing the categories used for analysis with simplified gene models representing the length the distal 3' UTRs. The complete 3' UTR sequences were used for comparison between genes with affected APA patterns after CFI depletion with the genes not affected, depicted by the vertical double arrow. **(B)** Significant motif enrichment analysis by MEME after comparison of the long 3' UTRs showed in **(A),** Motif 2 is selected for further analysis. The same analysis done with the shuffled 3' UTRs does not detect enrichment (Data not shown) **(C)** Magnification of motif 2 from **(B)**, the table represents the results from SELEX experiments as done in (K. M. Brown and Gilmartin 2003), note that 'T' is used instead of 'U' because 3' UTR sequences from Flybase are retrieved as DNA code. Note that the first nucleotides of Motif 2 show similarity with the mammalian sequences. Also, the high enrichment of 'T' is consistent with the broad description of the insect CFI binding site from (Hutchins *et al.* 2008).

of this binding site has been reported within the range of 50 – 100 nucleotides upstream of the PAS (Hutchins *et al.* 2008). Hence, I scanned the position of Motif 2 in the sensitive genes with respect to the location of the different PASs. In this way, if the motif is (i) found upstream of the PAS and (ii) at a distance of 50-100 nucleotides, this will suggest that Motif 2 acts as a true CFI binding motif in that gene. The results of this analysis are shown in Figure 5.10. I observe that these two conditions are achieved only for *Abd-B*, in which motif 2 is found 144 bp upstream of the last PAS. In the other cases, Motif 2 is found upstream of the second PAS of *nrg*, the first PAS of *abd-A* and the last PAS of *ADAR*, *hrb27C*, *imp* and *brat*. However, its location is too far away from the expected distance between this motif and the PAS to act as the binding site of CFI in these genes. These results show that this enrichment in Motif 2 can only partially explain why this subset of genes is sensitive to CFI depletion and suggests that potentially there are other mechanisms involved, which will be discussed in the next section.

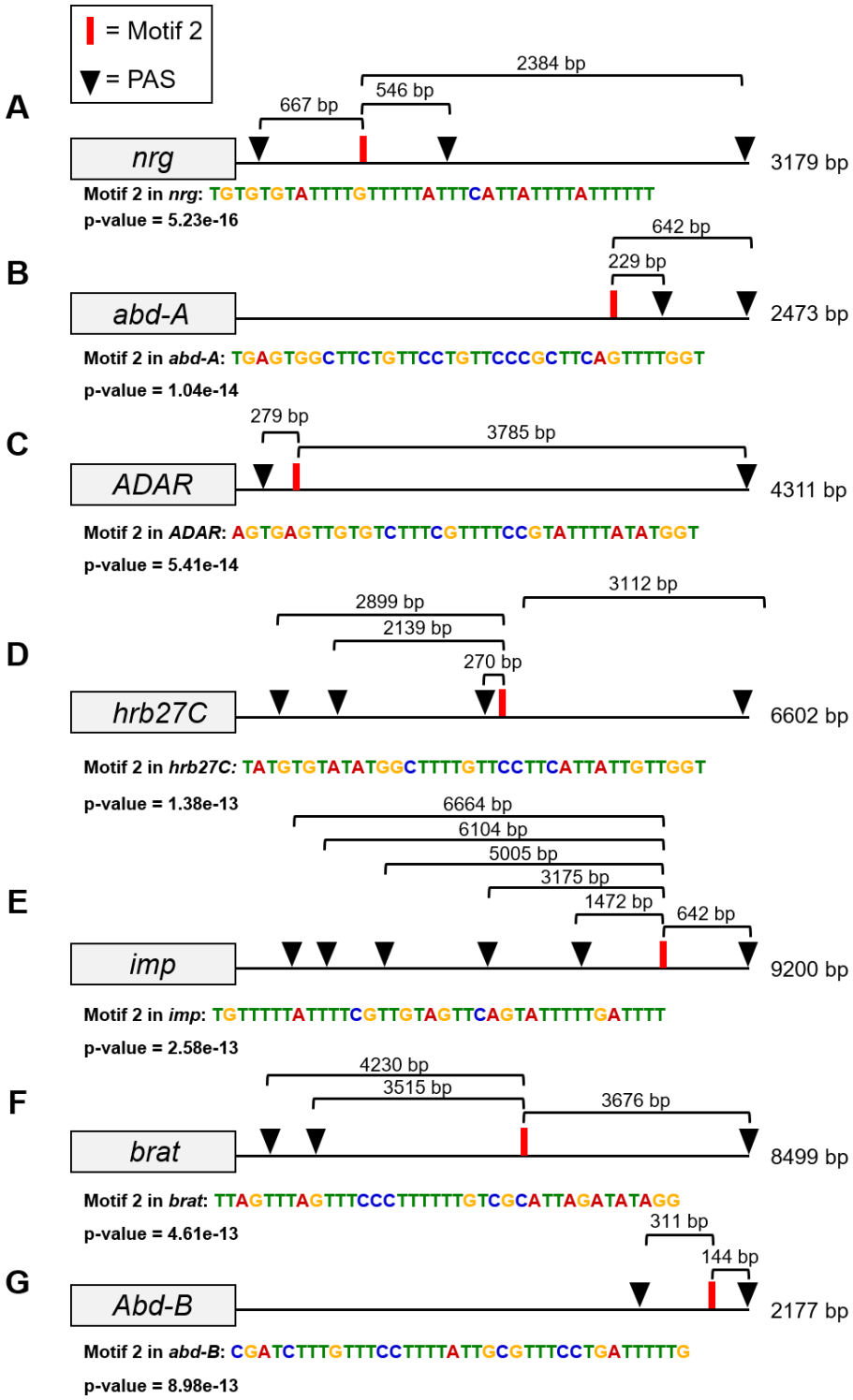


Figure 5.10 Location of Motif 2 with respect to the PAS in the 3' UTRs of genes affected by CFI depletion

(Legend on the following page)

Figure 5.10 Location of Motif 2 with respect to the PAS in the 3' UTRs of genes affected by CFI depletion

(A-G) Gene diagrams of *nrg* **(A)**, *abd-A* **(B)**, *ADAR* **(C)**, *hrb27C* **(D)**, *imp* **(E)**, *brat* **(F)** and *Abd-B* **(G)**, ranked from top to down according to the p-values for motif 2 enrichment (p-values represent the probability that an equal or better site would be found in a random sequence of the same length conforming to the background letter frequencies). The grey box represents the coding sequence of each gene, the 3' UTR is represented as a black line with its length indicated at the end for each gene, the location of the PAS is depicted by a black inverted triangle and the location of Motif 2 is depicted by a red square. The distance from Motif 2 to each one of the PAS, as well as the specific sequence of Motif 2 in each gene are shown. Note that only in the case of *Abd-B* **(G)**, the location of Motif 2 is consistent with it acting as a true CFI binding motif given its location upstream of the PAS and the distance (144 bp). These results suggest that motif 2 can only partially explain the sensitivity to CFI depletion in these target genes and that other mechanisms may be involved.

5.3 Discussion

The work presented in this chapter shows that the neural knockdown of CFI25 and CFI68 in late embryos affects APA patterns in a subset of genes reported for neural extensions in 3' UTRs (Hilgers *et al.* 2011). Relatedly, the annotation of the length of these genes does not match the current databases, and while the general change in length is seen as sequences shorter than previously annotated, specific cases for longer sequences are observed (Table 5.1), with *mei-P26* being the most dramatic example by having a curated extended 3' UTR of 17.650 kb, similarly to mammalian transcripts (Miura *et al*. 2013). As mentioned earlier, the biological roles played by these extensively long 3' UTRs remain unknown, given that differential miRNA and RBP targeting does not require such extensions to be achieved. This is what happens in the *Hox* genes (Patraquim *et al.* 2011). One scenario could be that extensively long 3' UTRs play roles beyond post-transcriptional regulation, by tethering transcripts to their DNA templates and releasing them in a timely-controlled way that can complement transcriptional control by promoters and enhancers. Another potential scenario is that this extensively long 3' UTR may serve as transcriptional checkpoints, where pre-mRNAs require assessment by different molecular machineries such as the ones controlling splicing, 5' capping and non-sense mediated RNA decay before finally releasing the transcripts for translation. These processes ensure that the produced proteins work in an optimal way. This hypothesis is consistent with the fact that hundreds of neural transcripts in *Drosophila* (Smibert *et al.* 2012) and thousands in mammals (Miura *et al.* 2013) undergo these processes, given that the complexity of the nervous system may require extra layers of gene expression checkpoints to ensure functionality.

Here I have also shown that neural-specific depletion of CFI factors within the developing nervous system in late embryos can affect APA patterns in reported neural-extended genes (Hilgers *et al*. 2011). Interestingly, while *abd-A*, *Abd-B* and *Imp* showed an increase in the production of proximal amplicon and therefore a reduction in their "Distal/Universal" ratios after CFI knockdown. *Nrg*, *nmo*, *ADAR*, *nej*, *shep*, *hrb27C* and *brat* show an increase in the production of the distal amplicon and therefore an increase in their "Distal/Universal" ratios after CFI knockdown. These results are not simply explainable by 3' UTR length, given that we cannot see any clear relationship between these two variables (Figure 5.8). Likewise, after analysing the full 3' UTR sequences of the genes affected by CFI depletion, a motif was found that resembles the binding site of CFI, according to comparison with descriptions in mammals (Brown and Gilmartin 2003) and insects (Hutchins *et al.* 2008). Nonetheless, this motif only fulfils the conditions to function as the binding site of CFI in the 3' UTR of *Abd-B*, given that is located 144 bp upstream of the second PAS (Figure 5.10 Panel G). One way of further testing if this and other discovered motifs function as true binding sites for CFI is to use mutagenesis. For example, by applying CRISPR (Hsu *et al*. 2014; Bassett and Liu 2014) to mutate these sites and then analyse whether the patterns of APA of these genes during neural development are affected.

The results observed in this study suggest that other mechanisms may be involved in the differential sensitivity of neural-extended genes to CFI depletion. For example, differential transcript stability or transcriptional input. Further exploration of the different 3' UTRs of all the genes that undergo neural-extensions in combination with manipulation of CPA factor expression during

neural development can be therefore a good way of exploring the mechanisms that control APA in this developmental context in *Drosophila*.

Most of the effects observed in our knockdown experiments produced an increase in the Dis/Uni ratio. Although effects of this nature are rare, they have been reported for several genes in human cells (Li *et al.* 2015). In the literature, both CFI25 and CFI68 have been described broadly as repressors of proximal PAS (Martin *et al*. 2012), this means that when their expression levels are downregulated, they favour the use of proximal PAS, which leads to a global shortening of 3' UTRs. The effects we observe for *abd-A*, *Abd-B* and *imp* are totally in line with this model, because an upregulation of the proximal 3' UTRs is observed (Figure 5.4 and 5.7). Given that in this study we only used 20 specific genes, we cannot rule out the possibility that the low sampling size when compared with the previous studies generated a bias in our interpretation, by detecting outliers in an otherwise opposite trend. One way of expanding these experiments is to subject the samples to transcriptome-wide analysis as the ones described previously (Li *et al*. 2015; Masamha *et al.* 2014) and to purify RNA that comes specifically from neurons. This could be achieved by using TU-tagging, which is a technique used in *Drosophila* to isolate cell type-specific RNA from complex tissues by expressing uracil phosphoribosyltransferase (UPRT) in a tissue specific manner. Then, its substrate 4-thiouracil is delivered to the organism by soaking the embryos in it, generating Thio-RNA only in the cells expressing UPRT, which can be subsequently purified (Miller *et al*. 2009).

In summary, here I have shown that the expression levels of CFI factors can control APA within the nervous system. This suggests that the observed enrichment of CFI expression in this tissue (shown in Chapter 3 Figures 3.10 and

3.11) can be directly linked to the reported 3' UTR extensions. This evidence is similar to that which was shown *in vitro.* Future experiments to establish this model can be the ectopic expression of CFI members in non-neural tissues by using the Gal4-UAS system (Brand and Perrimon 1993) to see if there is a shift at the transcriptomic level by genes acquiring "neural" profiles of APA. This same approach can be used to boost the expression of CFI factors within the nervous system to test the model of CPA factor abundance and PAS selection proposed by James Manley and colleagues (Takagaki *et al*. 1996), where we would expect the opposite results as those observed by RNAi knockdown as have been shown here.

Chapter 6

General Discussion

6.1 General discussion

The work presented in this thesis provides novel insights on the regulation of APA during neural development and highlights the biological roles of CPA factors at the organismal level beyond their well-established molecular function by studying them in the context of the *in vivo* formation and function of the nervous system in *Drosophila melanogaster*.

APA is a pervasive mechanism of post-transcriptional control that generates mRNAs with different 3' ends in more than 60% of all human and mammalian genes (Derti *et al.* 2012; Hoque *et al.* 2013) (Chapter 1). As well as in approximately half of all genes in other vertebrates, such as zebrafish, and invertebrates, such as *C. elegans* and *Drosophila* (Ulitsky *et al.* 2012; Smibert *et al.* 2012; Jan *et al.* 2011). This selection of different 3' UTRs can have different biological functions, including mRNA localization, stability and translation efficiency, having a direct impact on the diversity of the transcriptome and the proteome (Di Giammartino *et al*. 2011). Interestingly, the expression of long 3' UTR isoforms during development in hundreds of genes in *Drosophila* and thousands of genes in mammals is restricted to the nervous system (Thomsen *et al.* 2010; Smibert *et al.* 2012; Miura *et al.* 2013), in which differential targeting by miRNAs and RBPs have been the main purposes explained in the literature for their biological function within neural tissues.

The molecular mechanisms that control APA are not well understood. Experiments *in vitro* in mouse B-Lymphocytes showed that the IgM heavy chain gene produces two mRNA isoforms by APA: A membrane bound protein or a secreted one (Takagaki *et al*. 1996). This shift between the two isoforms was

shown to be a product of the abundance of the CPA factor CSTF64, in which high levels of CSTF64 favoured the use of the proximal PAS and therefore a secreted protein, while low levels of CSTF64 favoured the use of a distal PAS and therefore a membrane-bound protein. This evidence led to a mechanistic model that is prevalent today as one of the molecular mechanisms that control APA by means of CPA factor abundance and differential PAS strength. Nonetheless, experiments *in vitro* present limitations in which we cannot reproduce what is observed in multicellular organisms, let alone to understand the mechanisms by which these extensively long 3' UTRs are expressed within a developing nervous system.

In this thesis I focused on the molecular mechanisms that control APA during neural development, as well as the biological relevance of CPA factors for neural formation and function. The work was centred on the following questions: (i) What are the molecular factors that control APA during *Drosophila* development? (ii) Does the nervous system express different levels of CPA factors in comparison to other tissues to express comparatively long 3' UTRs during development? (iii) Can the manipulation on CPA factor within the developing nervous system affect APA in neural extended genes and so provide evidence for the "CPA factor abundance" model in this novel context? And (iv) what are the biological roles of CPA factors for neural function and behaviour? To address these questions I used *Drosophila melanogaster* as a novel model-system to study APA in contrast with work *in vitro*. Considering therefore the physiological context of neural development.

6.2 Cleavage and polyadenylation factor expression and extension of 3' UTRs during neural development

As mentioned earlier, since 2010 it has been reported that the nervous system in both vertebrates and invertebrates undergoes an extensive lengthening of 3' UTRs in the nervous system (Thomsen *et al.* 2010; Hilgers *et al.* 2011; Smibert *et al.* 2012; Miura *et al.* 2013). Yet, these studies have not explored the mechanisms by which this phenomenon is achieved by the CPA machinery.

A study in 2013 addressed the question on what is the biological meaning of variation in 3' UTR length in different human tissues. They show that ubiquitously expressed genes use APA to achieve tissue-specific expression of these 3' UTR isoforms, which are then targeted by ubiquitously expressed miRNAs to act on them differentially (Lianoglou *et al.* 2013). Still, one key question that emerges from this study is: What is the mechanism by which these tissues express long 3' UTR isoforms in the first place? This question can be addressed by analysing CPA factor expression levels in the tissues sampled and looking for a correlation that can be experimentally tested.

This work shows that in *Drosophila* there is a relationship between tissue-specific CPA factor expression and neural 3' UTR length in which neural tissues express higher levels of CPA factors in comparison with others during embryonic and larval life (Chapter 3). Also, the most conserved CPA complex between humans and *Drosophila*, CFI, which is also the most neural-enriched CPA complex in *Drosophila* larvae and adults, shows a restricted expression pattern in the embryonic nervous system at the time where extensively long 3' UTRs are reported (Chapter 3). This work also shows that variations in the expression

levels of the members of CFI during this stage by RNAi knockdown have a knock-on effect on neural-extended genes APA (Chapter 5), strongly suggesting that this neural-enrichment of CFI factor expression act as a cue within neurons to express these long 3' UTR isoforms. This takes the "CPA factor abundance" model to a novel developmental context that has not been explored before in the field. These results encourage us to take this approach to a genome-wide transcriptomic level by which we can explore further the mechanisms by which neural-3' UTR extensions are achieved involving also more members of the identified *Drosophila* CPA machinery (Chapter 3). Further, *Drosophila* allows the use of well-established molecular tools that have not been applied in the literature to address the question on the molecular mechanisms that control APA. For example, by overexpressing subsets of CPA factors in different tissues or by expressing synthetic constructs with collections of experimental 3' UTRs to then study their APA effects after genetic manipulation of CPA factor abundance. Thus, experiments like the ones proposed above can allow us to use the embryonic nervous system or other tissues with biased 3' end isoforms as test tubes to study this pervasive post-transcriptional modification.

6.3 Biological roles of cleavage and polyadenylation factors

When genes with well-established molecular functions are studied in a multicellular organism in different contexts, including behaviour, development or physiology, the phenotypes that are observed in mutant conditions at times can be remarkable. A good example of this is the miRNA locus *iab4/8*. This miRNA locus is located within the Bithorax complex in *Drosophila* and controls the expression of Hox genes located in this region, including *Ubx*. Interestingly,

although somehow expected, its ectopic expression in the halteres can lead to homeotic transformations from halteres to wings through precisely effects in *Ubx.* (Ronshaugen *et al.* 2005). Nevertheless, recent work has shown that this same miRNA locus can affect larval behaviour dramatically, in which a mutation in this miRNA locus makes the larvae unable to correct its position when turned upside-down. This behaviour is known as “Self-righting” (Picao-Osorio *et al*. 2015). The authors show that the phenotype is caused by a derepression of its target *Ubx* in two metameric neurons of the central nervous system and therefore uncovers new behavioural roles for miRNAs, a family of transcriptional regulators that are generally not well described in their relation with behaviour.

This work shows that a mutation in the *Drosophila* orthologue of CFI25 leads to defects in larval feeding, which causes a developmental arrest at the L1 stage in these mutants (Chapter 4). This phenotype at the anatomical level is caused by defects in the morphology of the CPS as well as defects in food ingestion. At the molecular level, four feeding-related genes are affected in the mutant condition (*Ranbpm*, *S6k*, *lov* and *klu*) and can potentially link the observed phenotype with CFI25. Previous studies have related a small subset of CPA factors with other biological roles. For example, mutations in the *Drosophila* orthologue of PAP leads to defects in the patterning of the adult wing margin. Although the exact molecular mechanism was not uncovered in these studies, the authors offered evidence for PAP as an important regulator of the Notch signalling pathway in *Drosophila* (Murata *et al*. 2001; Murata and Ogura 1996). These observations demonstrate that, similarly as miRNA *iab 4/8*, whose molecular function is thoroughly described in the literature (Ronshaugen *et al*. 2005; Pease *et al*. 2013; Bender 2008; Tyler *et al*. 2008), novel roles for larval behaviour can still be

unexpectedly uncovered. Consequently, this work suggests that the CPA machinery in *Drosophila* is potentially another system to be explored in this context, which has been largely ignored as a result of doing experiments *in vitro*, where this fascinating aspect cannot be studied.

6.4 Concluding remarks

In conclusion, this thesis explores different aspects of the control of APA by abundance of core CPA factors, it also explores the biological function of these factors for the formation and function of the nervous system. I present a novel experimental approach to study the control of APA in the context of multicellularity and development, by which new research in the field can take into account this layer of complexity to address the mechanisms by which APA operates. These findings advance therefore the current understanding of the molecular mechanisms that control APA during neural development and identifies one of potentially many biologically relevant roles played by the CPA machinery for behaviour.

Appendix

Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals

This appendix aims to present a collaboration in which I participated together with my supervisor (Dr. Claudio Alonso) between our laboratory and Dr. Javier Caceres from the Institute of Genetics & Molecular Medicine at the University of Edinburgh (http://www.hgu.mrc.ac.uk/index.html) in the field of RNA processing in eukaryotes. This work was published in 2014 in “EMBO Reports” (Casadio *et al.* 2014) and my specific contribution is explained in the following section. Additionally, the published manuscript is attached in this thesis.

nompA as a novel NMD factor in *Drosophila*

NMD is a cellular surveillance mechanism that degrades transcripts with premature stop codons, avoiding the translation of truncated proteins which could be harmful for the cell, making it a relevant process for animal development (Metzstein and Krasnow 2006; Alonso 2005).

The core factors involved in NMD in *Drosophila* have been well described in the literature (Alonso 2003; Avery *et al.* 2011). Nevertheless, a new screening for novel NMD factors in the model organism *Caenorhabditis elegans* shows that *noah-2* is a worm novel NMD factor which has an ortholog with arthropods, but not with vertebrates. In *Drosophila,* the orthologue of *noah-2* is *nompA* (Chung *et al.* 2001), a peripheral nervous system (PNS) specific protein. Our study offers evidence for *nompA* as a novel NMD factor by using a GFP-NMD reporter (Metzstein and Krasnow 2006). This contribution is shown in figure 3 of the attached manuscript.

References

Adams, Md, Se Celniker, Ra Holt, Ca Evans, Jd Gocayne, Pg Amanatides, Se Scherer, et al. 2000. “The Genome Sequence of Drosophila Melanogaster.” *Science* 287 (5461): 2185–95. doi:10.1126/science.287.5461.2185.

Adelman, Karen, Wenxiang Wei, M Behfar Ardehali, Janis Werner, Bing Zhu, Danny Reinberg, and John T Lis. 2006. “Drosophila Paf1 Modulates Chromatin Structure at Actively Transcribed Genes.” *Molecular and Cellular Biology* 26 (1): 250–60. doi:10.1128/MCB.26.1.250-260.2006.

Agrawal, Neha, Nisha Padmanabhan, and Gaiti Hasan. 2009. “Inositol 1,4,5-Trisphosphate Receptor Function in Drosophila Insulin Producing Cells.” *PLoS ONE* 4 (8): 1–10. doi:10.1371/journal.pone.0006652.

Alonso, Claudio R. 2005. “Nonsense-Mediated RNA Decay: A Molecular System Micromanaging Individual Gene Activities and Suppressing Genomic Noise.” *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 27 (5): 463–66. doi:10.1002/bies.20227.

Alonso, Claudio R. 2003. “A Hox Gene Mutation That Triggers Nonsense-Mediated RNA Decay and Affects Alternative Splicing during Drosophila Development.” *Nucleic Acids Research* 31 (14): 3873–80. doi:10.1093/nar/gkg482.

Altschul, Stephen F, Thomas L Madden, Alejandto A Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J Lipman. 1997. “Gapped BLAST and PSI-BLAST:a New Generation of Protein Database Search Programs.” *Nucleic Acids Res* 25 (17): 3389–3402.

An, Juan Ji, Kusumika Gharami, Guey Ying Liao, Newton H. Woo, Anthony G. Lau, Filip Vanevski, Enrique R. Torre, et al. 2008. “Distinct Role of Long 3’ UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons.” *Cell* 134 (1): 175–87. doi:10.1016/j.cell.2008.05.045.

Aoyagi, Norikazu, and David A. Wassarman. 2000. “Genes Encoding Drosophila Melanogaster RNA Polymerase II General Transcription Factors: Diversity in TFIIA and TFIID Components Contributes to Gene-Specific Transcriptional Regulation.” *Journal of Cell Biology* 150 (2): 45–49. doi:10.1083/jcb.150.2.F45.

Armstrong, J. D., M. J. Texada, R. Munjaal, D. A. Baker, and K. M. Beckingham. 2006. “Gravitaxis in Drosophila Melanogaster: A Forward Genetic Screen.” *Genes, Brain and Behavior* 5 (3): 222–39. doi:10.1111/j.1601-183X.2005.00154.x.

Attrill, Helen, Kathleen Falls, Joshua L Goodman, Gillian H Millburn, Giulia Antonazzo, Alix J Rey, Steven J Marygold, and the FlyBase FlyBase Consortium. 2016. “FlyBase: Establishing a Gene Group Resource for Drosophila Melanogaster.” *Nucleic Acids Research* 44 (D1): D786-92. doi:10.1093/nar/gkv1046.

Audibert, Agnès, and Martine Simonelig. 1999. “The Suppressor of Forked Gene of Drosophila, Which Encodes a Homologue of Human CstF-77K Involved in mRNA 3’-end Processing, Is Required for Progression through Mitosis.” *Mechanisms of Development* 82 (1–2): 41–50. doi:10.1016/S0925-4773(99)00011-8.

Avery, Paul, Marta Vicente-Crespo, Deepthy Francis, Oxana Nashchekina, Claudio R Alonso, and Isabel M Palacios. 2011. “Drosophila Upf1 and Upf2 Loss of Function Inhibits Cell Growth and Causes Animal Death in a Upf3-Independent Manner.” *RNA (New York, N.Y.)* 17 (4): 624–38. doi:10.1261/rna.2404211.

Awasthi, Sita, and James C Alwine. 2003. “Association of Polyadenylation Cleavage Factor I with U1 snRNP.” *RNA*, 1400–1409. doi:10.1261/rna.5104603.berson.

Bagga, P S, L P Ford, F Chen, and J Wilusz. 1995. “The G-Rich Auxiliary Downstream Element Has Distinct Sequence and Position Requirements and Mediates Efficient 3’ End Pre-mRNA Processing through a Trans-Acting Factor.” *Nucleic Acids Research* 23 (9): 1625–31. doi:4t0727 [pii].

Bai, Yun, Thierry C. Auperin, Chi Yuan Chou, Gu Gang Chang, James L. Manley, and Liang Tong. 2007. “Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors.” *Molecular Cell* 25 (6): 863–75. doi:10.1016/j.molcel.2007.01.034.

Bailey, T. L, and C. Elkan. 1994. “Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers.” *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36. doi:citeulike-article-id:878292.

Balbo, Paul B., and Andrew Bohm. 2007. “Mechanism of Poly(A) Polymerase: Structure of the Enzyme-MgATP-RNA Ternary Complex and Kinetic Analysis.” *Structure* 15 (9): 1117–31. doi:10.1016/j.str.2007.07.010.

Banerjee, Santanu, Rohit Joshi, Gayatri Venkiteswaran, Neha Agrawal, Sonal Srikanth, Farhan Alam, and Gaiti Hasan. 2006. “Compensation of Inositol 1,4,5-Trisphosphate Receptor Function by Altering Sarco-Endoplasmic Reticulum Calcium ATPase Activity in the Drosophila Flight Circuit.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 26 (32): 8278–88. doi:10.1523/JNEUROSCI.1231-06.2006.

Bassett, Andrew R., and Ji Long Liu. 2014. “CRISPR/Cas9 and Genome Editing in Drosophila.” *Journal of Genetics and Genomics* 41 (1). Elsevier Limited and Science Press: 7–19. doi:10.1016/j.jgg.2013.12.004.

Bateman, Alex, Maria Jesus Martin, Claire O’Donovan, Michele Magrane, Rolf Apweiler, Emanuele Alpi, Ricardo Antunes, et al. 2015. “UniProt: A Hub for Protein Information.” *Nucleic Acids Research* 43 (D1): D204–12. doi:10.1093/nar/gku989.

Beckervordersandforth, Ruth M., Christof Rickert, Benjamin Altenhein, and Gerhard M. Technau. 2008. “Subtypes of Glial Cells in the Drosophila Embryonic Ventral Nerve Cord as Related to Lineage and Gene Expression.” *Mechanisms of Development* 125 (5–6): 542–57. doi:10.1016/j.mod.2007.12.004.

Belay, A.T, R Scheiner, A.K.-C SO, S.J Douglas, M Chakaborty-Chatterjee, J.D Levine, and M.B Sokolowski. 2007. “The Foraging Gene of Drosophila Melanogaster: Spatial-Expression Analysis and Sucrose Responsiveness.” *The Journal of Comparative Neurology* 504 (December 2006): 287–97. doi:10.1002/cne.

Bellen, Hugo J., Robert W. Levis, Guochun Liao, Yuchun He, Joseph W. Carlson, Garson Tsang, Martha Evans-Holm, et al. 2004. “The BDGP Gene Disruption Project: Single Transposon Insertions Associated with 40% of Drosophila Genes.” *Genetics* 167 (2): 761–81. doi:10.1534/genetics.104.026427.

Bellen, Hugo J, Robert W Levis, Yuchun He, Joseph W Carlson, Martha Evans-Holm, Eunkyung Bae, Jaeseob Kim, et al. 2011. “The Drosophila Gene Disruption Project: Progress Using Transposons with Distinctive Site Specificities.” *Genetics* 188 (3): 731–43. doi:10.1534/genetics.111.126995.

Bender, Welcome. 2008. “MicroRNAs in the Drosophila Bithorax Complex.” *Genes & Development* 22 (1): 14–19. doi:10.1101/gad.1614208.

Benoit, Béatrice, Géraldine Mitou, Aymeric Chartier, Claudia Temme, Sophie Zaessinger, Elmar Wahle, Isabelle Busseau, and Martine Simonelig. 2005. “An Essential Cytoplasmic Function for the Nuclear poly(A) Binding Protein, PABP2, in poly(A) Tail Length Control and Early Development in Drosophila.” *Developmental Cell* 9 (4): 511–22. doi:10.1016/j.devcel.2005.09.002.

Berger, Christian, Heike Harzer, Thomas R Burkard, Jonas Steinmann, Suzanne van der Horst, Anne-Sophie Laurenson, Maria Novatchkova, Heinrich Reichert, and Juergen a Knoblich. 2012. “FACS Purification and Transcriptome Analysis of Drosophila Neural Stem Cells Reveals a Role for Klumpfuss in Self-Renewal.” *Cell Reports* 2 (2): 407–18. doi:10.1016/j.celrep.2012.07.008.

Berger, Jurg, Kirsten-Andre Senti, Gabriele Senti, Timothy P. Newsome, Bengt Asling, Barry J. Dickson, and Takashi Suzuki. 2008. “Systematic Identification of Genes That Regulate Neuronal Wiring in the Drosophila Visual System.” *PLoS Genetics* 4 (5). doi:10.1371/journal.pgen.1000085.

Berkovits, Binyamin D., and Christine Mayr. 2015. “Alternative 3′ UTRs Act as Scaffolds to Regulate Membrane Protein Localization.” *Nature* 8. doi:10.1038/nature14321.

Berni, Jimena, Stefan R. Pulver, Leslie C. Griffith, and Michael Bate. 2012. “Autonomous Circuitry for Substrate Exploration in Freely Moving Drosophila Larvae.” *Current Biology* 22 (20). Elsevier Ltd: 1861–70. doi:10.1016/j.cub.2012.07.048.

Bieber, a J, P M Snow, M Hortsch, N H Patel, J R Jacobs, Z R Traquina, J Schilling, and C S Goodman. 1989. “Drosophila Neuroglian: A Member of the Immunoglobulin Superfamily with Extensive Homology to the Vertebrate Neural Adhesion Molecule L1.” *Cell* 59 (3): 447–60. doi:10.1016/0092-8674(89)90029-9.

Bienroth, S, E Wahle, C Suter-Crazzolara, and W Keller. 1991. “Purification of the Cleavage and Polyadenylation Factor Involved in the 3’ Processing of Messenger RNA Precursors.” *J. Biol. Chem.* 266 (5): 19768–76.

Bier, E. 2005. “Drosophila, the Golden Bug, Emerges as a Tool for Human Genetics.” *Nature Reviews Genetics* 6 (1): 9–23. doi:10.1038/nrg1503.

Bivik, Caroline, Shahrzad Bahrampour, Carina Ulvklo, Patrik Nilsson, Anna Angel, Fredrik Fransson, Erika Lundin, Jakob Renhorn, and Stefan Thor. 2015. “Novel Genes Involved in Controlling Specification of Drosophila FMRFamide Neuropeptide Cells.” *Genetics* 200 (4): 1229–44. doi:10.1534/genetics.115.178483.

Bjorum, Sonia M, Rebecca a Simonette, Raul Alanis, Jennifer E Wang, Benjamin M Lewis, Michael H Trejo, Keith a Hanson, and Kathleen M Beckingham. 2013. “The Drosophila BTB Domain Protein Jim Lovell Has Roles in Multiple Larval and Adult Behaviors.” *PloS One* 8 (4): e61270. doi:10.1371/journal.pone.0061270.

Black, Douglas L. 2003. “Mechanisms of Alternative Pre-Messenger RNA Splicing.” *Annual Review of Biochemistry* 72 (1): 291–336. doi:10.1146/annurev.biochem.72.121801.161720.

Blake, J. A., K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, et al. 2015. “Gene Ontology Consortium: Going Forward.” *Nucleic Acids Research* 43 (D1): D1049–56. doi:10.1093/nar/gku1179.

Blobel, G. 1973. “A Protein of Molecular Weight 78,000 Bound to the Polyadenylate Region of Eukaryotic Messenger RNAs.” *Proceedings of the National Academy of*

Sciences of the United States of America 70 (3): 924–28. doi:10.1073/pnas.70.3.924.

Brand, A H, and N Perrimon. 1993. “Targeted Gene Expression as a Means of Altering Cell Fates and Generating Dominant Phenotypes.” *Development* 118 (2): 401–15. doi:10.1101/lm.1331809.

Brown, Kirk M., and Gregory M. Gilmartin. 2003. “A Mechanism for the Regulation of Pre-mRNA 3’ Processing by Human Cleavage Factor Im.” *Molecular Cell* 12 (6): 1467–76. doi:10.1016/S1097-2765(03)00453-2.

Brown, Mark R, Joe W Crim, Ryan C Arata, Haini N Cai, Cao Chun, and Ping Shen. 1999. “Identification of a Drosophila Brain-Gut Peptide Related to the Neuropeptide Y Family.” *Peptides* 20: 1035–42.

Burd, Craig J, Christin E Petre, Lisa M Morey, Ying Wang, Monica P Revelo, Christopher A Haiman, Shan Lu, et al. 2006. “Cyclin D1b Variant Influences Prostate Cancer Growth through Aberrant Androgen Receptor Regulation.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (7): 2190–95. doi:10.1073/pnas.0506281103.

Buszczak, Michael, Shelley Paterno, Daniel Lighthouse, Julia Bachman, Jamie Planck, Stephenie Owen, Andrew D. Skora, et al. 2007. “The Carnegie Protein Trap Library: A Versatile Tool for Drosophila Developmental Studies.” *Genetics* 175 (3): 1505–31. doi:10.1534/genetics.106.065961.

Callebaut, Isabelle, Despina Moshous, Jean-Paul Mornon, and Jean-Pierre de Villartay. 2002. “Metallo-Beta-Lactamase Fold within Nucleic Acids Processing Enzymes: The Beta-CASP Family.” *Nucleic Acids Research* 30 (16): 3592–3601. doi:10.1093/nar/gkf470.

Carbone, Mary Anna, Katherine W. Jordan, Richard F. Lyman, Susan T. Harbison, Jeff Leips, Theodore J. Morgan, Maria DeLuca, Philip Awadalla, and Trudy F C Mackay. 2006. “Phenotypic Variation and Natural Selection at Catsup, a Pleiotropic Quantitative Trait Gene in Drosophila.” *Current Biology* 16 (9): 912–19. doi:10.1016/j.cub.2006.03.051.

Casadio, Angela, Dasa Longman, Nele Hug, Laurent Delavaine, Raúl Vallejos Baier, Claudio R Alonso, and Javier F Cáceres. 2014. “Identification and Characterization of Novel Factors That Act in the Nonsense-Mediated mRNA Decay Pathway in Nematodes, Flies and Mammals.” *EMBO Reports*, 1–9.

Chen, Cho-yi, Shui-tein Chen, Hsueh-fen Juan, and Hsuan-cheng Huang. 2012. “Lengthening of 3’UTR Increases with Morphological Complexity in Animal Evolution.” *Oxford University Press*, 3–6.

Chen, Dahong, Chunjing Qu, and Randall S. Hewes. 2014. *Neuronal Remodeling during Metamorphosis Is Regulated by the Alan Shepard (Shep) Gene in Drosophila Melanogaster*. *Genetics*. Vol. 197. doi:10.1534/genetics.114.166181.

Chen, M S, C C Burgess, R B Vallee, and S C Wadsworth. 1992. “Developmental Stage- and Tissue-Specific Expression of Shibire, a Drosophila Gene Involved in Endocytosis.” *Journal of Cell Science* 103 (Pt 3: 619–28.

Chen, Mo, and James L Manley. 2009. “Mechanisms of Alternative Splicing Regulation: Insights from Molecular and Genomics Approaches.” *Molecular Cell* 10 (11): 741–54. doi:10.1038/nrm2777.Mechanisms.

Cherry, J. Michael, Eurie L. Hong, Craig Amundsen, Rama Balakrishnan, Gail Binkley, Esther T. Chan, Karen R. Christie, et al. 2012. “Saccharomyces Genome

Database: The Genomics Resource of Budding Yeast." *Nucleic Acids Research* 40 (D1): 700–705. doi:10.1093/nar/gkr1029.

Chintapalli, Venkateswara R, Jing Wang, and Julian a T Dow. 2007. "Using FlyAtlas to Identify Better Drosophila Melanogaster Models of Human Disease." *Nature Genetics* 39 (6): 715–20. doi:10.1038/ng2049.

Chung, Yun Doo, Jingchun Zhu, Young-goo Han, Maurice J Kernan, Stony Brook, and New York. 2001. "nompA Encodes a PNS-Specific , ZP Domain Protein Required to Connect Mechanosensory Dendrites to Sensory Structures." *Neuron* 29: 415–28.

Consortium, The Gene Ontology. 2000. "Gene Ontology: Tool for the Identification of Biology." *Nature Genetics* 25 (may): 25–29.

Cruz, Cristina, Alvaro Glavic, Mar Casado, and Jose F. De Celis. 2009. "A Gain-of-Function Screen Identifying Genes Required for Growth and Pattern Formation of the Drosophila Melanogaster Wing." *Genetics* 183 (3): 1005–26. doi:10.1534/genetics.109.107748.

Danckwardt, S, M W Hentze, and A E Kulozik. 2008. "3' End mRNA Processing: Molecular Mechanisms and Implications for Health and Disease." *Embo J* 27 (3): 482–98. doi:7601932 [pii]\r10.1038/sj.emboj.7601932.

Dansereau, David a., and Paul Lasko. 2008. "RanBPM Regulates Cell Shape, Arrangement, and Capacity of the Female Germline Stem Cell Niche in Drosophila Melanogaster." *Journal of Cell Biology* 182 (5): 963–77. doi:10.1083/jcb.200711046.

Darmon, Sarah K, and Carol S Lutz. 2012. "mRNA 3' End Processing Factors: A Phylogenetic Comparison." *Comparative and Functional Genomics* 2012 (January): 876893. doi:10.1155/2012/876893.

Darnell, J E, R Wall, and R J Tushinski. 1971. "An Adenylic Acid-Rich Sequence in Messenger RNA of HeLa Cells and Its Possible Relationship to Reiterated Sites in DNA." *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1321–25. doi:10.1073/pnas.68.6.1321.

Dass, Brinda, K. Wyatt McMahon, Nancy A. Jenkins, Debra J. Gilbert, Neal G. Copeland, and Clinton C. MacDonald. 2001. "The Gene for a Variant Form of the Polyadenylation Protein CstF-64 Is on Chromosome 19 and Is Expressed in Pachytene Spermatocytes in Mice." *Journal of Biological Chemistry* 276 (11): 8044–50. doi:10.1074/jbc.M009091200.

Dávila López, Marcela, and Tore Samuelsson. 2008. "Early Evolution of Histone mRNA 3' End Processing." *RNA* 14 (1): 1–10. doi:10.1261/rna.782308.

de Vries, H, U Rüegsegger, W Hübner, a Friedlein, H Langen, and W Keller. 2000. "Human Pre-mRNA Cleavage Factor II(m) Contains Homologs of Yeast Proteins and Bridges Two Other Cleavage Factors." *The EMBO Journal* 19 (21): 5895–5904. doi:10.1093/emboj/19.21.5895.

Derti, Adnan, Philip Garrett-Engele, Kenzie D. MacIsaac, Richard C. Stevens, Shreedharan Sriram, Ronghua Chen, Carol A. Rohl, Jason M. Johnson, and Tomas Babak. 2012. "A Quantitative Atlas of Polyadenylation in Five Mammals." *Genome Research* 22 (6): 1173–83. doi:10.1101/gr.132563.111.

Di Giammartino, Dafne Campigli, Kensei Nishida, and James L Manley. 2011. "Mechanisms and Consequences of Alternative Polyadenylation." *Molecular Cell* 43 (6). Elsevier Inc.: 853–66. doi:10.1016/j.molcel.2011.08.017.

Dominski, Zbigniew, Xiao-cui Yang, Matthew Purdy, J Eric, William F Marzluff, and Eric J Wagner. 2005. “A CPSF-73 Homologue Is Required for Cell Cycle Progression but Not Cell Growth and Interacts with a Protein Having Features of CPSF-100.” *Molecular and Cellular Biology* 25 (4): 1489–1500. doi:10.1128/MCB.25.4.1489.

Dubnau, J, L Grady, T Kitamoto, and T Tully. 2001. “Disruption of Neurotransmission in Drosophila Mushroom Body Blocks Retrieval but Not Acquisition of Memory.” *Nature* 411 (6836): 476–80. doi:10.1038/35078077.

Edmonds, Mary, and Richard Abrams. 1960. “Polynucleotide Biosynthesis: Formation of a Sequence of Adenylate Units from Adenosine Triphosphate by an Enzyme from Thymus Nuclei.” *The Journal of Biological chemistryStry* 235 (4).

Edmonds, M, M H Vaughan, and H Nakazato. 1971. “Polyadenylic Acid Sequences in the Heterogeneous Nuclear RNA and Rapidly-Labeled Polyribosomal RNA of HeLa Cells: Possible Evidence for a Precursor Relationship.” *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1336–40. doi:10.1073/pnas.68.6.1336.

Erikson, L R. 1991. “Structure, Expression, and Regulation of Protein Kinases Involved in the Phosphorylation of Ribosomal Protein S6.” *J.Biol.Chem.* 266 (10): 6007–10.

Erson-Bensan, A. E., and T. Can. 2016. “Alternative Polyadenylation: Another Foe in Cancer.” *Molecular Cancer Research* 90 (312). doi:10.1158/1541-7786.MCR-15-0489.

Favier, B, and P Dollé. 1997. “Developmental Functions of Mammalian Hox Genes.” *Molecular Human Reproduction* 3 (2): 115–31. doi:10.1093/molehr/3.2.115.

Fitch, C. L., L. Girton, and J. R. Girton. 1992. “The Suppressor of Forked Locus in Drosophila Melanogaster: Genetic and Molecular Analyses.” *Genetica* 85 (3): 185–203. doi:10.1007/BF00132271.

Ford, J P, and M T Hsu. 1978. “Transcription Pattern of in Vivo-Labeled Late Simian Virus 40 RNA: Equimolar Transcription beyond the mRNA 3’ Terminus.” *Journal of Virology* 28 (3): 795–801.

Friedrich, Jana, Sebastian Sorge, Fatmire Bujupi, Michael P. Eichenlaub, Natalie G. Schulz, Jochen Wittbrodt, and Ingrid Lohmann. 2016. “Hox Function Is Required for the Development and Maintenance of the Drosophila Feeding Motor Unit.” *Cell Reports*. Elsevier Ltd, 1–11. doi:10.1016/j.celrep.2015.12.077.

Fushiki, Akira, Maarten F Zwart, Hiroshi Kohsaka, Richard D Fetter, Albert Cardona, and Akinao Nose. 2016. “A Circuit Mechanism for the Propagation of Waves of Muscle Contraction in Drosophila.” *eLife* 5: e13253. doi:10.7554/eLife.13253.

Ge, Honghua, Dongwen Zhou, Shuilong Tong, Yongxiang Gao, Maikun Teng, and Liwen Niu. 2008. “Crystal Structure and Possible Dimerization of the Single RRM of Human PABPN1.” *Proteins* 71 (3): 1539–45. doi:10.1002/prot.21973.

Gilmartin, G. M., and J. R. Nevins. 1989. “An Ordered Pathway of Assembly of Components Required for Polyadenylation Site Recognition and Processing.” *Genes and Development* 3 (12 B): 2180–90. doi:10.1101/gad.3.12b.2180.

Gohl, Daryl M, Marion a Silies, Xiaojing J Gao, Sheetal Bhalerao, Francisco J Luongo, Chun-Chieh Lin, Christopher J Potter, and Thomas R Clandinin. 2011. “A Versatile in Vivo System for Directed Dissection of Gene Expression Patterns.” *Nature Methods* 8 (3): 231–37. doi:10.1038/nmeth.1561.

Graveley, Brenton R, Angela N Brooks, Joseph W Carlson, Michael O Duff, Jane M

Landolin, Li Yang, Carlo G Artieri, et al. 2011. “The Developmental Transcriptome of Drosophila Melanogaster.” *Nature* 471 (7339). Nature Publishing Group: 473–79. doi:10.1038/nature09715.

Grunberg-Manago, Marianne, Priscilla J. Ortiz, and Severo Ochoa. 1955. “Enzymatic Synthesis of Nucleic Acidlike Polynucleotides.” *Science* 215 (1): 403–15.

———. 1956. “Polynucleotide Phosphorylase of Azotobacter Vinelandii.” *Biochimica et Biophysica Acta* 20: 269–85.

Gupta, Ishaan, Sandra Clauder-Münster, Bernd Klaus, Aino I Järvelin, Raeka S Aiyar, Vladimir Benes, Stefan Wilkening, Wolfgang Huber, Vicent Pelechano, and Lars M Steinmetz. 2014. “Alternative Polyadenylation Diversifies Post-Transcriptional Regulation by Selective RNA-Protein Interactions.” *Molecular Systems Biology* 10 (2): 719. doi:10.1002/msb.135068.

Haddad, Raphaël, Frédérique Maurice, Nicolas Viphakone, Florence Voisinet-Hakil, Sébastien Fribourg, and Lionel Minvielle-Sébastia. 2012. “An Essential Role for Clp1 in Assembly of Polyadenylation Complex CF IA and Pol II Transcription Termination.” *Nucleic Acids Research* 40 (3): 1226–39. doi:10.1093/nar/gkr800.

Hayashi, Shigeo, Kei Ito, Yukiko Sado, Misako Taniguchi, Ai Akimoto, Hiroko Takeuchi, Toshiro Aigaki, et al. 2002. “GETDB, a Database Compiling Expression Patterns and Molecular Locations of a Collection of gal4 Enhancer Traps.” *Genesis* 34 (1–2): 58–61. doi:10.1002/gene.10137.

He, Xiang-Jun, Qi Zhang, Li-Ping Ma, Na Li, Xiao-Hong Chang, and Yu-Jun Zhang. 2016. “Aberrant Alternative Polyadenylation Is Responsible for Survivin Up-Regulation in Ovarian Cancer.” *Chinese Medical Journal* 129 (10): 1140. doi:10.4103/0366-6999.181965.

Hilgers, Valérie, Michael W Perry, David Hendrix, Alexander Stark, Michael Levine, and Benjamin Haley. 2011. “Neural-Specific Elongation of 3’ UTRs during Drosophila Development.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (38): 15864–69. doi:10.1073/pnas.1112672108.

Hirose, Y, and J L Manley. 1998. “RNA Polymerase II Is an Essential mRNA Polyadenylation Factor.” *Nature* 395 (6697): 93–96. doi:10.1038/25786.

Hollingworth, David, Christian G Noble, Ian a Taylor, and Andres Ramos. 2006. “RNA Polymerase II CTD Phosphopeptides Compete with RNA for the Interaction with Pcf11.” *RNA* 12 (4): 555–60. doi:10.1261/rna.2304506.

Hoque, Mainul, Zhe Ji, Dinghai Zheng, Wenting Luo, Wencheng Li, Bei You, Ji Yeon Park, Ghassan Yehia, and Bin Tian. 2013. “Analysis of Alternative Cleavage and Polyadenylation by 3’ Region Extraction and Deep Sequencing.” *Nature Methods* 10 (2): 133–39. doi:10.1038/nmeth.2288.

Hsu, Patrick D., Eric S. Lander, and Feng Zhang. 2014. “Development and Applications of CRISPR-Cas9 for Genome Engineering.” *Cell* 157 (6). Elsevier: 1262–78. doi:10.1016/j.cell.2014.05.010.

Huckesfeld, Sebastian, Andreas Schoofs, Philipp Schlegel, Anton Miroschnikow, and Michael J. Pankratz. 2015. “Localization of Motor Neurons and Central Pattern Generators for Motor Patterns Underlying Feeding Behavior in Drosophila Larvae.” *PLoS ONE* 10 (8): 1–18. doi:10.1371/journal.pone.0135011.

Hummel, T, K Krukkert, J Roos, G Davis, and C Klämbt. 2000. “Drosophila Futsch/22C10 Is a MAP1B-like Protein Required for Dendritic and Axonal Development.” *Neuron* 26 (2): 357–70.

Hummel, Thomas, Kristina Schimmelpfeng, and Christian Kla. 1999. “Commissure Formation in the Embryonic CNS of Drosophila.” *Developmental Biology* 398: 381–98.

Hutchins, Lucie N., Sean M. Murphy, Priyam Singh, and Joel H. Graber. 2008. “Position-Dependent Motif Characterization Using Non-Negative Matrix Factorization.” *Bioinformatics* 24 (23): 2684–90. doi:10.1093/bioinformatics/btn526.

Ito, S, A Sakai, T Nomura, Y Miki, M Ouchida, J Sasaki, and K Shimizu. 2001. “A Novel WD40 Repeat Protein, WDC146, Highly Expressed during Spermatogenesis in a Stage-Specific Manner.” *Biochemical and Biophysical Research Communications* 280 (3): 656–63. doi:10.1006/bbrc.2000.4163.

Jan, Calvin H, Robin C Friedman, J Graham Ruby, and David P Bartel. 2011. “Formation, Regulation and Evolution of Caenorhabditis Elegans 3’UTRs.” *Nature* 469 (7328). Nature Publishing Group: 97–101. doi:10.1038/nature09616.

Jeibmann, Astrid, and Werner Paulus. 2009. “Drosophila Melanogaster: The Model Organism.” *Entomologia Experimentalis et Applicata* 10 (1823): 407–40. doi:10.3390/ijms10020407.

Ji, Zhe, Ju Youn Lee, Zhenhua Pan, Bingjun Jiang, and Bin Tian. 2009. “Progressive Lengthening of 3’ Untranslated Regions of mRNAs by Alternative Polyadenylation during Mouse Embryonic Development.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (17): 7028–33. doi:10.1073/pnas.0900028106.

Johnson, Mark, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L. Madden. 2008. “NCBI BLAST: A Better Web Interface.” *Nucleic Acids Research* 36 (Web Server issue): 5–9. doi:10.1093/nar/gkn201.

Jürgens, Gerd, Ruth Lehmann, Margit Schardin, and Christiane Nüsslein-Volhard. 1986. “Segmental Organisation of the Head in the Embryo of Drosophila Melanogaster - A Blastoderm Fate Map of the Cuticle Structures of the Larval Head.” *Roux’s Archives of Developmental Biology* 195: 359–77. doi:10.1007/BF00402870.

Kapsimali, Marika, Wigard P Kloosterman, Ewart de Bruijn, Frederic Rosa, Ronald H A Plasterk, and Stephen W Wilson. 2007. “MicroRNAs Show a Wide Diversity of Expression Profiles in the Developing and Mature Central Nervous System.” *Genome Biology* 8 (8): R173. doi:10.1186/gb-2007-8-8-r173.

Kaufmann, Isabelle, Georges Martin, Arno Friedlein, Hanno Langen, and Walter Keller. 2004. “Human Fip1 Is a Subunit of CPSF That Binds to U-Rich RNA Elements and Stimulates poly(A) Polymerase.” *The EMBO Journal* 23 (3): 616–26. doi:10.1038/sj.emboj.7600070.

Keegan, Liam P., Leeane McGurk, Juan Pablo Palavicini, James Brindle, Simona Paro, Xianghua Li, Joshua J C Rosenthal, and Mary A. O’Connell. 2011. “Functional Conservation in Human and Drosophila of Metazoan ADAR2 Involved in RNA Editing: Loss of ADAR1 in Insects.” *Nucleic Acids Research* 39 (16): 7249–62. doi:10.1093/nar/gkr423.

Keller, W., S. Bienroth, K. M. Lang, and G. Christofori. 1991. “Cleavage and Polyadenylation Factor CPF Specifically Interacts with the Pre-mRNA 3’ Processing Signal AAUAAA.” *The EMBO Journal* 10 (13): 4241–49.

Keon, Brigitte H., Stephan Schafer, Caecilia Kuhn, Christine Grund, and Werner W.

Franke. 1996. “Symplekin, a Novel Type of Tight Junction Plaque Protein.” *Journal of Cell Biology* 134 (4): 1003–18. doi:10.1083/jcb.134.4.1003.

Kerwitz, Yvonne, Uwe Ku, Hauke Lilie, Anne Knoth, Till Scheuermann, Henning Friedrich, Elisabeth Schwarz, and Elmar Wahle. 2003. “Stimulation of poly(A) Polymerase through a Direct Interaction with the Nuclear poly(A) Binding Protein Allosterically Regulated by RNA.” *EMBO Journal* 22 (14): 3705–14.

Kim, Sol, Junichi Yamamoto, Yexi Chen, Masatoshi Aida, Tadashi Wada, Hiroshi Handa, and Yuki Yamaguchi. 2010. “Evidence That Cleavage Factor Im Is a Heterotetrameric Protein Complex Controlling Alternative Polyadenylation.” *Genes to Cells* 15 (9): 1003–13. doi:10.1111/j.1365-2443.2010.01436.x.

Klein, Thomas, and José A Campos-Ortega. 1997. “Klumpfuss , a Drosophila Gene Encoding a Member of the EGR Family of Transcription Factors, Is Involved in Bristle and Leg Development.” *Development* 3134: 3123–34.

Knudsen, K E, J Alan Diehl, C a Haiman, and E S Knudsen. 2006. “Cyclin D1: Polymorphism, Aberrant Splicing and Cancer Risk.” *Oncogene* 25 (11): 1620–28. doi:10.1038/sj.onc.1209371.

Krol, S., O. Cavalleri, P. Ramoino, A. Gliozzi, and A. Diaspro. 2003. “Encapsulated Yeast Cells inside Paramecium Primaurelia: A Model System for Protection Capability of Polyelectrolyte Shells.” *Journal of Microscopy* 212 (3): 239–43. doi:10.1111/j.1365-2818.2003.01251.x.

Krumlauf, R., H. Marshall, M. Studer, S. Nonchev, Mai Har Sham, and A. Lumsden. 1993. “Hox Homeobox Genes and Regionalisation of the Nervous System.” *Journal of Neurobiology* 24 (10): 1328–40. doi:10.1002/neu.480241006.

Kubo, Tomohiro, Tadashi Wada, Yuki Yamaguchi, Akira Shimizu, and Hiroshi Handa. 2006. “Knock-down of 25 kDa Subunit of Cleavage Factor Im in Hela Cells Alters Alternative Polyadenylation within 3’-UTRs.” *Nucleic Acids Research* 34 (21): 6264–71. doi:10.1093/nar/gkl794.

Kunz, T., K. F. Kraft, G. M. Technau, and R. Urbach. 2012. “Origin of Drosophila Mushroom Body Neuroblasts and Generation of Divergent Embryonic Lineages.” *Development* 139 (14): 2510–22. doi:10.1242/dev.077883.

Kurusu, Mitsuhiko, Takeshi Awasaki, Liria M Masuda-Nakagawa, Hiroshi Kawauchi, Kei Ito, and Katsuo Furukubo-Tokunaga. 2002. “Embryonic and Larval Development of the Drosophila Mushroom Bodies: Concentric Layer Subdivisions and the Role of Fasciclin II.” *Development* 129: 409–19.

Kwon, Matt Hyoung, Heather Callaway, Jim Zhong, and Barry Yedvobnick. 2013. “A Targeted Genetic Modifier Screen Links the SWI2/SNF2 Protein Domino to Growth and Autophagy Genes in Drosophila Melanogaster.” *G3 (Bethesda, Md.)* 3 (5): 815–25. doi:10.1534/g3.112.005496.

Lackford, Brad, Chengguo Yao, Georgette M Charles, Lingjie Weng, Xiaofeng Zheng, Eun-A Choi, Xiaohui Xie, et al. 2014. “Fip1 Regulates mRNA Alternative Polyadenylation to Promote Stem Cell Self-Renewal.” *The EMBO Journal*, March, 1–12. doi:10.1002/embj.201386537.

Lee, Donghoon M, and Tony J C. Harris. 2013. “An Arf-GEF Regulates Antagonism between Endocytosis and the Cytoskeleton for Drosophila Blastoderm Development.” *Current Biology* 23 (21). Elsevier Ltd: 2110–20. doi:10.1016/j.cub.2013.08.058.

Lee, Kyu-Sun, Kwan-Hee You, Jong-Kil Choo, Yong-Mahn Han, and Kweon Yu. 2004.

"Drosophila Short Neuropeptide F Regulates Food Intake and Body Size." *The Journal of Biological Chemistry* 279 (49): 50781–89. doi:10.1074/jbc.M407842200.

Lee, S Y, J Mendecki, and G Brawerman. 1971. "A Polynucleotide Segment Rich in Adenylic Acid in the Rapidly-Labeled Polyribosomal RNA Component of Mouse Sarcoma 180 Ascites Cells." *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1331–35.

Lefkowitz, Robert J. 2005. "Transduction of Receptor Signals by B-Arrestins." *Science* 512 (2005): 512–18. doi:10.1126/science.1109237.

Legendre, Matthieu, and Daniel Gautheret. 2003. "Sequence Determinants in Human Polyadenylation Site Selection." *BMC Genomics* 4 (1): 7. doi:10.1186/1471-2164-4-7.

Lewis, Ed. 1978. "A Gene Complex Controlling Segmentation in Drosophila." *Nature* 276. doi:10.1017/CBO9781107415324.004.

Li, Heng, Shuilong Tong, Xu Li, Hui Shi, Zheng Ying, Yongxiang Gao, Honghua Ge, Liwen Niu, and Maikun Teng. 2011. "Structural Basis of Pre-mRNA Recognition by the Human Cleavage Factor Im Complex." *Cell Research* 21 (7): 1039–51. doi:10.1038/cr.2011.67.

Li, Jie, Rui Li, Leiming You, Anlong Xu, Yonggui Fu, and Shengfeng Huang. 2015. "Evaluation of Two Statistical Methods Provides Insights into the Complex Patterns of Alternative Polyadenylation Site Switching." *Plos One* 10 (4): e0124324. doi:10.1371/journal.pone.0124324.

Li, Wencheng, Bei You, Mainul Hoque, Dinghai Zheng, Wenting Luo, Zhe Ji, Ji Yeon Park, et al. 2015. "Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation." *PLoS Genetics* 11 (4): 1–28. doi:10.1371/journal.pgen.1005166.

Lianoglou, Steve, Vidur Garg, Julie L Yang, Christina S Leslie, and Christine Mayr. 2013. "Ubiquitously Transcribed Genes Use Alternative Polyadenylation to Achieve Tissue-Specific Expression." *Genes & Development*, 2380–96. doi:10.1101/gad.229328.113.Freely.

Licatalosi, Donny D., Masato Yano, John J. Fak, Aldo Mele, Sarah E. Grabinski, Chaolin Zhang, and Robert B. Darnell. 2012. "Ptbp2 Represses Adult-Specific Splicing to Regulate the Generation of Neuronal Precursors in the Embryonic Brain." *Genes and Development* 26 (14): 1626–42. doi:10.1101/gad.191338.112.

Littauer, Uriel, and Arthur Kornberg. 1957. "Reversible Synthesis of Polyribonucleotides with an Enzyme from Escherichia Coli." *Journal of Biological Chemistry* 226: 1077–92.

Liu, Yuting, Wenchao Hu, Yasuhiro Murakawa, Jingwen Yin, Gang Wang, Markus Landthaler, and Jun Yan. 2013. "Cold-Induced RNA-Binding Proteins Regulate Circadian Gene Expression by Controlling Alternative Polyadenylation." *Scientific Reports* 3 (June): 1–11. doi:10.1038/srep02054.

Lizcano, Jose M, Saif Alrubaie, Agnieszka Kieloch, Maria Deak, Sally J Leevers, and Dario R Alessi. 2003. "Insulin-Induced Drosophila S6 Kinase Activation Requires Phosphoinositide 3-Kinase and Protein Kinase B." *The Biochemical Journal* 374: 297–306. doi:10.1042/BJ20030577.

MacDonald, C. C., J. Wilusz, and T. Shenk. 1994. "The 64-Kilodalton Subunit of the CstF Polyadenylation Factor Binds to Pre-mRNAs Downstream of the Cleavage

Site and Influences Cleavage Site Location." *Molecular and Cellular Biology* 14 (10): 6647–54. doi:10.1128/MCB.14.10.6647.

Mallo, M., and C. R. Alonso. 2013. "The Regulation of Hox Gene Expression during Animal Development." *Development* 140 (19): 3951–63. doi:10.1242/dev.068346.

Mandel, Corey R, Syuzo Kaneko, Hailong Zhang, Damara Gebauer, Vasupradha Vethantham, James L Manley, and Liang Tong. 2006. "Polyadenylation Factor CPSF-73 Is the Pre-mRNA 3'-end-Processing Endonuclease." *Nature* 444 (7121): 953–56. doi:10.1038/nature05363.

Manley, James L., Phillip A. Sharp, and Malcolm L. Gefter. 1982. "RNA Synthesis in Isolated Nuclei: Processing of Adenovirus Serotype 2 Late Messenger RNA Precursors." *Journal of Molecular Biology* 159 (4): 581–99. doi:10.1016/0022-2836(82)90102-4.

Marek, K W, N Ng, R Fetter, S Smolik, C S Goodman, and G W Davis. 2000. "A Genetic Analysis of Synaptic Development: Pre- and Postsynaptic dCBP Control Transmitter Release at the Drosophila NMJ." *Neuron* 25 (3): 537–47. doi:10.1016/S0896-6273(00)81058-2.

Martin, G, W Keller, and S Doublié. 2000. "Crystal Structure of Mammalian poly(A) Polymerase in Complex with an Analog of ATP." *The EMBO Journal* 19 (16): 4193–4203. doi:10.1093/emboj/19.16.4193.

Martin, Georges, Andreas R Gruber, Walter Keller, and Mihaela Zavolan. 2012. "Genome-Wide Analysis of Pre-mRNA 3' End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3' UTR Length." *Cell Reports*, 753–63. doi:10.1016/j.celrep.2012.05.003.

Martin, Robert M., Heinrich Leonhardt, and M. Cristina Cardoso. 2005. "DNA Labeling in Living Cells." *Cytometry Part A* 67 (1): 45–52. doi:10.1002/cyto.a.20172.

Marzluff, William F, Eric J Wagner, and Robert J Duronio. 2008. "Metabolism and Regulation of Canonical Histone mRNAs: Life without a poly(A) Tail." *Nature Reviews. Genetics* 9 (11): 843–54. doi:10.1038/nrg2438.

Masamha, Chioniso P., Zheng Xia, Jingxuan Yang, Todd R. Albrecht, Min Li, Ann-Bin Shyu, Wei Li, and Eric J. Wagner. 2014. "CFIm25 Links Alternative Polyadenylation to Glioblastoma Tumour Suppression." *Nature* 0 (May). Nature Publishing Group. doi:10.1038/nature13261.

Masamha, Chioniso P, Zheng Xia, Natoya Peart, Scott Collum, W E I Li, Eric J Wagner, and Ann-bin Shyu. 2016. "CFIm25 Regulates Glutaminase Alternative Terminal Exon Definition to Modulate miR-23 Function." *RNA*, 1–9. doi:10.1261/rna.055939.116.

Mayr, Christine, and David P Bartel. 2009. "Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells." *Cell* 138 (4). Elsevier Ltd: 673–84. doi:10.1016/j.cell.2009.06.016.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S D Patterson, M Wickens, and D L Bentley. 1997. "The C-Terminal Domain of RNA Polymerase II Couples mRNA Processing to Transcription." *Nature*. doi:10.1038/385357a0.

McLennan, a. G. 2006. "The Nudix Hydrolase Superfamily." *Cellular and Molecular Life Sciences* 63 (2): 123–43. doi:10.1007/s00018-005-5386-7.

Medioni, Caroline, Mirana Ramialison, Anne Ephrussi, and Florence Besse. 2014. "Imp

Promotes Axonal Remodeling by Regulating Profilin mRNA during Brain Development." *Current Biology* 24 (7). Elsevier Ltd: 793–800. doi:10.1016/j.cub.2014.02.038.

Melcher, Christoph, and Michael J Pankratz. 2005. "Candidate Gustatory Interneurons Modulating Feeding Behavior in the Drosophila Brain." *PLoS Biology* 3 (9): e305. doi:10.1371/journal.pbio.0030305.

Merrill, V K L, and F R Turner. 1987. "A Genetic and Developmental Analysis of Mutations in the Deformed Locus in Drosophila Melaogaster." *Developmental Biology* 395: 379–95.

Metzstein, Mark M, and Mark a Krasnow. 2006. "Functions of the Nonsense-Mediated mRNA Decay Pathway in Drosophila Development." *PLoS Genetics* 2 (12): e180. doi:10.1371/journal.pgen.0020180.

Milchanowski, Allison B., Amy L. Henkenius, Maya Narayanan, Volker Hartenstein, and Utpal Banerjee. 2004. "Identification and Characterization of Genes Involved in Embryonic Crystal Cell Formation during Drosophila Hematopoiesis." *Genetics* 168 (1): 325–39. doi:10.1534/genetics.104.028639.

Miller, D F, B T Rogers, A Kalkbrenner, B Hamilton, S L Holtzman, and T Kaufman. 2001. "Cross-Regulation of Hox Genes in the Drosophila Melanogaster Embryo." *Mechanisms of Development* 102 (1–2): 3–16.

Miller, Michael R, Kristin J Robinson, Michael D Cleary, and Chris Q Doe. 2009. "TU-Tagging : Cell Type – Specific RNA Isolation from Intact Complex Tissues." *Nature Methods*, no. May: 1–5. doi:10.1038/NMETH.1329.

Millevoi, Stefania, Clarisse Loulergue, Sabine Dettwiler, Sarah Zeïneb Karaa, Walter Keller, Michael Antoniou, and Stéphan Vagner. 2006. "An Interaction between U2AF 65 and CF I(m) Links the Splicing and 3' End Processing Machineries." *The EMBO Journal* 25 (20): 4854–64. doi:10.1038/sj.emboj.7601331.

Miura, Pedro, Piero Sanfilippo, Sol Shenker, and Eric C Lai. 2014. "Alternative Polyadenylation in the Nervous System: To What Lengths Will 3' UTR Extensions Take Us?" *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, June, 1–12. doi:10.1002/bies.201300174.

Miura, Pedro, Sol Shenker, Celia Andreu-agullo, Pedro Miura, Sol Shenker, Celia Andreu-agullo, Jakub O Westholm, and Eric C Lai. 2013. "Widespread and Extensive Lengthening of 3′ UTRs in the Mammalian Brain." *Genome Research*, 812–25. doi:10.1101/gr.146886.112.

Monarez, Roberto R, Clinton C MacDonald, and Brinda Dass. 2007. "Polyadenylation Proteins CstF-64 and tauCstF-64 Exhibit Differential Binding Affinities for RNA Polymers." *The Biochemical Journal* 401 (3): 651–58. doi:10.1042/BJ20061097.

Moore, Melissa J. 2012. "From Birth to Death : The Complex Lives of Eukaryotic mRNAs." *Science* 1514 (2005). doi:10.1126/science.1111443.

Mount, S M, and H K Salz. 2000. "Pre-Messenger RNA Processing Factors in the Drosophila Genome." *The Journal of Cell Biology* 150 (2): F37-44.

Murata, T, H Nagaso, S Kashiwabara, T Baba, H Okano, and K K Yokoyama. 2001. "The Hiiragi Gene Encodes a poly(A) Polymerase, Which Controls the Formation of the Wing Margin in Drosophila Melanogaster." *Developmental Biology* 233 (1): 137–47. doi:10.1006/dbio.2001.0205.

Murthy, K. G K, and J. L. Manley. 1992. "Characterization of the Multisubunit Cleavage-

Polyadenylation Specificity Factor from Calf Thymus." *Journal of Biological Chemistry* 267 (21): 14804–11.

Murthy, Kanneganti G K, and James L Manley. 1995. "The 160-kD Subunit of Human Cleavage- Polyadenylation Specificity Factor Coordinates Pre-mRNA 3 â€™ -End Formation." *Genes and Development*, 2672–83. doi:10.1101/gad.9.21.2672.

Neely, G. Gregory, Andreas Hess, Michael Costigan, Alex C. Keene, Spyros Goulas, Michiel Langeslag, Robert S. Griffin, et al. 2010. "A Genome-Wide Drosophila Screen for Heat Nociception Identifies α2δ3 as an Evolutionarily Conserved Pain Gene." *Cell* 143 (4). Elsevier Inc.: 628–38. doi:10.1016/j.cell.2010.09.047.

Nei, M, P Xu, and G Glazko. 2001. "Estimation of Divergence Times from Multiprotein Sequences for a Few Mammalian Species and Several Distantly Related Organisms." *Proceedings of the National Academy of Sciences of the United States of America* 98 (5): 2497–2502. doi:10.1073/pnas.051611498.

Neuwald, A F, and A Poleksic. 2000. "PSI-BLAST Searches Using Hidden Markov Models of Structural Repeats: Prediction of an Unusual Sliding DNA Clamp and of Beta-Propellers in UV-Damaged DNA-Binding Protein." *Nucleic Acids Research* 28 (18): 3570–80. doi:10.1093/nar/28.18.3570.

Nevins, Joseph R., and James E. Darnell. 1978. "Steps in the Processing of Ad2 mRNA: Poly(A)+ Nuclear Sequences Are Conserved and poly(A) Addition Precedes Splicing." *Cell* 15 (4): 1477–93. doi:10.1016/0092-8674(78)90071-5.

Page, Scott L., Kim S. McKim, Benjamin Deneen, Tajia L. Van Hook, and R. Scott Hawley. 2000. "Genetic Studies of Mei-P26 Reveal a Link between the Processes That Control Germ Cell Proliferation in Both Sexes and Those That Control Meiotic Exchange in Drosophila." *Genetics* 155 (4): 1757–72.

Pan, Qun, Ofer Shai, Leo J Lee, Brendan J Frey, and Benjamin J Blencowe. 2008. "Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing." *Nature Genetics* 40 (12): 1413–15. doi:10.1038/ng.259.

Park, Jong-Eun, Hyerim Yi, Yoosik Kim, Hyeshik Chang, and V. Narry Kim. 2016. "Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle." *Mol Cell* 62 (3). Elsevier Inc.: 462–71. doi:http://dx.doi.org/10.1016/j.molcel.2016.04.007.

Patraquim, Pedro, Maria Warnefors, and Claudio R Alonso. 2011. "Evolution of Hox Post-Transcriptional Regulation by Alternative Polyadenylation and microRNA Modulation within 12 Drosophila Genomes." *Molecular Biology and Evolution* 28 (9): 2453–60. doi:10.1093/molbev/msr073.

Pease, Benjamin, Ana C Borges, and Welcome Bender. 2013. *Non-Coding RNAs of the Ultrabithorax Domain of the Drosophila Bithorax Complex*. *Genetics*. doi:10.1534/genetics.113.155036.

Pereira, H S, and M B Sokolowski. 1993. "Mutations in the Larval Foraging Gene Affect Adult Locomotory Behavior after Feeding in Drosophila Melanogaster." *Proceedings of the National Academy of Sciences of the United States of America* 90 (June): 5044–46. doi:10.1073/pnas.90.11.5044.

Perez Canadillas, Jose Manuel, and Gabriele Varani. 2003. "Recognition of GU-Rich Polyadenylation Regulatory Elements by Human CstF-64 Protein." *EMBO Journal* 22 (11): 2821–30. doi:10.1093/emboj/cdg259.

Perrimon, N., D. Smouse, and G. L. Miklos. 1989. "Developmental Genetics of Loci at

the Base of the X Chromosome of Drosophila Melanogaster." *Genetics* 121 (2): 313–31.

Peter, Annette, Petra Schoettler, Meike Werner, Nicole Beinert, Gordon Dowe, Peter Burkert, Foteini Mourkioti, et al. 2002. "Mapping and Identification of Essential Gene Functions on the X Chromosome of Drosophila." *EMBO Reports* 3 (1): 34–38.

Picao-Osorio, Joao, Jamie Johnston, Matthias Landgraf, Jimena Berni, and Claudio R Alonso. 2015. "MicroRNA-Encoded Behavior in Drosophila." *Science* 350 (6262): 815–20.

Preker, Pascal J., and Walter Keller. 1998. "The HAT Helix, a Repetitive Motif Implicated in RNA Processing." *Trends in Biochemical Sciences* 23 (1): 15–16. doi:10.1016/S0968-0004(97)01156-0.

Preker, Pascal J., Joachim Lingner, Lionel Minvielle-Sebastia, and Walter Keller. 1995. "The FIP1 Gene Encodes a Component of a Yeast Pre-mRNA Polyadenylation Factor That Directly Interacts with poly(A) Polymerase." *Cell* 81 (3): 379–89. doi:10.1016/0092-8674(95)90391-7.

Proudfoot, N J, and G G Brownlee. 1976. "3' Non-Coding Region Sequences in Eukaryotic Messenger RNA." *Nature* 263 (5574): 211–14. doi:10.1038/263211a0.

Pruitt, Kim D., Garth R. Brown, Susan M. Hiatt, Françoise Thibaud-Nissen, Alexander Astashyn, Olga Ermolaeva, Catherine M. Farrell, et al. 2014. "RefSeq: An Update on Mammalian Reference Sequences." *Nucleic Acids Research* 42 (D1): 756–63. doi:10.1093/nar/gkt1114.

Puck, B Y Theodore, Dimitry Morkovin, Philip I Marcus, and Steven J Cieclura. 1997. "The 30-kD Subunit of Mammalian Cleavage and Polyadenylation Specificity Factor and Its Yeast Homolog Are RNA-Binding Zinc Finger Proteins." *Genes and Development* 11 (60): 1703–16.

Qu, Xiangping, Jose Manuel Perez-Canadillas, Shipra Agrawal, Julia De Baecke, Hailing Cheng, Gabriele Varani, and Claire Moore. 2007. "The C-Terminal Domains of Vertebrate CstF-64 and Its Yeast Orthologue Rna15 Form a New Structure Critical for mRNA 3'-end Processing." *Journal of Biological Chemistry* 282 (3): 2101–15. doi:10.1074/jbc.M609981200.

Raabe, Tobias; Bollum F.J.; Manley, James L. 1991. "Primary Structure and Expression of Bovine poly(A) Polymerase." *Nature* 353: 229–34. doi:10.1038/353229a0.

Raghu, Padinjat, and Gaiti Hasan. 1995. "The Inositol 1,4,5-Triphosphate Receptor Expression in Drosophila Suggests a Role for IP3 Signalling in Muscle Development and Adult Chemosensory Functions." *Developmental Biology*.

Raj, Bushra, and Benjamin J. Blencowe. 2015. "Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles." *Neuron* 87 (1). Elsevier Inc.: 14–27. doi:10.1016/j.neuron.2015.05.004.

Ramirez, Alejandro, Stewart Shuman, and Beate Schwer. 2008. "Human RNA 5'-kinase (hClp1) Can Function as a tRNA Splicing Enzyme in Vivo." *RNA* 14 (9): 1737–45. doi:10.1261/rna.1142908.

Ransom, B, S a Goldman, J Meldolesi, L Zhou, K K Murai, K M Harris, K D Mccarthy, et al. 2008. "Proliferating Cells Express mRNAs with Shortened 3' Unstranslated Regions and Fewer microRNA Target Sites." *Science* 320 (June): 1643–48.

Rehwinkel, J A N, Isabelle Behm-ansmant, and David Gatfield. 2005. “miRNA-Mediated Gene Silencing A Crucial Role for GW182 and the DCP1 : DCP2 Decapping Complex in miRNA-Mediated Gene Silencing.” *RNA* 11: 1640–47. doi:10.1261/rna.2191905.van.

Richardson, John M., K. Wyatt McMahon, Clinton C. MacDonald, and George I. Makhatadze. 1999. “MEARA Sequence Repeat of Human CstF-64 Polyadenylation Factor Is Helical in Solution. A Spectroscopic and Calorimetric Study.” *Biochemistry* 38 (39): 12869–75. doi:10.1021/bi990724r.

Rogulja-Ortmann, A., J. Picao-Osorio, C. Villava, P. Patraquim, E. Lafuente, J. Aspden, S. Thomsen, G. M. Technau, and C. R. Alonso. 2014. “The RNA-Binding Protein ELAV Regulates Hox RNA Processing, Expression and Function within the Drosophila Nervous System.” *Development* 141 (10): 2046–56. doi:10.1242/dev.101519.

Rogulja-ortmann, Ana, and Gerhard M Technau. 2008. “Multiple Roles for Hox Genes in Segment-Specific Shaping of CNS Lineages.” *Landes Bioscience* 2 (6): 316–19. doi:10.1242/dev.023986.316.

Rollmann, Stephanie M., Liesbeth Zwarts, Alexis C. Edwards, Akihiko Yamamoto, Patrick Callaerts, Koenraad Norga, Trudy F C Mackay, and Robert R H Anholt. 2008. “Pleiotropic Effects of Drosophila Neuralized on Complex Behaviors and Brain Structure.” *Genetics* 179 (3): 1327–36. doi:10.1534/genetics.108.088435.

Roman, Gregg, Jin He, and Ronald L. Davis. 2000. “Kurtz, a Novel Nonvisual Arrestin, Is an Essential Neural Gene in Drosophila.” *Genetics* 155: 1281–95.

Ronshaugen, Matthew, Frédéric Biemar, Jessica Piel, Mike Levine, and Eric C. Lai. 2005. “The Drosophila microRNA Iab-4 Causes a Dominant Homeotic Transformation of Halteres to Wings.” *Genes and Development* 19 (24): 2947–52. doi:10.1101/gad.1372505.

Rüegsegger, U, D Blank, and W Keller. 1998. “Human Pre-mRNA Cleavage Factor Im Is Related to Spliceosomal SR Proteins and Can Be Reconstituted in Vitro from Recombinant Subunits.” *Molecular Cell* 1 (2): 243–53.

Ruepp, Marc David, Daniel Schümperli, and Silvia M L Barabino. 2011. “mRNA 3’ End Processing and More-Multiple Functions of Mammalian Cleavage Factor I-68.” *Wiley Interdisciplinary Reviews: RNA* 2 (1): 79–91. doi:10.1002/wrna.35.

Ruepp, Marc David, Christoph Schweingruber, Nicole Kleinschmidt, and Daniel Schümperli. 2011. “Interactions of CstF-64, CstF-77, and Symplekin: Implications on Localisation and Function.” *Molecular Biology of the Cell* 22 (1): 91–104. doi:10.1091/mbc.E10-06-0543.

Ryder, Edward, Fiona Blows, Michael Ashburner, Rosa Bautista-Llacer, Darin Coulson, Jenny Drummond, Jane Webster, et al. 2004. “The DrosDel Collection: A Set of P-Element Insertions for Generating Custom Chromosomal Aberrations in Drosophila Melanogaster.” *Genetics* 167 (2): 797–813. doi:10.1534/genetics.104.026658.

Sarov, Mihail, Christiane Barz, Helena Jambor, Marco Y. Hein, Christopher Schmied, Dana Suchold, Bettina Stender, et al. 2016. “A Genome-Wide Resource for the Analysis of Protein Localisation in Drosophila.” *eLife* 5: 1–38. doi:10.7554/eLife.12068.

Scantlebury, Nadia, Xiao Li Zhao, Verónica G Rodriguez Moncalvo, Alison Camiletti, Stacy Zahanova, Aidan Dineen, Ji Hou Xin, and Ana Regina Campos. 2010. “The

Drosophila Gene RanBPM Functions in the Mushroom Body to Regulate Larval Behavior." *PLoS ONE* 5 (5). doi:10.1371/journal.pone.0010652.

Schalet, a, and G Lefevre. 1973. "The Localization Of 'ordinary' sex-Linked Genes in Section 20 of the Polytene X Chromosome of Drosophila Melanogaster." *Chromosoma* 44 (2): 183–202.

Schonemann, L., U. Kuhn, Georges Martin, P. Schafer, Andreas R. Gruber, Walter Keller, M. Zavolan, et al. 2014. "Reconstitution of CPSF Active in Polyadenylation : Recognition of the Polyadenylation Signal by WDR33." *Genes & Development*, October, 2381–93. doi:10.1101/gad.250985.114.proteins.

Schoofs, Andreas, Sebastian Hückesfeld, Philipp Schlegel, Anton Miroschnikow, Marc Peters, Malou Zeymer, Roland Spieβ, Ann-Shyn Chiang, and Michael J Pankratz. 2014. "Selection of Motor Programs for Suppressing Food Intake and Inducing Locomotion in the Drosophila Brain." *PLoS Biology* 12 (6): e1001893. doi:10.1371/journal.pbio.1001893.

Schoofs, Andreas, Sebastian Hückesfeld, Sandya Surendran, and Michael Pankratz. 2014. "Serotonergic Pathways in the Drosophila Larval Enteric Nervous System." *Journal of Insect Physiology* 69C (October). Elsevier Ltd: 118–25. doi:10.1016/j.jinsphys.2014.05.022.

Schoofs, Andreas, Senta Niederegger, and Roland Spieß. 2009. "From Behavior to Fictive Feeding: Anatomy, Innervation and Activation Pattern of Pharyngeal Muscles of Calliphora Vicina 3rd Instar Larvae." *Journal of Insect Physiology* 55 (3): 218–30. doi:10.1016/j.jinsphys.2008.11.011.

Schoofs, Andreas, Senta Niederegger, Andrè van Ooyen, Hans-Georg Heinzel, and Roland Spiess. 2010. "The Brain Can Eat: Establishing the Existence of a Central Pattern Generator for Feeding in Third Instar Larvae of Drosophila Virilis and Drosophila Melanogaster." *Journal of Insect Physiology* 56 (7): 695–705. doi:10.1016/j.jinsphys.2009.12.008.

Sheets, M D, S C Ogg, and M P Wickens. 1990. "Point Mutations in AAUAAA and the Poly (A) Addition Site: Effects on the Accuracy and Efficiency of Cleavage and Polyadenylation in Vitro." *Nucleic Acids Research* 18 (19): 5799–5805. doi:10.1093/nar/18.19.5799.

Shi, X, A Finkelstein, a J Wolf, P a Wade, Z F Burton, and J a Jaehning. 1996. "Paf1p, an RNA Polymerase II-Associated Factor in Saccharomyces Cerevisiae, May Have Both Positive and Negative Roles in Transcription." *Molecular and Cellular Biology* 16 (2): 669–76.

Shi, Yongsheng, Dafne Campigli Di Giammartino, Derek Taylor, Ali Sarkeshik, William J Rice, John R Yates, Joachim Frank, and James L Manley. 2009. "Molecular Architecture of the Human Pre-mRNA 3' Processing Complex." *Molecular Cell* 33 (3). Elsevier Ltd: 365–76. doi:10.1016/j.molcel.2008.12.028.

Siddiqi, O, and S Benzer. 1976. "Neurophysiological Defects in Temperature-Sensitive Paralytic Mutants of Drosophila Melanogaster." *Proceedings of the National Academy of Sciences of the United States of America* 73 (9): 3253–57. doi:10.1073/pnas.73.9.3253.

Simonelig, Martine, Kate Elli, Andrew Mitchelsont, and Kevin O Haret. 1996. "Interallelic Complementation at the Supressor of Fork Locus of Drosophila Reveals Complementation Between Suppressor of Forked Proteins Mutated in Different Region." *Genetics*.

Smibert, Peter, Pedro Miura, Jakub O. Westholm, Sol Shenker, Gemma May, Michael O. Duff, Dayu Zhang, et al. 2012. “Global Patterns of Tissue-Specific Alternative Polyadenylation in Drosophila.” *Cell Reports* 1 (3). The Authors: 277–89. doi:10.1016/j.celrep.2012.01.001.

Sokolowski, M B. 1980. “Foraging Strategies of Drosophila Melanogaster: A Chromosomal Analysis.” *Behavior Genetics* 10 (3): 291–302.

Soshnikova, Natalia, Romain Dewaele, Philippe Janvier, Robb Krumlauf, and Denis Duboule. 2013. “Duplications of Hox Gene Clusters and the Emergence of Vertebrates.” *Developmental Biology* 378 (2). Elsevier: 194–99. doi:10.1016/j.ydbio.2013.03.004.

Subtelny, Alexander O, Stephen W Eichhorn, Grace R Chen, Hazel Sive, and David P Bartel. 2014. “Poly(A)-Tail Profiling Reveals an Embryonic Switch in Translational Control.” *Nature* 508 (7494). Nature Publishing Group: 66–71. doi:10.1038/nature13007.

Sun, F L, M H Cuaycong, C A Craig, L L Wallrath, J Locke, and S C Elgin. 2000. “The Fourth Chromosome of Drosophila Melanogaster: Interspersed Euchromatic and Heterochromatic Domains.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (10): 5340–45. doi:10.1073/pnas.090530797.

T. Murata, K. Ogura, R. Murakami et al. 1996. “Hiiragi, a Gene Essential for Wing Development in Drosophila Melanogaster, Affects the Notch Cascade.” *Genes Genet. Syst.*

Takagaki, Y., and J. L. Manley. 1997. “RNA Recognition by the Human Polyadenylation Factor CstF.” *Molecular and Cellular Biology* 17 (7): 3907–14.

Takagaki, Y, C C MacDonald, T Shenk, and J L Manley. 1992. “The Human 64-kDa Polyadenylylation Factor Contains a Ribonucleoprotein-Type RNA Binding Domain and Unusual Auxiliary Motifs.” *Roceedings of the National Academy of Sciences of the United States of America* 89 (4): 1403–7.

Takagaki, Y, and James L Manley. 2000. “Complex Protein Interactions within the Human Polyadenylation Machinery Identify a Novel Component.” *Molecular and Cellular Biology* 20 (5): 1515–25. doi:10.1128/MCB.20.5.1515-1525.2000.

Takagaki, Y, R L Seipelt, M L Peterson, and J L Manley. 1996. “The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation.” *Cell* 87 (5): 941–52.

Takagaki, Yoshio, James L. Manley, Clinton C. MacDonald, Jeffrey Wilusz, and Thomas Shenk. 1990. “A Multisubunit Factor, CstF, Is Required for Polyadenylation of Mammalian Pre-mRNAs.” *Genes and Development* 4 (12): 2112–20. doi:10.1101/gad.4.12a.2112.

Takagaki, Yoshio, Lisa C. Ryner, and James L. Manley. 1988. “Separation and Characterization of a poly(A) Polymerase and a Cleavage/specificity Factor Required for Pre-mRNA Polyadenylation.” *Cell* 52 (5): 731–42. doi:10.1016/0092-8674(88)90411-4.

Takagaki, Yoshio, Lisa C Ryner, and James L Manley. 1989. “Four Factors Are Required for 3’-end Cleavage of Pre-mRNAs.” *Genes and Development*, no. 1988: 1711–24. doi:10.1101/gad.3.11.1711.

Thomsen, Stefan, Ghows Azzam, Richard Kaschula, Lucy S Williams, and Claudio R Alonso. 2010. “Developmental RNA Processing of 3’UTRs in Hox mRNAs as a Context-Dependent Mechanism Modulating Visibility to microRNAs.” *Development*

137 (17): 2951–60. doi:10.1242/dev.047324.

Toba, G, T Ohsako, N Miyata, T Ohtsuka, K H Seong, and T Aigaki. 1999. “The Gene Search System. A Method for Efficient Detection and Rapid Molecular Identification of Genes in Drosophila Melanogaster.” *Genetics* 151 (2): 725–37.

Toma, Daniel P, Kevin P White, Jerry Hirsch, and Ralph J Greenspan. 2002. “Identification of Genes Involved in Drosophila Melanogaster Geotaxis, a Complex Behavioral Trait.” *Nature Genetics* 31 (4): 349–53. doi:10.1038/ng893.

Topalian, Suzanne L, Syuzo Kaneko, Monica I Gonzales, Gareth L Bond, Yvona Ward, and James L Manley. 2001. “Identification and Functional Characterization of Neo-Poly (A) Polymerase , an RNA Processing Enzyme Overexpressed in Human Tumors.” *Molecular and Cellular Biology* 21 (16): 5614–23. doi:10.1128/MCB.21.16.5614.

Trésaugues, Lionel, Pål Stenmark, Herwig Schüler, Susanne Flodin, Martin Welin, Tomas Nyman, Martin Hammarström, Martin Moche, Susanne Gräslund, and Pär Nordlund. 2008. “The Crystal Structure of Human Cleavage and Polyadenylation Specific Factor-5 Reveals a Dimeric Nudix Protein with a Conserved Catalytic Site.” *Proteins* 73 (4): 1047–52. doi:10.1002/prot.22198.

Tyler, David M, Katsutomo Okamura, Wei-Jen Chung, Joshua W Hagen, Eugene Berezikov, Gregory J Hannon, and Eric C Lai. 2008. “Functionally Distinct Regulatory RNAs Generated by Bidirectional Transcription and Processing of microRNA Loci.” *Genes & Development* 22 (1): 26–36. doi:10.1101/gad.1615208.

Ulitsky, Igor, Alena Shkumatava, Calvin H. Jan, Alexander O. Subtelny, David Koppstein, George W. Bell, Hazel Sive, and David P. Bartel. 2012. “Extensive Alternative Polyadenylation during Zebrafish Development.” *Genome Research* 22 (10): 2054–66. doi:10.1101/gr.139733.112.

Ursula Ruegsegger, Katrin Beyer, and Walter Keller. 1996. “Purification and Characterization of Human Cleavage Factor Im Involved in the 3’ □End Processing of Messenger RNA Precursors” 271 (11): 6107–13. doi:10.1074/jbc.271.11.6107.

Venkataraman, Krishnan, Kirk M Brown, and Gregory M Gilmartin. 2005. “Analysis of a Noncanonical poly(A) Site Reveals a Tripartite Mechanism for Vertebrate poly(A) Site Recognition.” *Genes & Development* 19 (11): 1315–27. doi:10.1101/gad.1298605.

Venkatesh, K, and G Hasan. 1997. “Disruption of the IP3 Receptor Gene of Drosophila Affects Larval Metamorphosis and Ecdysone Release.” *Current Biology* 7: 500–509. doi:10.1016/S0960-9822(06)00221-1.

Venken, Koen J T, and Hugo J Bellen. 2005. “Emerging Technologies for Gene Manipulation in Drosophila Melanogaster.” *Nature Reviews. Genetics* 6 (3): 167–78. doi:10.1038/nrg1553.

Venken, Koen J T, Karen L Schulze, Nele a Haelterman, Hongling Pan, Yuchun He, Martha Evans-Holm, Joseph W Carlson, et al. 2011. “MiMIC: A Highly Versatile Transposon Insertion Resource for Engineering Drosophila Melanogaster Genes.” *Nature Methods* 8 (9): 737–43. doi:10.1038/nmeth.1662.

Wahle, Elmar. 1991a. “A Novel poly(A)-Binding Protein Acts as a Specificity Factor in the Second Phase of Messenger RNA Polyadenylation.” *Cell* 67 (3): 640. doi:10.1016/0092-8674(91)90537-9.

———. 1991b. “Purification and Characterization of a Mammalian Polyadenylate

Polymerase Involved in the 3' End Processing of Messenger RNA Precursors." *Journal of Biological Chemistry* 266 (5): 3131–39.

———. 1995. "Poly(A) Tail Length Control Is Caused by Termination of Processive Synthesis." *The Journal of Biological Chemistry* 270 (6): 2800–2808. doi:10.1017/CBO9781107415324.004.

Wang, Ying, Jeffry L. Dean, Ewan K A Millar, Hong Tran Thai, Catriona M. McNeil, Craig J. Burd, Susan M. Henshall, et al. 2008. "Cyclin D1b Is Aberrantly Regulated in Response to Therapeutic Challenge and Promotes Resistance to Estrogen Antagonists." *Cancer Research* 68 (14): 5628–38. doi:10.1158/0008-5472.CAN-07-3170.

Watson, K L, M M Chou, J Blenis, W M Gelbart, and R L Erikson. 1996. "A Drosophila Gene Structurally and Functionally Homologous to the Mammalian 70-kDa S6 Kinase Gene." *Proceedings of the National Academy of Sciences of the United States of America* 93 (24): 13694–98. doi:10.1073/pnas.93.24.13694.

Weitzer, S, and J Martinez. 2007. "The Human RNA Kinase hClp1 Is Active on 3' Transfer RNA Exons and Short Interfering RNAs." *Nature* 447 (7141): 222–26. doi:10.1038/nature05777.

West, Steven, and Nicholas J. Proudfoot. 2008. "Human Pcf11 Enhances Degradation of RNA Polymerase II-Associated Nascent RNA and Transcriptional Termination." *Nucleic Acids Research* 36 (3): 905–14. doi:10.1093/nar/gkm1112.

Wheeler, Guy, Sofia Ntounia-Fousara, Begona Granda, Tina Rathjen, and Tamas Dalmay. 2006. "Identification of New Central Nervous System Specific Mouse microRNAs." *FEBS Letters* 580 (9): 2195–2200. doi:10.1016/j.febslet.2006.03.019.

Wilusz, J, S M Pettine, and T Shenk. 1989. "Functional Analysis of Point Mutations in the AAUAAA Motif of the SV40 Late Polyadenylation Signal." *Nucleic Acids Research* 17 (10): 3899–3908.

Wu, Qi, Tieqiao Wen, Gyunghee Lee, Jae H. Park, Haini N. Cai, and Ping Shen. 2003. "Developmental Control of Foraging and Social Behavior by the Drosophila Neuropeptide Y-like System." *Neuron* 39: 147–61. doi:10.1016/S0896-6273(03)00396-9.

Wu, Qi, Yan Zhang, Jie Xu, and Ping Shen. 2005. "Regulation of Hunger-Driven Behaviors by Neural Ribosomal S6 Kinase in Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 102 (37): 13289–94. doi:10.1073/pnas.0501914102.

Wu, Qi, Zhangwu Zhao, and Ping Shen. 2005. "Regulation of Aversion to Noxious Food by Drosophila Neuropeptide Y- and Insulin-like Systems." *Nature Neuroscience* 8 (10): 1350–55. doi:10.1038/nn1540.

Xiang, Kehui, Liang Tong, and James L Manley. 2014. "Delineating the Structural Blueprint of the Pre-mRNA 3'-end Processing Machinery." *Molecular and Cellular Biology* 34 (11): 1894–1910. doi:10.1128/MCB.00084-14.

Yamamoto, Misato, Ryu Ueda, Kuniaki Takahashi, Kaoru Saigo, and Tadashi Uemura. 2006. "Control of Axonal Sprouting and Dendrite Branching by the Nrg-Ank Complex at the Neuron-Glia Interface." *Current Biology* 16 (16): 1678–83. doi:10.1016/j.cub.2006.06.061.

Yamanaka, Naoki, Kim F Rewitz, and Michael B O 'Connor. 2013. "Ecdysone Control of Developmental Transitions: Lessons from Drosophila Research." *Annu. Rev.*

Entomol 58: 497–516. doi:10.1146/annurev-ento-120811-153608.

Yang, Qin, Molly Coseno, Gregory M Gilmartin, and Sylvie Doublié. 2011. “Crystal Structure of a Human Cleavage Factor CFI(m)25/CFI(m)68/RNA Complex Provides an Insight into poly(A) Site Recognition and RNA Looping.” *Structure* 19 (3): 368–77. doi:10.1016/j.str.2010.12.021.

Yang, Qin, Gregory M Gilmartin, and Sylvie Doublié. 2010. “Structural Basis of UGUA Recognition by the Nudix Protein CFI(m)25 and Implications for a Regulatory Role in mRNA 3’ Processing.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (22): 10062–67. doi:10.1073/pnas.1000848107.

Yang, Yan, Wencheng Li, Mainul Hoque, Liming Hou, Steven Shen, Bin Tian, and Brian D. Dynlacht. 2016. “PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation.” *PLOS Genetics* 12 (1): e1005794. doi:10.1371/journal.pgen.1005794.

Yates, Andrew, Wasiu Akanni, M. Ridwan Amode, Daniel Barrell, Konstantinos Billis, Denise Carvalho-Silva, Carla Cummins, et al. 2016. “Ensembl 2016.” *Nucleic Acids Research* 44 (D1): D710–16. doi:10.1093/nar/gkv1157.

Ye, Bing, Claudia Petritsch, Ira E. Clark, Elizabeth R. Gavis, Lily Yeh Jan, and Yuh Nung Jan. 2004. “Nanos and Pumilio Are Essential for Dendrite Morphogenesis in Drosophila Peripheral Neurons.” *Current Biology* 14 (4): 314–21. doi:10.1016/S0960-9822(04)00046-6.

Yu, Fengwei, Yu Cai, Rachna Kaushik, Xiaohang Yang, and William Chia. 2003. “Distinct Roles of Gai and GB13F Subunits of the Heterotrimeric G Protein Complex in the Mediation of Drosophila Neuroblast Asymmetric Divisions.” *Journal of Cell Biology* 162 (4): 623–33. doi:10.1083/jcb.200303174.

Zaharieva, Emanuela, Irmgard U Haussmann, Ulrike Bräuer, and Matthias Soller. 2015. “Concentration and Localization of Coexpressed ELAV/Hu Proteins Control Specificity of mRNA Processing.” *Molecular and Cellular Biology* 35 (18): 3104–15. doi:10.1128/MCB.00473-15.

Zhao, Jing, Linda Hyman, and C Moore. 1999. “Formation of mRNA 3’ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis.” *Microbiology and Molecular Biology Reviews* 63 (2): 405–45.

Zhao, Xiao Li, and Ana Regina Campos. 2012. “Insulin Signalling in Mushroom Body Neurons Regulates Feeding Behaviour in Drosophila Larvae.” *The Journal of Experimental Biology* 215 (Pt 15): 2696–2702. doi:10.1242/jeb.066969.

Zinke, I, C Kirchner, L C Chao, M T Tetzlaff, and M J Pankratz. 1999. “Suppression of Food Intake and Growth by Amino Acids in Drosophila: The Role of Pumpless, a Fat Body Expressed Gene with Homology to Vertebrate Glycine Cleavage System.” *Development* 126 (23): 5275–84.

Books cited in this study:

Ashburner, M., 1989. *Drosophila*. A laboratory handbook. Cold Spring Harbor Laboratory Press.

Campos-Ortega, J.A. and Hartenstein, V., 2013. The embryonic development of *Drosophila melanogaster*. Springer Science & Business Media.

Hartenstein, V., 1993. Atlas of *Drosophila* development (Vol. 328). Cold Spring Harbor Laboratory Press.

Sadler, T.W., 2011. Langman's medical embryology. Lippincott Williams & Wilkins.

Personal communications to FlyBase cited in this study:

Gallant, P. (2016.4.27). Sym and Madm deletions and rescue constructs (Personal communication to FlyBase)

Kaufman, T. (2009.6.18). drk, cbc, bbc & cid mutations (Personal communication to FlyBase)