Investigation of pathways responsible for repeat RNA-mediated cellular perturbation in *Drosophila* models of dominant expanded repeat disease

A thesis submitted for the degree of Doctor of Philosophy

September 2011

Kynan Thomas Lawlor, B.Sc. (Mol. Biol.) (Hons.)

Discipline of Genetics

School of Molecular and Biomedical Science

The University of Adelaide

Table of Contents

Corrections

Chapter 1
Page 10, paragraph 2 should read "rather *than* enhancement"

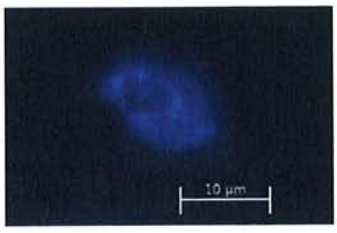
Chapter 2
Page 35, **Quantification of tergite disruption**, should include the paragraph:
The scoring scheme was based on the number and severity of disrupted tergites, using particular morphological attributes to define each category, thus minimising any experimenter bias. Preliminary data showed no significant difference (data not shown) between populations when scoring 'experimenter blind'. As such, remaining experiments were not scored blind. The order in which genotypes were scored each day was randomised and data from multiple sets of progeny obtained from multiple sets of parents on different days was used in each case.

Page 36, **Quantification of locomotion phenotype**, should include the paragraph:
Scoring involved reviewing the video to tally the time in seconds that each fly spent either upright (walking or standing) or on its back As the possibility for experimenter bias in this case appeared negligible scoring was not done 'blind'.

Page 40, **Climbing assays**, should include the clarification :
n = 3 biological replicates (sets), with 20-25 animals per genotype, per biological replicate (set), for a total of 60-75 animals examined for each genotype. A climbing score representing each biological replicate (set) was obtained by calculating the mean from 5 consecutive trials for each genotype. A final genotype score was obtained by calculating the mean of all 3 biological replicates.

Chapter 3
Page 46, **Figure 3.1** legend should include the paragraph:
Fisher's exact test does not include a calculation of standard deviation, or standard error, however 95% confidence intervals were calculated for each particular proportion. As this involved a separate calculation these values are included in Appendix 1, rather than as error bars.

Chapter 4
Page 69, In **Figure 4.1 C**, DAPI staining was poorly reproduced in the printed version. Images were chosen based on being representative of each genotype in regard to repeat RNA staining (Cy3 signal), with DAPI included as a guide to the location of the nucleus only. As such the relative levels of DAPI signal do not change the interpretation of the data. A modified version (to improve visibility in printed form) of the DAPI staining shown in 4.1 C is included below.

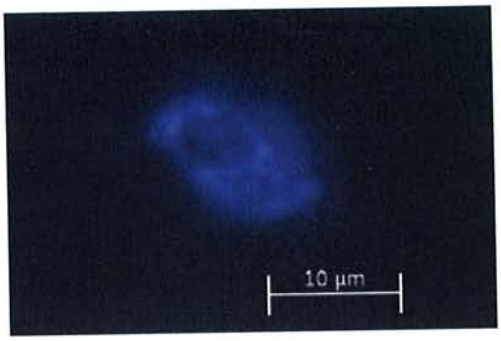


Page 77, paragraph 1, should include the paragraph:
In this study CUG-specific RNA localisation patterns were observed in independent samples from independent transgenic lines and thus the result appears robust. However, as quantification of foci was not performed, further analysis would be necessary to confirm the more subtle differences in CUG-specific localisation patterns observed in different repeat expression contexts.

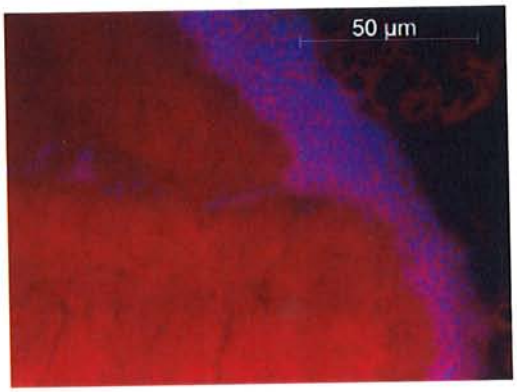
Page 80, paragraph 2, should include the sentences:
Confocal microscopy was not performed in this case. Techniques allowing higher imaging resolution may confirm the absence of neuronal foci in *Drosophila* with more certainty.

Scale bars were initially not included in fluorescent micrographs. Examples for each type of image taken are included below to aid in interpretation of these results.

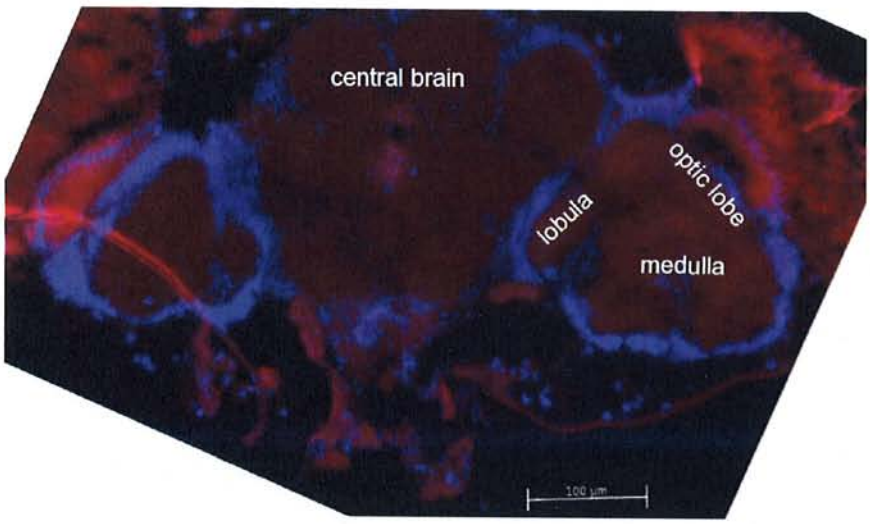
An example of muscle nuclei (As in 4.1, 4.3, 4.4, 4.5). All images were captured and cropped in the same way such that scale is identical :



An example of an adult brain at higher magnification (as in 4.6 B-D) :



An example of an adult brain at lower magnification (as in 4.6 A). In this case landmarks within the brain are annotated to further aid in interpretation.



Page 82, paragraph 1, should read :
"... support *the conclusion* that pathways ....."

Chapter 5
Page 89, paragraph 3, should read:
" ... indicate that rather *than* the insertion directly disrupting ...."

Chapter 6
Page 118, Figure 6.2 figure legend, should state:
All flies were aged for 35 days before sectioning (Materials and Methods).

Chapter 7
Page 135, paragraph 2, should read:
"In support *of* this we see alterations to miRNA profiles...."

Page 135, paragraph 3, should read:
".... indicating that, as in our model, complementary transcripts form double-stranded RNA that is processed."

Index of Figures and Tables

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Kynan Lawlor and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Thesis Program (ADTP) and the also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Kynan Thomas Lawlor

Acknowledgements

I wish to thank the following for their contribution to this project : my supervisor Robert Richards for wise guidance and support; Louise O'Keefe, my second supervisor, for excellent ideas, mentorship and support, especially during the writing of this thesis. Past and present members of the Richards lab for expert assistance with experiments and sharing their wisdom and good humour over the years, especially Clare van Eyk, Saumya Samaraweera, Amanda Choo and Sonia Dayan ; my friends and family, particularly Merridy Lawlor for always being there and finally, my wife Jessica, as this work would not have been possible without her endless encouragement, support and patience.

Abbreviations

°C : degrees Celsius
% : percentage
µg : micrograms
µL : microlitre
µm : micrometre
A : adenosine
AR : androgen receptor
ATN1 : atrophin 1
ATXN1 : ataxin 1
ATXN2 : ataxin 2
ATXN3 : ataxin 3
ATXN7 : ataxin 7
ATXN8OS : ataxin 8 opposite strand
ATXN10 : ataxin 10
BAC : bacterial artificial chromosome
bp : base pairs
C : cytosine
CACNA1A : calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
cDNA : complementary DNA
CLC-1 : Chloride channel 1
CNBP : CCHC-type zinc finger, nucleic acid binding protein
CUG-BP : CUG binding protein
da : daughterless
DAPI : 4'-6-diamido-2-phenylindole
DIC : differential interference contrast
DM1 : myotonic dystrophy type 1
DM2 : myotonic dystrophy type 2
DMPK : dystrophia myotonica protein kinase
DNA : deoxyribonucleic acid
dNTP : deoxyribonucleoside triphosphate
DRPLA : dentatorubral-pallidoluysian atrophy
dsRNA : double-stranded RNA
DTT : dithiothreitol
EDTA : ethylene diamine tetra-acetic acid
elav : embryonic lethal abnormal vision
ERG : electroretinogram
FMR1 : fragile X mental retardation 1
FXTAS : fragile X tremor-ataxia syndrome
G : guanosine
GFP : green fluorescent protein
GMR : Glass multimer reporter
HD : Huntington's disease
HDL-2 : Huntington's disease like 2
hnRNP : heterogenous nuclear ribonucleoprotein
HTT : huntingtin
JPH3 : juntophilin 3
kb : kilobase

KLHL1 : kelch-like 1
M : molar
mbl : muscleblind
MBNL : muscleblind-like
mg : milligrams
miRNA : micro RNA
mL : millilitres
mM : millimolar
MQ H_2O : Milli-Q (Millipore) ultrapure H_2O
mRNA : messenger RNA
mV : millivolts
ng : nanograms
PBS : phosphate buffered saline
PCR : polymerase chain reaction
pmol : picomole
polyQ : polyglutamine
polyL : polyleucine
PPP2R2B : protein phosphatase 2, regulatory subunit B, beta isoforms
RNA : ribonucleic acid
RNAi : RNA interference
rpm : revolutions per minute
RT-PCR : reverse transcription polymerase chain reaction
SBMA : spinal bulbar muscular atrophy
SCA : spinocerebellar ataxia
siRNA : small interfering RNA
SSC : saline sodium citrate
T : thymine
TAE : tris-acetate EDTA
TBE : tris-borate EDTA
TBP : TATA box binding protein
U : uracil
UAS : upstream activating sequence
UTR : untranslated region

Abstract

The expansion of polymorphic repeat sequences within unrelated genes is responsible for pathology in a family of dominant human diseases. Based on clinical and genetic similarities, it is hypothesised that common pathways may contribute to all of these diseases, with evidence for a number of mechanisms mediated by the expanded repeat. Where the repeats are translated, a long polyglutamine protein has been shown to have pathogenic properties. However, the identification of diseases caused by untranslated repeats has led to the discovery of repeat RNA-mediated pathogenic pathways. As expanded repeat-containing transcripts are present in the case of both translated and untranslated repeats, repeat RNA is a candidate common pathogenic agent. Therefore, determining its contributions to pathology will be important in understanding these diseases.

Using the model organism *Drosophila melanogaster,* this study identifies common CUG and CAG repeat RNA-mediated phenotypes, enabling the investigation of common pathways of cellular perturbation. Ubiquitous expression of either repeat sequence led to reduced viability and disruption to the development of the adult dorsal abdominal tergites through a specific effect on histoblast cells. This phenotype provides a biological read-out of common RNA-mediated effects, enabling examination of the pathways involved by quantifying the changes in the phenotype when specific candidate genes are genetically altered. Tergite disruption was not strongly modified by reducing activity of the well-characterised *muscleblind* mediated pathway. Furthermore, the presence of specific nuclear RNA foci, an indicator of repeat RNA-mediated protein sequestration, was not correlated with the phenotype. Results indicate that tergite disruption is not strongly dependent on muscleblind sequestration and may involve an alternative pathway. Ectopic expression of either repeat did not cause significant phenotypes in the eye, or neurons, except in the case of one fortuitous transgene insertion. In this case, bi-directional transcription of the repeat tract facilitated by an endogenous promoter was necessary for pathology, providing support for a novel pathway of pathology involving the formation of double-stranded RNA. Subsequent comparison of the pathways involved in hairpin-forming single stranded RNA, and bi-directional double-stranded RNA mediated phenotypes in *Drosophila* supports the existence of multiple distinct pathways that contribute to cellular perturbation.

CHAPTER 1 : Introduction and background

1.1 Human expanded repeat disease

The genetic basis for many human diseases is now well characterised, and knowledge of the mechanisms leading from mutation to disease is rapidly increasing. Discoveries 20 years ago identified a novel form of 'dynamic mutation' involving the expansion of polymorphic repeat sequences within the genome [1, 2]. In subsequent years, at least 22 genetic diseases have been identified that are caused by the expansion of polymorphic repeat sequences [3-5]. These include Huntington's disease (HD), fragile X syndrome, Friedreich's ataxia, spinal and bulbar muscular atrophy (SBMA), dentatorubral-pallidoluysian atrophy (DRPLA), spinocerebellar ataxia (SCA) type 1, 2, 3, 6, 7, 8, 10, 12 and 17, myotonic dystrophy type 1 (DM1) and 2 (DM2), fragile X tremor/ataxia syndrome (FXTAS) and Huntington's disease like-2 (HDL-2). In effected individuals, repeats expand beyond a pathogenic threshold resulting in a series of diseases that are distinct, but clinically and genetically similar in many cases (Table 1.1). The majority of diseases are caused by trinucleotide CAG or CUG repeat transcripts, with diseases also identified that are associated with CGG, CCUG and AUUCU repeats (Table 1.1).

1.1.1 A common molecular basis for dominant expanded repeat disease

With the discovery of an increasing number of expanded repeat diseases, attempts have been made to understand the basis for pathology in each case, with a number of common and distinct features identified. In a specific group of diseases including fragile X syndrome and Friedreich's ataxia repeat expansion leads to disease through loss of function of the repeat containing gene [6, 7]. In these cases, disease is associated with pathogenic repeat lengths greater than approximately 200, resulting in an effect on the repeat containing gene [4]. In the case of fragile X syndrome, this involves hypermethylation, and silencing of the repeat containing FMR1 gene [8, 9], while Friedreich's ataxia involves the inhibition of transcriptional elongation due to

an expanded repeat within the first intron of the affected gene [10]. Disease is developmentally based and results from loss of function of the repeat containing gene, leading to subsequent recessive or X-linked inheritance [11-13]. Consistent with this, knocking out gene function in animal models replicates many of the features of these diseases [14, 15].

Table 1.1 Common features of the human dominant expanded repeat diseases

Disease (gene)	Repeat sequence	Clinical symptoms*	Normal length	Pathogenic length
DM1 (*DMPK*)	CUG (3'UTR}	M, I, Ca, Co	5-37	50- >3500
DM2 (*CNBP*)	CCUG (intron)	M, I, Ca	10-26	75- ~11000
DRPLA (*ATN1*)	CAG (exon)	A, Co, T, D	3-36	49-88
FXTAS (*FMR1*)	CGG (5'UTR)	A, Co, Pa, T	6-52	60-200
HD (*HTT*)	CAG (exon)	Ch, Co, D, Dy, Ps	6-35	36-121
HDL-2 (*JPH3*)	CUG (3'UTR)	Ch, Dy, Co	6-28	40-59
SBMA (*AR*)	CAG (exon)	Mo	9-36	40-55
SCA1 (*ATXN1*)	CAG (exon)	A, Co, Dr	6-39	39-81
SCA2 (*ATXN2*)	CAG (exon)	A, T, Pa	13-33	>34
SCA3 (*ATXN3*)	CAG (exon)	A, Dy, D	13-44	>55
SCA6 (*CACNA1A*)	CAG (exon)	A, Dr, T	4-18	20-29
SCA7 (*ATXN7*)	CAG (exon)	A, Dr	4-35	37-306
SCA8 (*ATXN8OS*)	CUG (3'UTR)	A, Dr, N	<50	74-1300
SCA10 (*ATXN10*)	AUUCU (intron)	A, Dr, T, S	10-29	800-4500
SCA12 (*PPP2R2B*)	CAG (5'UTR)	A, D, T, Ps	4-32	51-78
SCA17 (*TBP*)	CAG (exon)	A, Ch, Co, D, Dy, S	25-42	47-63

*Clinical symptoms : A – Ataxia, Ca – Cardiac defects, Ch – Chorea, Co - Cognitive defects, D – Dementia, Dr – Dysarthria, Dy – Dystonia, I – Insulin resistance, M – Myotonia, Mo – Motor weakness, N – Nystagmus, Pa – Parkinsonism, Ps – Psychiatric problems, S – Seizures, T – Tremor,
DMPK - dystrophia myotonica-protein kinase, CNBP - CCHC-type zinc finger, nucleic acid binding protein, FMR1 - fragile X mental retardation 1, ATN1 - atrophin 1, HTT - huntingtin, JPH3 - junctophilin 3, AR - androgen receptor, ATXN1 - ataxin 1, ATXN2 - ataxin 2, ATXN3 - ataxin 3, CACNA1A - calcium channel, voltage-dependent, P/Q type, alpha 1A subunit, ATXN7 - ataxin 7, ATXN8OS - ataxin 8 opposite strand, ATXN10 - ataxin 10, PPP2R2B - protein phosphatase 2, regulatory subunit B, beta isoform, TBP - TATA box binding protein.

Features of a number of the human dominant expanded repeat diseases. Each disease is associated with a repeat sequence located within a different gene, where the location within the gene is shown in brackets. UTR – untranslated region. Common and distinct clinical symptoms associated with each disease are shown, where letters refer to conditions as detailed below the table. Normal and pathogenic repeat length are shown, where normal lengths are associated with unaffected individuals in the population, while pathogenic lengths have been observed in individuals with the disease. Information from [3, 4, 13, 16-30]

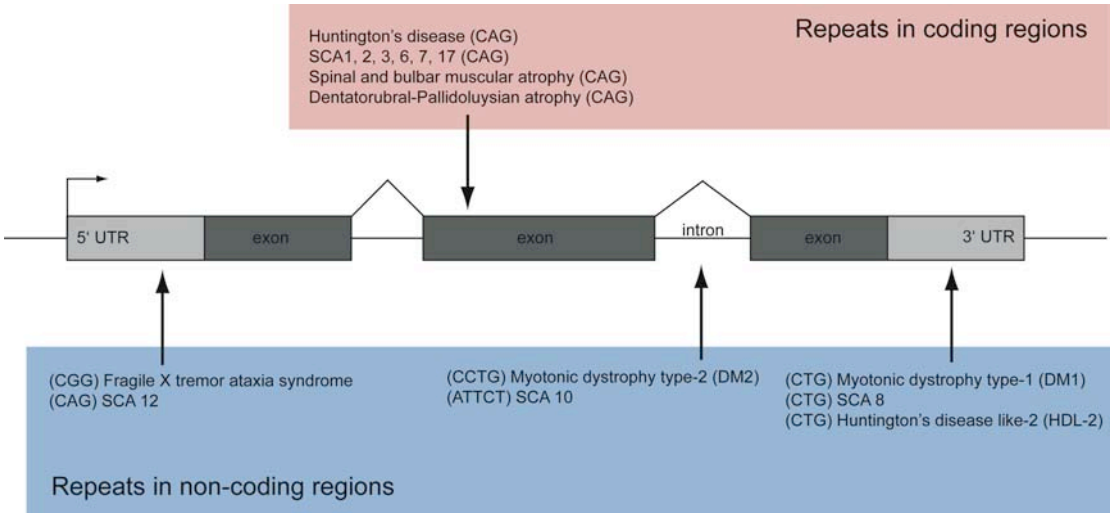
However, a majority of the repeat associated diseases are inherited in a dominant manner that is not consistent with a loss-of-function mechanism. In these cases identical, or similar, repeat sequences are found in a series of unrelated genes (Table 1.1). These genes have distinct, diverse functions, and knocking out particular genes in animal models does not appear to significantly reproduce the symptoms of disease [4, 13, 31, 32]. In many cases these diseases have overlapping clinical features involving late onset neurodegeneration, movement defects and specific cellular loss (Table 1.1). For example, DRPLA and the SCAs all share clinical features, such as ataxia, and can be difficult to distinguish based on observation alone, while HD, DRPLA, and some of the SCAs lead to cognitive defects and dementia (Table 1.1). In most expanded repeat diseases, onset occurs at middle age and results in death within 10-20 years [13]. Additionally, different repeat sequences in unrelated genes give rise to similar diseases, for example, in the case of HD (translated CAG repeat) and HDL-2 (untranslated CUG), and DM1 (untranslated CUG) and DM2 (untranslated CCUG) [33-35].

The dominant repeat diseases also share a number of genetic features. In most cases the repeat copy number threshold at which disease is observed is between 30-50 (Table 1.1), with increasing disease severity and decreasing age of onset correlated with longer repeat lengths [4, 13]. In families a phenomenon known as anticipation occurs, whereby successive generations tend to show earlier age of onset, and increased severity [5]. The basis for this is the expansion of repeat copy number across generations, through a mechanism that appears to be common to all diseases [5, 36]. Based on the similarities between the dominant expanded repeat diseases, it has been proposed that a common pathogenic mechanism may contribute to pathology and this has been a focus of work in our group [5, 37, 38]. Nonetheless, many diseases show unique clinical and genetic features, and a growing body of evidence describes the role of mechanisms specific to each repeat-containing gene [3, 4, 13].

One difficulty in defining a common pathogenic agent is that repeat sequences are found in both protein coding and non-coding regions, yet give highly similar pathology. Disease causing repeats within protein coding regions lead to the production of an expanded protein tract that has been shown to be pathogenic (Figure

1.1). However, the identification of pathogenic repeats within non-coding regions (Figure 1.1), producing an expanded but untranslated repeat, has led to evidence that repeat RNA is also able to mediate pathology. It seems increasingly likely that multiple pathways contribute to disease, including common repeat protein and/or RNA-mediated pathways, as well as specific interactions unique to each repeat-containing gene. Understanding the mechanisms by which each pathway acts, and the relative contribution of each to disease is now a focus within the field. The work described in this thesis focuses on the role of expanded repeat RNA as a common contributor to pathology by using the model organism *Drosophila melanogaster* to examine the genetic pathways leading from expanded repeat to cellular perturbation.

Figure 1.1 Gene location of repeat tracts causing dominant expanded repeat disease



Locations of repeats that cause dominant expanded repeat disease. Schematic indicates the gene region in which the repeat is located in each case, in relation to a hypothetical gene, where each repeat is actually located in a different gene. 5' and 3' untranslated regions (UTR) are shown in light grey, while exons are shown in dark grey. Introns are indicated as lines between exons. Repeats that encode a polyglutamine protein are within the pink box, while repeats within non-coding regions are shown within the blue box.

1.2 Expanded polyglutamine protein repeat-mediated pathology

HD, SBMA, SCA type 1, 2, 3, 6, 7 and 17 and DRPLA are all caused by CAG repeat expansions within coding regions of unrelated genes (Table 1.1, Figure 1.1). This leads to translation of an expanded polyglutamine tract within the resulting protein in each case, however, loss of protein function does not appear to account for a majority of the symptoms of disease [4, 39]. In animal models, expression of a translated CAG repeat within a number of disease protein contexts, or within an unrelated gene, leads to neurodegeneration, suggesting that pathology is the result of a dominant repeat mediated effect [40-42]. Likewise, expression of a polyglutamine tract alone leads to repeat length-dependent pathology, suggesting that polyglutamine protein is inherently toxic [37, 43].

Several mechanisms have been proposed to explain how polyglutamine exerts a toxic effect. Post-mortem brain tissue from polyglutamine disease patients and tissue from animal models, has been shown to contain protein inclusions that contain polyglutamine protein and are present in either the nucleus, cytoplasm, or both [44]. Therefore, the role of inclusions in pathology has been a focus of research [45, 46]. Ubiquitin is found within polyglutamine inclusions, and molecular chaperones such as Hsp70 have been shown to modify polyglutamine toxicity in *Drosophila* and mouse [47, 48, 49]. Inhibition of the ubiquitin proteasome system has therefore been investigated as a possible pathogenic pathway however conflicting results exist both in support of and against a role in pathology [50, 51]. Similarly the role, if any, of protein inclusions in pathology remains unclear with contrasting results suggesting they may be either pathogenic, or protective [52-55]. More recently, it has been proposed that polyglutamine may act through an effect on autophagy, with some evidence for this pathways involvement in HD [3, 56]. Additionally, a number of polyglutamine interactions have been identified with transcription factors such as CREB-binding protein, with some evidence that transcriptional dysregulation may be a general polyglutamine mediated pathogenic pathway [16, 57, 58]. In this case it has been proposed that soluble polyglutamine, rather then aggregates, may mediate this interaction, providing an explanation for observations that polyglutamine aggregates are not toxic [59].

In addition to intrinsic toxicity, expanded polyglutamine protein may contribute to disease by perturbing the normal function of the protein in which the repeat resides [3]. Studies indicate that in some cases different phosphorylation states of the repeat-containing protein can alter toxicity based on modulation of normal protein interactions [60-62]. Post-translation modification in general is proposed to be a contributor to expanded repeat mediated pathology, and may account for context specific effects in some diseases [63, 64].

In support of expanded repeat-containing protein toxicity as a primary contributor to pathology, in some cases translocation of repeat containing protein to the nucleus is necessary to observe effects. In SBMA, where the repeat is found within the *androgen receptor (AR)* gene, testosterone-mediated nuclear translocation is required for disease [65]. Likewise, mutation of the nuclear localisation sequence in ataxin 1 (ATXN1), the repeat-containing protein responsible for SCA1, is able to prevent pathology [66]. Caspase cleavage of huntingtin (HTT), the repeat-containing protein that causes HD, is required for pathology in some cases, indicating that protein mediated toxicity plays an important role in this disease also [67]. In each case these studies were performed in mice, and determining the extent to which this situation is relevant to human patients will be important in understanding the relative contribution of polyglutamine protein to human disease.

-

1.3 Expanded repeat RNA-mediated pathology

A second group of expanded repeat diseases, including DM1, DM2, FXTAS, SCA8, 10 and 12, and HDL-2, are caused by repeats within non-coding regions (Table 1.1, Figure 1.1). Repeat tracts associated with these diseases are within regions that are not predicted to be translated, and some consist of tetra and pentanucleotide repeats that cannot encode a pure polyglutamine protein tract (Table 1.1, Figure 1.1). The effects of polyglutamine protein are therefore not expected to account for pathology, however most of these diseases are clinically and genetically similar to the polyglutamine diseases (Table 1.1), suggesting that common pathogenic pathways may contribute in each case. Examining the basis for pathology has identified repeat RNA-mediated pathogenic mechanisms and there is now evidence supporting a role for these in most of the non-coding repeat diseases, as well as emerging evidence for these pathways in polyglutamine disease. The remainder of this chapter will summarise evidence for repeat RNA-mediated pathology in each of the non-coding expanded repeat diseases, and proposed pathways through which repeat RNA may contribute as a common pathogenic agent.

1.3.1 Myotonic dystrophy type 1 (DM1) and 2 (DM2)

DM1 is a dominantly inherited degenerative muscular disease that begins later in life and results in myotonia, cardiac defects, insulin resistance and cognitive dysfunction [18, 35, 68-70]. DM1 shows clinical anticipation, whereby successive generations tend to have a more severe disease, with an earlier age of onset that is correlated with repeat tract length [4, 36]. Disease is caused by the expansion of a CUG repeat within the 3'UTR of the *dystrophia myotonica protein kinase* (*DMPK*) gene transcript [71]. *DMPK* knockout mice do not reproduce all symptoms of DM1, suggesting that loss of function of this gene is not the primary basis for pathology [72]. However, mice expressing only the untranslated CUG repeat within the context of *human skeletal actin*, a gene unrelated to *DMPK*, show a phenotype similar to the clinical features of DM1 [73]. This result suggests that expression of the CUG repeat RNA is sufficient to cause pathology, indicative of a dominant repeat RNA-mediated effect. DM2 is

caused by the expansion of a CCUG repeat within an intron of the *zinc finger protein 9* (now known as *CCHC-type zinc finger, nucleic acid binding protein, [CNBP]*) gene transcript [74]. *CNBP* and *DMPK* are unrelated genes, yet both DM1 and DM2 show similar clinical features, providing further support for the theory that expanded repeat RNA is responsible for pathology.

In examining the basis for DM1 pathology, it was observed that the expanded CUG repeat transcript is retained in the nucleus where it forms discrete RNA foci [75, 76]. CUG RNA is also able to form hairpin structures that bind to the Mucleblind-like (MBNL) proteins [77, 78]. Consistent with this, MBNL-1 protein co-localises with nuclear foci, suggesting that it is sequestered to these sites by repeat RNA [77].

MBNL-1 function was first described in *Drosophila*, where the ortholog *muscleblind (mbl)* was originally identified as a protein involved in photoreceptor and neural development and muscle function [79, 80]. It has since been shown to be an RNA binding protein that regulates alternative splicing [81]. MBNL-1 regulates alternative splicing by binding CHHG and CHG (where H=U, A or C) repeat containing mRNA [81-83]. MBNL-1 sequestration leads to dysregulation of alternative splicing, such that specific MBNL-1 regulated genes are mis-spliced, accounting for some of the clinical symptoms of the disease [81, 84]. For example, changes in chloride conductance leading to myotonia are caused by mis-splicing of the *chloride channel 1* (*ClC-1)* gene which encodes a chloride channel protein [77]. CCUG RNA in DM2 also forms foci that co-localise with MBNL-1 and alterations in splicing similar to those in DM1 are observed, suggesting that this may be a common pathogenic mechanism [74, 85].

Supporting a role for CUG RNA as a pathogenic agent, multiple studies have shown that expression of untranslated CUG repeats in *Drosophila*, *Caenorhabditis elegans* and cell culture models replicates the altered splicing and muscle cell perturbation seen in DM1 patients [86-89]. Furthermore, loss of MBNL-1 function in mouse, or muscleblind in *Drosophila*, leads to muscle abnormalities and alternative splicing defects similar to those seen in DM1 [90, 91]. Similarly, overexpression of MBNL-1 in mouse and *muscleblind* in *Drosophila* suppresses CUG repeat RNA effects [87, 89, 92].

A second protein involved in alternative splicing, CUG-binding protein (CUG-BP), has also been shown to play a role in DM1. CUG-BP was initially identified for its ability to bind CUG repeat RNA, although it is not found in nuclear foci in DM1 [93, 94]. Steady state levels of CUG-BP are increased in disease, with some evidence that this may occur through hyper-phosphorylation by Protein Kinase C (PKC) [95]. A mouse model over-expressing CUG-BP reproduces several features of DM1 supporting the involvement of this pathways in disease [96]. CUG-BP is a member of the CELF family of alternative splicing factors that is antagonised by MBNL-1 and hence CUG-BP and MBNL have opposing roles, however the dynamics of splicing dysregulation in DM remains a subject of investigation [34, 68].

Although much work has focused on pathology in muscle cells, DM1 also involves cognitive defects, with some evidence for the involvement of similar pathogenic pathways in neuronal cells [18, 97, 98]. Expanded *DMPK* is expressed in the brain and forms RNA foci that co-localise with MBNL-1, with subsequent alterations in splicing observed in neurons [99, 100]. Therefore, the CUG repeat RNA-mediated mechanism identified in DM1 may also be relevant to other expanded repeat diseases that involve primarily neuronal pathology.

1.3.2 Fragile X tremor ataxia syndrome (FXTAS)

FXTAS is a late onset neurodegenerative disease caused by intermediate length CGG expansions in the 5' UTR of the *Fragile X Mental Retardation 1* (*FMR1)* gene [101]. Longer CGG expansions lead to hypermethylation and loss of gene transcription resulting in Fragile X Syndrome, however, in FXTAS transcription is not prevented and a different, dominantly inherited disease occurs [6, 101, 102]. In many cases FXTAS patients appear to have an elevated level of *FMR1* transcript and only a slightly reduced or normal level of protein, highlighting the possibility of RNA-mediated pathology [101]. Consistent with this, expression of an untranslated CGG repeat in the cells of the *Drosophila* eye causes degeneration that is dependent on age, repeat length and dosage [103]. Subsequently, observations in mouse models further support that expanded CGG repeat RNA is toxic in neurons [104, 105].

Intranuclear protein inclusions have been found in post-mortem brain tissue from FXTAS patients and in brains from a mouse model of FXTAS [104, 106]. Likewise, repeat expresson in *Drosophila* also induces the formation of ubiquitin positive inclusions [103]. Examination of these inclusions revealed that expanded *FMR1* RNA is present as well as RNA binding proteins heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP A2/B1) and MBNL-1, suggesting that a mechanism involving repeat RNA-mediated sequestration of RNA binding proteins, as in DM1, may be involved [107]. Consistent with this, structural studies suggest that CGG RNA is able to form stable secondary structures, similar to those formed by CUG RNA [108]. In further support of RNA mediated protein sequestration in FXTAS, Pur-alpha and hnRNP A2/B1 have been identified as CGG binding proteins and overexpression of each in *Drosophila* is able to suppress CGG mediated toxicity [109, 110]. Pur-alpha is found in inclusions in FXTAS patient brains, and phenotypes of *Pur-alpha* knockout mice indicate that this protein is necessary for neuronal survival, supporting its involvement in disease [109, 111]. HnRNP A2/B1 has multiple roles in RNA metabolism, including mRNA localisation and splicing, and therefore the ability of hnRNP A2/B1 to bind CGG repeats suggests a role for alterations in normal RNA metabolism in FXTAS [110, 112]. Consistent with this pathogenic pathway, a recent study suggests that overall impairment of nuclear mRNA export may contribute to FXTAS [113].

HnRNP A2/B1 is also able to form a complex with CUG-BP, and therefore sequestration may lead to dysregulation of similar pathways to those observed in DM1 [110]. In further support of common pathogenic pathways, CGG repeat RNA responsible for FXTAS forms RNA aggregates in cell culture that sequester the RNA binding protein Sam68, which is involved in alternative splicing, leading to sequestration of MBNL-1 and dysregulation of splicing [114]. However, in a *Drosophila* model, overexpression of CUG-BP led to suppression of toxicity, rather then enhancement as observed in models of DM1, suggesting that this pathway may be differently perturbed in FXTAS [110]. Likewise, Pur-alpha and hnRNP A2/B1 appear to interact with CGG RNA, but not CAG or CUG RNA [109], suggesting that FXTAS pathology may involve distinct pathways unique to this disease.

1.3.3 Spinocerebellar ataxia type 8 (SCA8)

SCA8 is a late onset, dominant, neurodegenerative disease that is caused by an expanded CUG repeat in the 3' UTR of the *ATXN8OS* transcript [115]. This transcript does not encode a protein, and given that CUG repeats have been shown to cause disease in DM1, it is hypothesised that a similar mechanism may be responsible for SCA8 [115]. In support of this, expression of the expanded *ATXN8OS* transcript in the *Drosophila* eye caused cellular disorganisation and degeneration [116]. This model was used to identify the RNA binding proteins muscleblind and staufen as modifiers of pathology [116]. The identification of muscleblind suggests that SCA8 may share aspects of pathology with DM1 [116]. Additionally, staufen is a double-stranded RNA binding protein involved in *Drosophila* development, while the human ortholog is involved in neuronal mRNA transport, and regulation of translation [117, 118]. It is therefore possible that staufen may be sequestered or dysregulated by the expanded SCA8 transcript, preventing its normal role in neuronal maintenance.

Mice expressing the expanded *ATXN8OS* gene within a bacterial artificial chromosome (BAC) under control of its normal regulatory sequences show a progressive neurological phenotype [119]. While no RNA foci were observed in this model, ubiquitin-positive protein inclusions were found in the brain [119]. The authors discovered that the expanded repeat is transcribed from the opposite strand to produce a CAG repeat encoded polyglutamine protein and suggest that the phenotype is caused by both RNA and protein-mediated mechanisms [119]. Further work using a mouse model identified CUG RNA foci in the brain that co-localise with MBNL-1 [120]. Reducing MBNL-1 levels enhanced motor deficits in mice, suggesting a role for this pathway in pathogenesis [120]. Consistent with this, specific alterations in splicing and subsequent upregulation in the CNS of *GABA-A*, a gene required for neuronal function, were identified in both mouse brains and human autopsy tissue [120].

1.3.4 Huntington's disease like-2 (HDL-2)

HDL-2 is a late onset neurodegenerative disease with symptoms frequently indistinguishable from HD [121]. HDL-2 is caused by an expanded CUG repeat in the 3' UTR of the *Junctophillin 3* (*JPH3*) transcript [122]. Expression of *JPH3* with an expanded CUG repeat in PC12 cells leads to cellular toxicity and the formation of nuclear foci that co-localise with MBNL-1 protein [123]. This suggests that HDL-2 may be caused by a dominant repeat RNA-mediated mechanism and that expanded RNA is able to cause pathology similar to that caused by polyglutamine protein in HD. However, this similarity may also be explained by evidence for polyglutamine mediated pathology in HDL-2, with ubiquitin-positive protein inclusions detected in post-mortem brain tissue from patients [33, 121]. Subsequent work in a BAC transgenic mouse model identified an antisense transcript spanning the repeat such that a polyglutamine protein may be produced, resulting in nuclear protein inclusions similar to those observed in patients [124]. This transcript has not been previously identified in human patients [122, 125], however, if this mechanism is conserved both repeat RNA and protein may contribute to pathology in HDL-2.

1.3.5 Spinocerebellar ataxia type 10 (SCA10)

SCA10 is a dominant, late onset neurodegenerative disease caused by an untranslated AUUCU repeat in intron 9 of the *ATXN10* gene transcript [126]. In lymphoblastoid cell lines from affected patients, *ATXN10* transcription is not reduced compared to unaffected individuals, suggesting that expanded repeat RNA rather than loss of gene transcription is responsible for the disease [126]. In further support of this, a mouse knock-out of the *SCA10* gene is homozygous lethal but gives no phenotype in heterozygotes [127].

Structural studies suggest that the AUUCU repeat RNA is able to form a hairpin structure under physiological conditions and therefore the transcript may be able to bind RNA binding proteins, with polypyrimidine tract binding protein suggested as a candidate [127]. Preliminary studies suggest that expanded *SCA10* transcripts are able to form foci when transfected into cell cultures [128]. Subsequent work has

confirmed in cell culture and mouse that AUUCU RNA is able to form foci and bind heterogeneous nuclear ribonucleoprotein K (hnRNP K) [129]. In this case pathology occurs through activation of apoptosis due to hnRNP K downregulation, and overexpression of hnRNP K is able to rescue pathology, suggesting a novel pathway not previously identified in RNA-mediated disease [129]. In *Drosophila,* expression of AUUCU repeat RNA led to transcriptional changes in neurons that were also observed with CUG and CAG repeat RNA expression, as well as AUUCU nuclear RNA foci [38, 130]. Support therefore also exists for pathogenic pathways in SCA10 that are similar to those responsible for the other repeat RNA-mediated diseases.

1.3.6 Spinocerebellar ataxia type 12 (SCA12)

SCA12 is a rare neurodegenerative disease caused by a CAG repeat within the 5' UTR of the *protein phosphatase 2, regulatory subunit B, beta isofom* (*PPP2R2B*) transcript [131]. The repeat does not appear to prevent *PPP2R2B* transcription and therefore it is hypothesised that expanded RNA toxicity may be responsible for SCA12 [131]. Although direct evidence in SCA12 is limited, in *Drosophila* CAG RNA expression alone is toxic, suggesting that the expanded RNA may contribute to SCA12 [132].

1.3.7 RNA-mediated pathology in the polyglutamine diseases

Although limited, these is some evidence in support of a pathogenic role for repeat-containing RNA in the polyglutamine diseases. In a *Drosophila* model of SCA3, expression of the expanded CAG repeat containing transcript with CAA interruptions led to a reduction in toxicity when expressed in the eye [132]. CAA and CAG repeats both encode polyglutamine protein, while only CAG repeats are able to form a stable RNA secondary structure, suggesting that RNA-mediated interactions may play a role in SCA3. However, this has yet to be confirmed in human patients. Similarly, the CAG repeat within the *HTT* transcript, responsible for HD, is able to form a hairpin secondary structure *in vitro* suggesting the possibility of RNA-mediated interactions [133]. Consistent with this, the CAG repeat-containing *HTT* transcript in human

fibroblast cells from HD patients forms nuclear foci, similar to those observed in DM1, that co-localise with MBNL-1 protein. [133]. Further work is required to confirm the contribution of repeat RNA in SCA3 and HD, however these studies provide preliminary evidence for RNA-mediated pathology in the polyglutamine diseases.

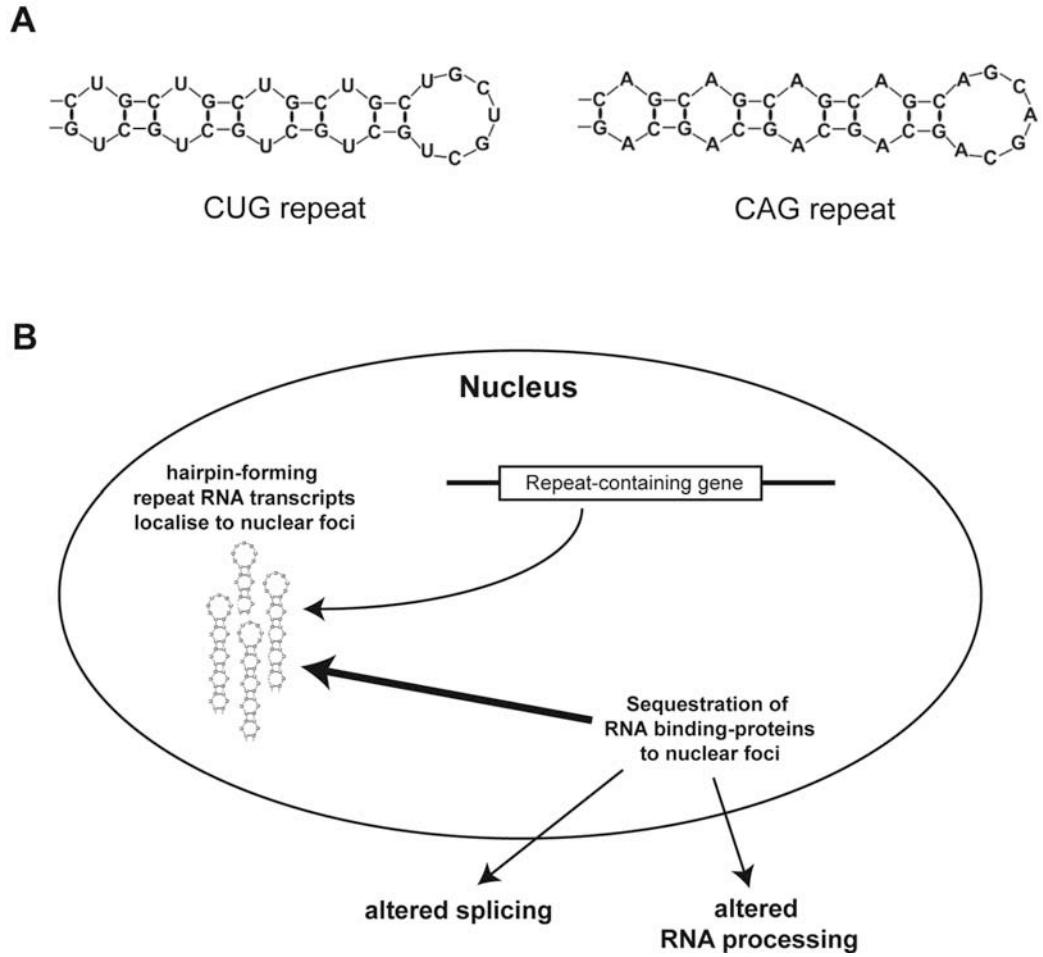
1.4 Pathways of repeat RNA-mediated pathology

1.4.1 Hairpin-forming RNA as a pathogenic agent

One mechanism proposed to account for RNA-mediated pathology involves the ability of repeat transcripts to form specific, stable secondary structures. Structural studies show that all CNG repeats (where N is any nucleotide), as well as the CCUG repeat responsible for DM2 and the SCA10 AUUCU repeat are able to form stable secondary structures [78, 134-138]. In the case of CNG trinucleotides this involves the formation of a hairpin structure as a result of complementary binding of the C and G nucleotides with a mismatch at every second base (Figure 1.2 A).

Hairpin RNA secondary structures have been proposed to cause pathology through the binding and dysregulation of specific RNA-binding proteins. Given the shared ability of repeats to form stable secondary structures, and the presence of a repeat RNA in all diseases, dysregulation of RNA binding-proteins has been proposed as general mechanism contributing to pathology in both translated and untranslated repeat disease. The most extensively characterised example of this involves sequestration of MBNL-1 to nuclear foci, and subsequent dysregulation of alternative splicing (Figure 1.2 B). However different RNA-binding protein preferences caused by the specific structure of each repeat and the influence of repeat transcript context, may lead to both common and unique interactions. Consistent with this, proteins have been identified that show both common and sequence specific RNA-binding interactions [82, 109, 110, 114, 129]. A mechanism based on protein dysregulation by repeat RNA may therefore contribute to the common and distinct clinical features of each of the expanded repeat diseases.

Figure 1.2 Hairpin-forming repeat RNA-mediated sequestration



A, Representation of hairpin structures formed by CUG and CAG repeat RNA. **B**, Schematic of the mechanism by which hairpin repeat RNA is proposed to cause pathology. A repeat containing-gene is transcribed, leading to the production of multiple repeat-containing RNA transcripts that localise to nuclear RNA foci, which are sites of RNA binding-protein sequestration. Sequestration leads to altered splicing and RNA processing, leading to pathology.

1.4.2 Sequestration of MBNL-1 and other proteins

Dysregulation of splicing through sequestration of MBNL-1 was initally shown to contribute to pathology in DM1 and DM2 [84]. Subsequently, sequestration and dysregulation of MBNL-1 and other proteins has been proposed as a candidate pathway for common RNA-mediated pathology. In support of this, the hallmarks of this mechanism, including the presence of RNA foci, co-localisation of MBNL-1 with foci and the mis-splicing of specific transcripts, have been identified in multiple diseases (Figure 1.2 B, Table 1.2). However, some uncertainties exist and further work is required to understand both the specifics of this mechanism, and the manner in which it contributes in each case.

One aspect of the pathways involving MBNL-1 dysregulation in need of further investigation is the correlation between RNA foci formation and pathology. Experiments in cell culture show that expression of either a long CUG, or CAG RNA repeat leads to the formation of foci that sequester MBNL-1, while alternative splicing is only altered by CUG repeat expression [88]. Similarity, a *Drosophila* model showed pathology due to CAG RNA expression, with the formation of foci, but identified no alterations in splicing [132]. In *Drosophila* CUG expression in one case has been shown to lead to nuclear foci that co-localise with muscleblind, with no associated pathology [139]. Conversely, over-expression of the normal length *DMPK* 3'UTR in mouse has been shown to cause mis-regulation of splicing without the formation of foci [140]. These results suggest that there does not always appear to be a clear link between the formation of foci, sequestration of MBNL-1 and pathology.

One explanation for these observations may be that MBNL-1 is only one of a number of proteins that are sequestered by repeat RNA as part of this pathogenic mechanism. A number of other proteins have been identified that may play a role in RNA-mediated pathology, including multiple hnRNPs and other proteins involved in aspects of RNA processing (Table 1.2). Similarly, MBNL-1 itself has been shown to localise in stress granules, and undergo interactions suggestive of a wider role in RNA metabolism [141]. It is therefore possible that MBNL-1 may be only one contributor to a wider pathogenic mechanism involving RNA-mediated dysregulation of multiple aspects of RNA processing, including splicing.

Table 1.2 Evidence for pathways involving sequestration of MBNL-1 as a common contributor to pathology.

Disease	Nuclear foci	Evidence for Mbnl-1 pathway*	Other protein interactions
DM1	Yes	C, S, M	CUG-BP
DM2	Yes	C, S, M	CUG-BP
FXTAS	Aggregates containing transcript	C, S, M#	Pur-alpha, HnRNP A2/B1, Sam68
SCA8	Yes	C, S, M	
HDL-2	Yes	C, S	
SCA10	Yes		HnRNP K
HD	Yes	C	
SCA3	-	M#	

*C = Evidence for co-localisation of MBNL-1/mucleblind with foci ; S = Evidence for RNA-induced alterations in splicing, M = modification of phenotypes by altering the levels of MBNL-1 or othologs.
= modification is the opposite of that expected [110, 132].

Each disease is shown, along with specific evidence in each case for the MBNL-1 sequestration mechanism as a contributor to disease. Evidence includes whether nuclear foci have been observed, and specific evidence implicating MBNL-1 including specific co-localisation of MBNL-1 with foci ; observations of repeat RNA-induced splicing alterations, and modification of phenotypes in animal modes by altering the level of MBNL-1, or orthologs. Other protein interactions are indicated that may not involve the same pathway. Evidence in each case is based on observations in human disease or animal/cell based models. For more details see references, DM1/DM2, [74-76, 85, 92, 99, 142]; FXTAS, [103, 107, 109, 110, 114]; SCA8 [116, 120]; HDL-2, [123]; SCA10, [129, 130]; HD, [133]; SCA3, [132].

Further explanation for the inconsistent link between RNA foci and pathology is that other MBNL-1-independent pathways are important in disease. This is highlighted by the RNA-binding protein CUG-BP, which is not identified in foci, but has been shown to play a role in pathology [94]. Although this role may relate to its opposing regulation of MBNL-1 splicing, some evidence suggests that increased CUG-BP levels in disease may be brought about activation of signalling cascades independent

of MBNL-1 [95]. Similarly, comparison of transcriptional changes in MBNL-1 loss of function mice to those expressing expanded CUG RNA suggests that a number of MBNL-1-independent changes occur in DM1 pathology [143]. Repeat RNA has also been proposed to sequester transcription factors, with a possible link between this and a role for CUG-BP in transcriptional regulation [144, 145]. This mechanism is supported by evidence that specific transcription factors are depleted from active chromatin at a sufficient level to induce specific transcriptional changes [144]. It is not clear at present whether these proposed MBNL-1 independent pathways involve sequestration that occurs without the formation of foci, or an entirely different RNA-mediated mechanism. Understanding the molecular pathways contributing to MBNL-1 dependent and independent pathology will be important in determining the general role for RNA-mediated protein sequestration in expanded repeat disease.

1.4.3 Small RNA processing pathways and bi-directional transcription

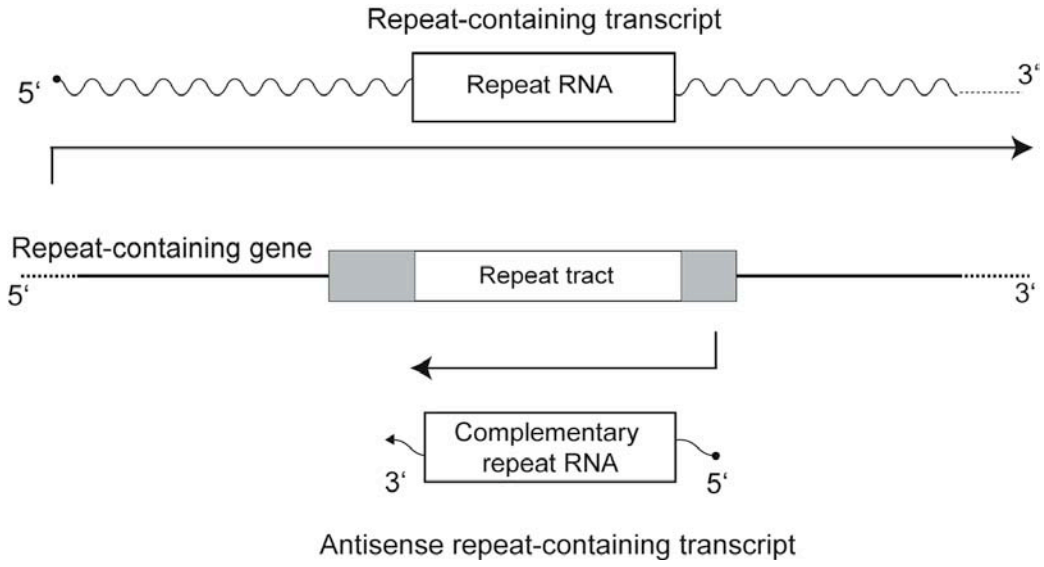
Emerging evidence now suggests that pathways involving small RNA processing may play a role in expanded repeat disease. As for hairpin RNA-mediated pathology this pathway involves the expanded repeat transcript that is present in all cases and could therefore play a common role in both coding and non-coding repeat-mediated disease. Small RNA processing involves a complex pathway of enzymes including the Dicer and Argonaute proteins, leading to the production of a growing class of functional small RNAs that includes short-interfering RNA (siRNA) and micro-RNA (miRNA) [146, 147]. Different precursor double-stranded RNA (dsRNA) structures formed by endogenous transcripts are recognised and processed to form small, usually 21-25nt RNAs [146, 147]. These small RNAs are involved in a growing number of pathways, with a well characterised role in translational inhibition and transcript stability [146, 148].

Biochemical studies have shown that hairpin structures formed by CNG (where N is any nucleotide) RNA repeats can bind and become substrates for the Dicer processing pathway, leading to cleavage of the repeat transcript [149, 150]. In cell culture, CUG repeats are cleaved to form a short dsRNA that is able to enter the siRNA pathway and down-regulate long complementary repeats, although this has yet to be observed in human patients [150].

In addition, increasing evidence now suggests that a number of repeat transcripts are bi-directionally transcribed [3, 151, 152]. This involves the production of a complementary repeat transcript in the opposite direction to the initially observed repeat transcript (Figure 1.3), highlighting the possibility of the complementary repeats hybridising to form double-stranded repeat RNA. Bi-directional transcription was initially observed in a mouse model of SCA8, which is caused by an untranslated CUG repeat, where a short complementary CAG transcript was identified [119]. Evidence also exists for the existence of complementary repeat transcripts in DM1[153], FXTAS [154], SCA7[155] and HDL-2 [124]. A large proportion of the genome is believed to be bi-directionally transcribed [156], and therefore this may be a common property of all expanded repeat transcripts. Antisense transcripts have functional roles in the genome through regulation of transcription, changes to

chromatin and hybridisation to form dsRNA [157]. It is therefore of interest to determine if any of these pathways contribute to repeat RNA-mediated pathology.

Figure 1.3 Bi-directional transcription of repeat-containing genes



Schematic depicting bi-directional transcription from a hypothetical repeat-containing gene. The repeat tract is transcribed in opposite directions from both strands to give two transcripts that each contain complementary repeat RNA sequences.

Bi-directional repeat transcription has been proposed to play a role in disease through the regulation of chromatin modification [153, 158]. In the case of DM1, antisense transcripts have been shown to produce 21nt fragments that direct specific methylation, leading to alterations in transcription levels that are repeat length dependent [153]. Bi-directional transcription is regulated in DM1 by a CTCF-binding element, with evidence that this mechanism may be involved generally at sites of CTG/CAG repeat expansion [155, 159]. Further work is required to determine the mechanism by which altered chromatin dynamics contributes to pathology, and if this pathway is common to all diseases.

Additionally, the formation of double-stranded repeat RNA that becomes a substrate for Dicer processing may contribute to pathology through siRNA and miRNA pathways. In support of a role for small RNA processing in repeat mediated

pathology, modulation of miRNA regulation by genetically altering Dicer levels has been shown to modify toxicity in a *Drosophila* model of SCA3 [160]. Similarly, miRNA dysregulation is observed in DM1 and HD [161-163], suggesting that aspects of small RNA processing may be perturbed. miRNA regulation is important for neuronal function and survival [164] and therefore this pathway provides an interesting candidate for future studies.

In a *Drosphila* model of FXTAS, expression of either CGG repeat RNA, or CCG repeat RNA, led to pathology, however, co-expression suppressed pathology in a manner that was Argonuate-2 dependent [165]. This work suggests that the presence of complementary transcripts may reduce repeat RNA-mediated pathology. However, work presented in this thesis and other results from our lab suggest that bi-directional transcription of complementary CUG and CAG repeat RNA leads to toxicity that is more severe than when either repeat is expressed alone [166]. This is supported by an independent study published during the writing of this thesis that observed a similar effect with complementary CUG and CAG repeat RNA expression [167].

1.5 Repeat RNA as a common contributor to dominant expanded repeat disease.

Given the similar type of mutation, mode of inheritance, pathogenic threshold and clinical features seen in the untranslated and translated dominant expanded repeat diseases, it is hypothesised that a common pathogenic pathway may contribute in all cases. Polyglutamine-mediated pathology has been well characterised and expanded polyglutamine tracts have been shown to have strong intrinsic cytotoxicity. However, the absence of polyglutamine protein in diseases caused by non-coding repeats has limited the evidence for this mechanism to account for the common features of disease. Nonetheless, it may be that common pathology is explained by an as yet unidentified polyglutamine being expressed in some cases where the repeat was considered untranslated. Emerging evidence suggests that bi-directional transcription may produce a polyglutamine tract in HDL-2 and SCA8, both of which are associated with an untranslated CUG repeat [119, 124]. Furthermore, recent novel findings suggest that in some cases repeats within non-coding regions may be translated through an ATG-independent translation mechanism [168]. Further studies will be required to determine if this mechanism has a role in disease, particularly in cases that cannot otherwise encode polyglutamine.

However, even if protein-mediated pathology is identified in the untranslated repeat diseases this does not discount a potential role for repeat RNA as a contributor to pathology in some or all diseases. Expanded repeat RNA transcripts are present in all cases, regardless of whether the repeat is translated or not and in cases that do not involve trinucleotide repeats. Repeat RNA is thus a potential mediator of common pathology. Furthermore, the increasing evidence for repeat RNA-mediated pathology in at least seven diseases highlights the importance of examining the contribution made by repeat RNA in these cases at least.

Central to examining the role of repeat RNA as a common pathogenic agent is determining whether CAG repeat RNA contributes to pathology in cases where it encodes polyglutamine. CAG repeat RNA is able to form hairpin secondary structures and evidence suggests that expanded CAG RNA is able to bind specific RNA binding proteins, including MBNL-1, such that sequestration may be possible [83, 134, 136, 169]. Further evidence is provided by the existence of SCA12, which

appears to be caused by an untranslated CAG repeat tract [123].

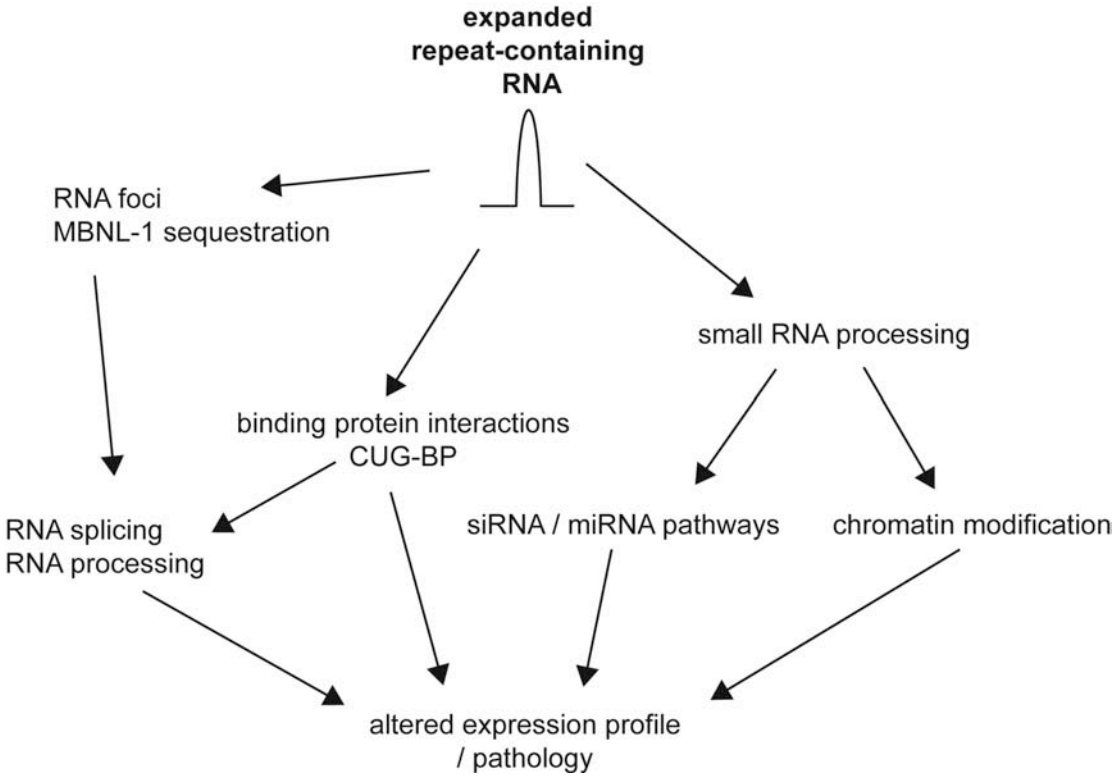
Previous work in our group has examined the contribution of RNA secondary structure to pathology in *Drosophila* by comparing phenotypes caused by ectopic expression of translated CAG and CAA repeats in the eye [37]. While both repeats are translated to form polyglutamine, only the CAG is predicted to form a hairpin secondary structure that may bind specific proteins. In this case, expression of either repeat leads to strong pathology, suggesting that polyglutamine makes a greater contribution to pathology than repeat RNA [37]. Furthermore, expression of an untranslated CAG repeat had no detectable effect [37].

However, similar experiments in an independent *Drosophila* model for SCA3 showed that translated pure CAG repeats, in the context of the repeat containing gene *ataxin-3*, were more toxic than when the repeat tract was interrupted with CAA repeats, suggesting that hairpin RNA contributes to pathology in this case [132]. Further work showed that expression of untranslated CAG RNA alone was able to induce neurodegeneration in this *Drosophila* model, suggesting that CAG RNA is able to contribute to pathology [132]. Following from this, studies have identified CAG RNA-mediated toxicity in *C. elegans* and mouse, where hallmarks of RNA-mediated pathology such as the formation of RNA foci were observed in each case [170, 171]. Additionally, fibroblasts from HD patients show RNA foci that co-localise with MBNL-1, a hallmark of RNA-mediated pathology [133], supporting a role for CAG repeat RNA in polyglutamine disease pathology.

Further identification of particular pathways and biomarkers involved in RNA-mediated pathology will help to identify a contribution by this mechanism to each disease. Currently multiple mechanisms have been proposed to account for RNA-mediated pathology (Figure 1.4). CUG RNA-mediated dysregulation of splicing has been well characterised, however, animal models suggest that CAG RNA may not induce splicing changes, and pathology may occur via an alternative mechanism [132, 170, 171]. Likewise, CGG RNA pathology characterised in FXTAS appears to involve aspects in common with, as well as distinct from, those observed in CUG RNA pathology. It may be that repeat RNA contributes to pathology through a number of pathways, involving mechanisms that are common and others that are

repeat-sequence or context specific. Furthermore, emerging evidence now highlights interesting candidate pathways involving bi-directional transcription and small RNA processing of repeat RNA. Understanding the pathways involved, and relative contribution of each to pathology, therefore presents a new challenge in understanding the basis for human disease.

Figure 1.4 Proposed pathways of RNA-mediated pathology



Proposed pathways of expanded repeat RNA-mediated pathology. Repeat RNA sequesters RNA binding-proteins including MBNL-1 and others, leading to disregulation of splicing and other aspects of RNA processing. Another pathway involves dysregulation of CUG-BP and other proteins that does not involve the formation of foci. Small RNA processing may also contribute to pathology through chromatin modification, or alterations in siRNA and miRNA pathways. A number of pathways lead to altered expression profiles and pathology.

1.6 *Drosophila* as a model for dominant expanded repeat disease

Investigating the pathways through which expanded repeat RNA contributes to cellular perturbation is complicated by the inability to access human tissue, and the slow onset and neurodegenerative nature of many expanded repeat diseases. Animal models such as mouse and *Drosophila* have therefore been essential in identifying a number of important pathogenic pathways.

The model organism *Drosophila* is small, has a short lifespan and produces large numbers of progeny. The genome has been sequenced and annotated and many transgenic lines and genetic resources are available, including the well developed UAS/GAL4 system for ectopic expression of genes [172]. This enables GAL4-dependent expression of transgenes under the control of an upstream activating sequence (UAS). GAL4 protein, originally from yeast, is ectopically expressed via a cell specific promoter, where it binds to the UAS and activates transcription of the transgene in the same cells (Figure 1.5 A). In this way the same set of transgenic lines can be ectopically expressed in different tissues by using different tissue-specific GAL4 lines (Figure 1.5 A).

Another powerful approach enabled by *Drosophila* is the examination of genetic modifiers [173, 174]. This involves identifying a quantifiable morphological phenotype due to particular genetic alterations such as ectopic expression of a transgene. The biological pathways giving rise to this phenotype can be determined by identifying candidate genes that will modify the phenotype when increased or decreased. In this way novel interacting genes can be identified, and this technique has provided a number of important findings in the study of expanded repeat pathology [38, 43, 87, 103, 110]

In many cases studying expanded repeat pathology is limited by the fact that in post-mortem tissue affected cells may have already degenerated and been lost. Using animal models such as *Drosophila* enables the examination of early changes that underlie pathology in living cells. Similarly, using the UAS-GAL4 system enables expression to be directed only in certain tissues so that cellular perturbation can be examined without lethality. One commonly used tissue is the *Drosophila* eye, as this

contains a complex external pattern such that cellular perturbation leads to disorganisation that can be readily detected [173, 174]. Similarly, the *Drosophila* eye is not essential for survival or reproduction and therefore flies exhibiting eye phenotypes can be propagated.

The features of *Drosophila* make it an ideal system in which to study neurodegenerative disease, and models have been created for SCA3, HD, SCA1 and general polyglutamine induced toxicity as well as FXTAS, SCA8, DM1, SCA10 and intrinsic CUG and CAG RNA mediated pathology [38, 43, 103, 116, 132, 139, 142, 175, 176]. *Drosophila* allows the fast generation of multiple transgenic lines and higher throughput assays, thus providing a useful system to examine and compare the multiple pathways that have been proposed to contribute to expanded repeat disease.

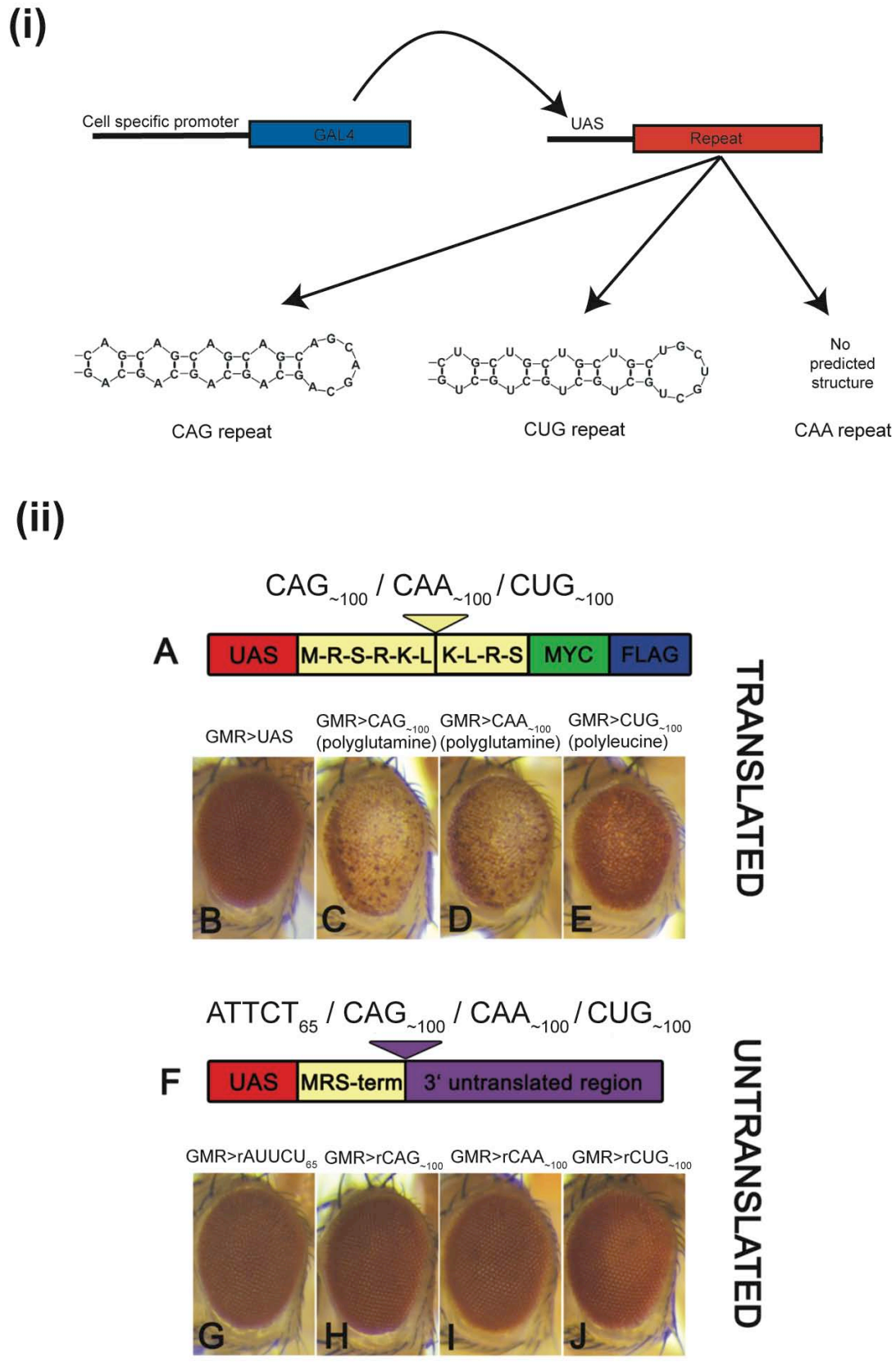
1.6.1 A *Drosophila* model to examine pathways of repeat RNA-mediated pathology.

Studies leading to the work described in this thesis examined a role for hairpin-forming repeat RNA as a common pathogenic agent in expanded repeat disease. Initially this involved comparing translated CAG and CAA toxicity in the *Drosophila* eye, which both encode polyglutamine, however only the CAG repeat is able to form an RNA hairpin structure [37]. In this case, both repeats gave a strong phenotype involving disruption and disorganisation of the eye suggesting that polyglutamine is the primary pathogenic agent, with no detectable contribution due to repeat RNA structure [37]. However, this does not rule out an ability for repeat RNA to have more subtle effects that are not detected in the eye, or are not observed in this assay given the already strong polyglutamine effect. As part of this study a number of constructs were generated to enable GAL4-dependent expression of translated CAA and CAG repeats, as well as untranslated CAG repeats, by inserting a stop codon such that the repeat is now within the 3'UTR of the transcript (Figure 1.5 (ii)) [37]. In this case expression of untranslated CAG RNA had no consistent observable effect on the eye [37].

Following from this initial work, constructs were generated to compare the expression of translated hairpin-forming CAG and CUG repeats and untranslated CAG, CUG, and AUUCU repeat RNA in the same transcript context (Figure 1.5 (ii)) [38]. Expression of polyglutamine from CAG or CAA repeats led to strong phenotypes in the eye, while expression of polyleucine from a translated CUG led to a milder disruption [38]. None of the untranslated repeats led to consistent disruption when expressed in the *Drosophila* eye (Figure 1.5 (ii)), however transcriptional changes were detected compared to controls, including a non-hairpin forming CAA repeat, when expressed in *Drosophila* neurons [38]. In this case both common and distinct changes were identified, suggesting that both shared and sequence-specific pathways are altered by repeat RNA expression [38]. Further examination of candidates identified specific genetic interactions in the eye, and common repeat-RNA mediated perturbation of Akt/Gsk3-β signalling, further supporting an ability for repeats to induce cellular changes in this *Drosophila* system [38]. It may therefore be possible to observe specific phenotypes when expanded repeat RNA is expressed in *Drosophila* under different ectopic expression conditions via the UAS/GAL4 system. As this model enables the direct comparison of different RNA repeat sequences, identifying common phenotypes would enable investigation of the pathways leading to cellular perturbation. This approach forms the basis of work described in Chapter 3 and 4 of this thesis.

Figure 1.5 : (i) The GAL4/UAS system as used to express different repeat RNA sequences in *Drosophila*. A transgenic line carrying the GAL4 gene under control of a cell specific promoter is crossed to a line carrying either repeat under control of a UAS sequence. Progeny carrying both transgenes express the repeat in the same cells that express GAL4, leading to ectopic expression of the repeat in a GAL4-dependent manner. **(ii)** Model system developed in our lab to examine expanded repeat expression, adapted from [38]. **A**, UAS contruct to express different translated repeat sequences, consisting of the repeat flanked by a short peptide sequence on either side, as well as MYC and FLAG tags for detection. **B-E**, Phenotypes when each translated repeat is expressed ectopically in the eye with GMR-GAL4. **B**, UAS control. **C**, Expression of a translated CAG repeat to produce hairpin-forming RNA and polyglutamine protein causes degeneration and loss of pigment in the eye. **D**, Expression of a translated CAA repeat, not forming a stable RNA stucture, but producing polyglutamine protein causes similar degeneration and loss of pigment in the eye. **E,** Expression of a translated CUG repeat to produce hairpin-forming RNA and polyleucine protein gives a milder eye phenotype. **F,** Expression construct used to express untranslated repeat sequences. Construct is as for **A** but with a stop codon introduced so that the repeat is now within the 3'UTR. **G-J**, Expression of **G**, AUUCU, **H**, CAG, **I**, CAA or **J**, CUG repeat RNA does not lead to detectable disruption to the external surface of the eye.

Figure 1.5 A system to examine repeat RNA pathology in Drosophila



Previous studies did not observe consistent repeat RNA-induced phenotypes in the eye except in the case of a single line which carries two insertions of the CAG RNA expression transgene. This line gave a mild phenotype when ectopically expressed in the eye, and locomotion defects when ectopically expressed in neurons [177, 178]. Phenotypes are not due to transgene expression levels, as expression of four transgene copies of each repeat gives no observable phenotype [38, 166]. Further studies were therefore undertaken to examine whether these phenotypes are due to an insertion mutation associated with generation of transgenic lines, or a specific repeat-mediated effect. The investigation and subsequent characterisation of repeat-mediated pathways in this line is described in Chapter 5 and 6 of this thesis.

CHAPTER 2 : Materials and Methods

2.1 Materials

Enzymes

BigDye® terminator mix (Applied Biosystems)

DNase I (Invitrogen)

Restriction endonucleases (New England Biolabs)

RNase H (Invitrogen)

RNase out™ RNase inhibitor (Invitrogen)

Superscript® III reverse transcriptase (Invitrogen)

SYBR Green® PCR master mix (Applied Biosystems)

T4 DNA ligase (New England Biolabs)

Taq DNA polymerase (Invitrogen)

Kits

QIAquick® PCR clean up kit (Qiagen)

RNeasy® mini kit (Qiagen)

Expand Long Template® PCR kit (Roche)

Molecular weight markers

All DNA gels use 1kb+ DNA ladder (Invitrogen)

Oligonucleotides

All oligonucleotides were synthesised by Geneworks (Adelaide, South Australia), and were cleaned up using the companies 'Sequencing/PCR grade' option. Sequences are listed 5' to 3'.

<u>Primers for inverse PCR</u>

5' P element, Fwd :CGCACACAACCTTTCCTC

5' P element, Rev: ATGAACCACTCGGAACCA

3' P element, Fwd: CCTTAGCATGTCCGTGG

3' P element, Rev: AGTGGATGTCTCTTGCCG

Primers for RT-PCR analysis of *rCAG*$_{\sim 100}$ *[line C]*

(informal lab name in brackets)

(i) Fwd: CCTGAAGACCATCGATCGG (ex2F)

(i) Rev: CAGTCCACAATGTCCGACG (ex3R)

(ii) Fwd: ATCCGAATGCCGGTGATC (Fil1F)

(ii) Rev: ACAGGACCTAACGCACAG (Fil1R)

(iii) Fwd: TGAGAGTCAGCAGTAGCC (SV40-R1)

(iii) Rev: CAGCGTTGTTACTCGAGC (5UTR-R2)

(iv) Fwd: CGGCGACAAACTTCAACG (5UTR-F2)

(iv) Rev: CAGTCCACAATGTCCGACG (ex3R)

Primers for real time PCR analysis of *cheerio* transcripts

rp49 Fwd :GACGCTTCAAGGGACAGTATCTG

rp49 Rev : AAACGCGGTTCTGCATGAG

cheerio Fwd: AACATTCCGCAGCCAGAAACT

cheerio Rev: AGTCCACAATGTCCGACGAGTC

Fluorescent probes for *in situ* hybridisation experiments

Cy3-CTG_{10} : Cy3-CTGCTGCTGCTGCTGCTGCTGCTGCTG

Cy3-CAG_{10} : Cy3-CAGCAGCAGCAGCAGCAGCAGCAGCAG

Cy3-TTG_{10} : Cy3-TTGTTGTTGTTGTTGTTGTTGTTG

Cy3-GFP : Cy3-CCTTCACCCTCTCCACTGACAGAAAATTTGTGCCC

***Drosophila* media**

11% sucrose, 5.2% polenta, 5% bakers yeast, 5% treacle, 1.8% tegosept (10% methyl-parahydroxybenzoate) and 0.9% J-grade agar in reverse osmosis purified H_2O.

Buffers and solutions

1xSSC : 150mM NaCl, 15mM sodium citrate

In situ hybridisation buffer : 4xSSC, 0.2g/mL dextran sulphate, 50% formamide, 0.25 mg/mL poly(A) RNA, 0.25mg/mL single stranded DNA, 0.1M DTT, 0.5x Denhardt's reagent.

PBS : 7.5mM Na_2HPO_4, 2.5mM NaH_2PO_4, 145 mM NaCl.

***Drosophila* stocks**

Stocks were obtained from the Bloomington Drosophila Stock Centre (Bloomington, IN, USA) unless otherwise noted. Nomenclature used for *Drosophila* genotypes is as used on Flybase (www.flybase.org).

Act5c-GAL4 (Bloomington stock #4414), originally described in Ito et al 1997 [179]. Drives expression ubiquitously.

da-GAL4 (#8641), originally described in Wodarz et al 1995 [180]. Drives expression ubiquitously.

eip71CD-GAL4 (#6871), originally described in Cherbas et al 2003. Drives expression in larval epithelial cells [181, 182]

elav^{C155}*-GAL4* (#458), originally described by Lin and Goodman 1994 [183]. Drives expression pan-neuronally.

GMR-GAL4 (#1104), originally described by Freeman 1996 [184]. Drives expression in all cells of the developing and adult eye.

T155-GAL4 (#5076), originally described by Harrison et al 1995 [185]. Drives expression in the histoblasts [181, 186].

dcr1^{Q1147X} and *dcr2*^{L811fsX} were obtained from Prof. Richard Carthew and are described in Lee et al 2004 [187].

mbl^{E27} (#7318) is caused by imprecise excision of a p-element, removing exon 1 and 2 [80].

Df(3R)Exel6176 (#7655) deficiency removes 89E11-89F1 including *cheerio.*

EP(3)3715 (FBti0011794, Szeged stock centre) consists of an EP element within the *cheerio* gene allowing UAS driven overexpression.

Repeat expression lines

Repeat constructs and exact repeat copy numbers are described in McLeod et al [37] and van Eyk et al [38]. In this thesis the abbreviated nomenclature 'rCUG$_{\sim 100}$' is used, as in Lawlor et al [166]. Four copy repeat lines were generated by L. O'Keefe and S. Samaraweera. Generation of rCAG$_{\sim 100}$ and rCUG$_{\sim 100}$ is described in Lawlor et al [166] (Appendix 4), and identical methods were used to generate *4xrCAA$_{\sim 100}$*. Complementary repeat expression *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* lines used in Chapter 6 were generated in the same manner by S. Samaraweera.

Lines used (including original working alphabetical nomenclature given to lines)
rCAG$_{\sim 100}$ [line 1] (line A, E; G, I)
rCAG$_{\sim 100}$ [line 1a] (line A, E)
rCAG$_{\sim 100}$ [line 1b] (line G, I)
rCAG$_{\sim 100}$ [line 2] (line J, K; D, H)
rCAG$_{\sim 100}$ [line 3] (line A, B; G, H)

rCUG$_{\sim 100}$ [line 1] (line C, D; E, F)
rCUG$_{\sim 100}$ [line 2] (line H, I; J, G)

rCAA$_{\sim 100}$ [line 1] (line C, I; E, F)
rCAA$_{\sim 100}$ [line 2] (line A, H; B, G)

Complementary lines used in Chapter 6:
rCAG$_{\sim 100}$ [line G].rCUG$_{\sim 100}$ [line D]
rCAG$_{\sim 100}$ [line E].rCUG$_{\sim 100}$ [line J]
rCAG$_{\sim 100}$ [line D].rCUG$_{\sim 100}$ [line H]

4xUAS (attB insertion sites : 22A, 58A ; 68E, 96E)

Repeat-GFP lines described in Chapter 4 were generated by S.Samaraweera in a similar manner to those above, using the attB system [188] to generate transgenics. Insertion sites were as for 4xUAS above. In this case the repeat was inserted within the 5'UTR of GFP so that the repeat tract was not translated.

2.2 Methods

Drosophila husbandry

Unless otherwise noted flies were raised at 25°C in standard vials (Genesee Scientific), with a 12h light/dark cycle.

DNA manipulation

Restriction enzyme digestion, agarose gel electrophoresis and other standard techniques were performed as in Sambrook and Russell 2001 [189] unless otherwise noted.

Quantification of tergite disruption

Female adults only, to avoid morphological differences between genders, were scored for tergite disruption by examining the dorsal abdomen under a standard dissecting microscope. The phenotype was categorised on the scale: 1, like wild-type ; 2, tergites mildly disrupted but not split ; 3, at least one tergite split ; 4, two or more tergites split. Counts from multiple crosses scored under identical conditions were pooled to give a final tally for each genotype. To compare the effect of modifiers on the tergite phenotype categories were pooled into 2 groups, those in category one or two, and those in category three or four. This represents those with a mild, or no phenotype, and those with a strong phenotype, and appeared to be the most robust way to determine if modification was significant. Statistical significance was determined using Fisher's exact test (Graphpad Prism). This enabled direct comparison of different sized populations, and determined the probability that the distribution of progeny within categories differ between genotypes by chance alone. $p=0.05$ was used as a cut off for significance.

In situ hybridisation of frozen sections

Whole wandering third instar larvae, or dissected 0-24 hour adult heads were placed in tissue molds, covered with Tissue-Tek® O.C.T.™ freezing medium, frozen using dry ice and stored at -80ºC till cutting. Sections were cut using a Leica CM1900 cryostat, with chuck and chamber temperature between -16ºC and -19ºC, adjusted as

necessary to ensure good quality sections. 10μm sections were cut and placed on poly-lysine coated slides, then either immediately fixed for hybridisation, or stored at -80ºC. For each genotype multiple larvae were frozen per mould such that each section contained multiple animals.

Prior to hybridisation sections were fixed 15 minutes in ice cold 4% paraformaldehyde in PBS, washed 3 x 15 minutes in room temperature PBS and quickly rinsed in 100% ethanol. Slides were dried and hybridised for 2 hours, or overnight at 37ºC in a humid chamber with 0.5ng/ul fluorescent oligonucleotide probe in hybridisation buffer. Slides were washed 2x15 minutes in 2xSSC, 2x15 minutes in 0.5xSSC at 37ºC, air-dried and mounted in Vectashield™ (Vector labs) with 1ng/μL DAPI to visualise nuclei. Multiple nuclei from multiple sections, each containing multiple animals ($n \approx 10$) were examined per genotype.

Microscopy

Light microscopy of external eye structure, and adult abdomen was performed on an Olympus SZX7 microscope fitted with a SZX-AS aperture. Images were captured using an Olympus ColourView IIIU Soft Imaging System camera and AnalysisRuler image acquisition software. Adobe Photoshop CS was used for image preparation.

Fluorescent microscopy for *in situ* hybridisation experiments was performed using a Zeiss Axioplan 2 microscope with 63x PlanApo objective. Images were captured using Axiovision 4.5 software with an Axiocam MRm camera. Further preparation of images was done using Axiovision, or Adobe Photoshop CS.

Quantification of locomotion phenotype

0-3day old flies raised at 29ºC were separated into individual vials then allowed to recover for 30 min. Only males were examined as these showed a more obvious phenotype. Vials were placed at 37ºC in a humid box for approximately 60 minutes before each vial was knocked and filmed for 60 seconds using a standard home video camera. Films were reviewed in Apple iMovie to manually score the time that each fly spent on its back in the 60 seconds after the vial was knocked.

DNA extraction

DNA was extracted based on the method of Huang et al, Protocol 23 in *Drosophila Protocols*, ed. Sullivan et al [190]. Briefly 30 whole flies were ground with 400µL of 100mM Tris-Cl (pH 7.5), 100mM EDTA, 100mM NaCl, 0.5% SDS, incubated 30 minutes at 65°C, 800µL 2 : 5 mix of 5M potassium acetate : 6M lithium chloride was added, mixed and incubated on ice for 10 minutes minimum. Tubes were centrifuged at 12000 rpm for 15 minutes at room temperature. 1mL of supernatant was transferred and mixed with 600µL isopropanol then centrifuged at 12 000 rpm, 15minutes at room temperature. The supernatant was discarded and the pellet washed with 70% ethanol, air dried and re-suspended in 150µL MQ H_2O.

Inverse PCR

Inverse PCR was performed using the method of Huang et al, Protocol 23 in *Drosophila Protocols*, ed. Sullivan et al [190]. Enzyme *Msp1* (New England Biolabs) was used to digest DNA, and ligations were performed using T4 DNA Ligase and included buffer (New England Biolabs). PCR conditions were 94ºC for 3 minutes, 35 cycles of 94ºC for 30 seconds, 62ºC for 30 seconds, 68ºC for 2 minutes, then 72ºC for 10 minutes. Products were sequenced and BLAST (NCBI) used to identify the genomic site of insertion.

DNA Sequencing

DNA was sequenced using ABI Prism™ Big Dye Terminator v3.1 Cycle Sequencing Ready Reaction Mix (Perkins-Elmer) as per the manufacturer's protocol, except that half the amount of reaction mix was used. 20µL reactions were performed with approximately 400-800ng of DNA template and 100ng of primer on a MJ Research PTC-200 peltier thermal cycler. Conditions were 25 cycles of : 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes.

Cleanup was performed by : precipitation with 80µL 75% isopropanol for 15 minutes at room temperature ; centrifugation at 13,000 rpm in a standard desktop centrifuge for 20 minutes ; removal of supernatant ; washing with 250µL 75% isopropanol ; 10 minutes centrifugation under the same conditions ; removal of supernatant and air drying of pellet. Sequencing was performed by the Institute of Medical and Veterinary Science (IMVS), Adelaide.

RNA purification

Samples were frozen in liquid nitrogen, 1ml of Trizol (Invitrogen) added and homogenised by passing through a 19 guage, then 26 guage needle. Homogenate was centrifuged at 13 000 rpm for 10 minutes at 4°C and supernatant decanted. A further 500µl Trizol was added and incubated 15 minutes at room temperature. 300µL chloroform was added and vortexed for 1 minute, then centrifuged for 15 minutes at 13 000 rpm. The upper aqueous phase was collected and mixed with an equal volume of 100% ethanol, vortexed, and 700µL loaded onto an RNeasy column (Qiagen) and purification carried out as per the manufacturers instructions.

Reverse transcription to produce cDNA

Reverse transcription was performed using the Superscript III kit (Invitrogen) following manufacturers directions. Random hexamers were used at a concentration of 250ng/µL.

Quantitative real-time PCR (qRT-PCR)

RNA was reverse transcribed using Superscript III (Invitrogen) and random hexamers. Power SYBR® Green master mix (Applied Biosystems) was used to carry out qRT-PCR in triplicate using an Applied Biosystems ABI Prism 7000 Sequence Detection System (Applied Biosystems) and quantified using the Relative Standard Curve Method as described by the manufacturer. Expression levels were normalised to the house-keeping gene *ribosomal protein 49* (*rp49*). Experiments used three independent RNA preparations for each genotype. Data was exported to Microsoft Excel for statistical analysis.

***cheerio* RT-PCR**

RNA from five whole one-day old adult male flies was purified, DNase I (Invitrogen) treated to remove any contaminating genomic DNA and reverse transcribed. cDNA was used as template with primers described earlier, and in Chapter 5 (Figure 5.4). Non-repeat spanning reactions were performed using Taq polymerase (Invitrogen) following manufacturers directions. Conditions were 94ºC for 3 minutes, 35 cycles of 94ºC for 30 seconds, 55ºC for 30 seconds, 72ºC for 2 minutes then 72ºC for 10

minutes. Repeat spanning transcripts were amplified with the Expand Long Template kit (Roche) following manufacturers directions. 'Buffer 1' was used with conditions : 94ºC for 2 minutes, 10 cycles of 94ºC for 45 seconds, 60ºC for 45 seconds (reduced by 1ºC per cycle), 68ºC for 5 minutes ; then, 25 cycles of 94ºC for 45 seconds, 50ºC for 45 seconds, 68ºC for 5 minutes (increased by 20 seconds per cycle) ; finally 68ºC for 7 minutes. Products were cleaned up and sequenced, and analysis performed using Geneious software (Biomatters, NZ).

Eye sections and counting

Eyes were aged for at least 35 days, or 0 days, and dissected by removing the head and cutting it in half so that each eye was attached to a separate piece of tissue. Tissue was fixed overnight in cold 2.5% glutaraldehyde, 0.1M sodium phosphate buffer (pH 7.2), post fixed in 2% osmium tetroxide, washed in water, and dehydrated in acetone. Tissue was mounted in epoxy resin, sectioned at 1 μm using an RMC Mt7 ultramicrotome, mounted on slides and stained with toluidine blue. Post fixation processing, and sectioning was performed by H. Irving-Rodgers (Uni. of Adelaide, School of Health Sciences).

Photographs were taken at 100x using a Zeiss Axioplan 2 microscope and adjusted using Axiovision LE 4.8.1 to ensure rhabdomeres were clearly visible. Ommatidia within each image were scored for the number of rhabdomeres present, with a possible range of 1 to 8. Multiple images, each representing individual animals, were analysed per genotype with each image containing at least 20 ommatidia. For each genotype a mean of the frequency of each category across all animals was calculated. Values for each category were then compared to the corresponding control value using a Student's t-test in GraphPad Prism 5.0a.

Survival assays

Adult male flies were collected within 5 days post-eclosion and separated into populations no greater then 20 per vial. Flies were transferred to fresh food every 2 days at 25ºC and any deaths or escapes recorded. Survival analysis and log-rank tests were performed using Graphpad Prism 5.0a for Mac OS X (GraphPad Software, San Diego California USA). Escaping flies were entered as censored data.

Climbing assays

Flies were collected and maintained as for survival assays except that a culture temperature of 29°C was used. Each genotype was separated into multiple sets with at most 25 animals per population. Climbing assays were performed by transferring flies to a 500ml measuring cylinder (diameter 48mm) sealed at the top with Parafilm. Multiple identical cylinders were used so that all genotypes could be tested in the same session. Tests were conducted by gently tapping the cylinder on the bench-top to knock flies to the bottom and scoring the number of flies that remained below the 50ml mark (height 27mm) after 25 seconds. Five consecutive trials were conducted per set of flies, with an approximate 3 minute rest between trials. Trials were averaged to get a representative proportion for that population. Proportions were then averaged across all populations (n = 3) to determine the mean for each genotype. Genotypes were compared to the control using a 2-tailed Student's t-test in GraphPad Prism with a cut off of p=0.05 for determining significance.

Electroretinograms

Flies were briefly anaesthetised with CO_2 and mounted on a glass cover-slip using President (Coltene) dental wax, so that one eye was facing upwards. Flies were allowed to recover for 30 minutes before testing. Readings were taken using a pulled glass electrode filled with 2M potassium chloride (resistance approximately 10-20 Mega Ohms) and inserted just within the surface of the eye, amplified using a BA-01X amplifier (NPI electronic). Photoreceptors were stimulated using a white LED placed 1 cm from the eye, and controlled by a custom build LED driver. Stimulus signalling and data capture were performed with a National Instruments USB-6211 interface, and custom Labview (National Instruments) program written by R. Brinkworth (Uni. of Adelaide, School of Mechanical Engineering). Multiple readings (n=20) were taken to get a mean value for each fly, and this was used to obtain a mean for multiple flies of the same genotype. Analysis was performed in Microsoft Excel and GraphPad Prism.

Summary of results

Chapter 3

The hypothesis that hairpin-forming CUG and CAG repeat RNA contributes to pathology through a common pathway is examined. Experiments aim to identify common phenotypes caused by repeat RNA expression in *Drosophila*. Ubiquitous expression of CUG or CAG repeat RNA leads to a reduction in viability and disruption to the patterning of the abdominal tergites that is dependent on the perturbation of specific cells. Expression of either repeat leads to common effects suggesting a shared pathway of perturbation. Attempts are made to examine the pathways involved by determining whether reducing *muscleblind* is able to modify the phenotype. Results do not support that muscleblind sequestration is a major rate-limiting contributor to the phenotype.

Chapter 4

The contribution of a mechanism of repeat RNA-mediated sequestration to the tergite phenotype is further examined by determining whether nuclear RNA foci are formed, and if this correlates with the phenotype. CUG, but not CAG or CAA repeat RNA forms specific nuclear foci in muscle cells only. In other cells all repeats localise to sites of RNA concentration within the nucleus that are not repeat sequence dependent. Overall no correlation is found between specific nuclear RNA localisation and tergite disruption, providing further support that this phenotype is not caused by a mechanism of sequestration of muscleblind to nuclear foci.

Chapter 5

The basis for dominant phenotypes in a particular CAG expressing transgenic line is examined. Results suggest that phenotypes are dependent on expression of CAG RNA from a specific insertion and the cells in which expression occurs and are not caused by loss of function of the inserted gene. Further experiments reveal that phenotypes are dependent on bi-directional transcription of the repeat from both the UAS sequence and an endogenous promoter. Results support a novel pathway of cellular perturbation involving the formation of double-stranded repeat RNA due to bi-directional repeat transcription.

Çhapter 6

The pathways responsible for different repeat RNA-mediated phenotypes are examined. Phenotypes caused by bi-directional repeat expression from a single locus and complementary repeat expression from different loci are compared. Further experiments examine whether pathways involving Dicer processing, which are able to modify double-stranded repeat RNA-mediated pathology, are also responsible for hairpin-forming repeat RNA phenotypes. Results support that distinct pathways are responsible in each case.

CHAPTER 3 : Specific cellular perturbation due to ubiquitous expression of expanded repeat RNA in *Drosophila*

Introduction

Growing evidence supports a role for repeat RNA-mediated pathology in human dominant expanded repeat disease. Understanding the extent of this role, and the pathways responsible for pathology will be important in determining the basis for disease. Multiple studies highlight the relevance of *Drosophila* as a useful model organism in which to study expanded repeat RNA-mediated pathology [37, 38, 103, 110, 132, 139, 142]. By identifying repeat RNA-induced phenotypes in *Drosophila* it will be possible to examine potential genetic modifiers of cellular perturbation by quantifying phenotypic changes when specific candidate genes are altered. In this way the *Drosophila* system can be used to identify candidate pathways that will inform further studies in human patients.

Previous *Drosophila* studies indicate that CUG repeat RNA is able to cause pathology when ectopically expressed in the eye and muscles, while an independent model shows that CAG repeat RNA is able to cause pathology in neurons [87, 132]. In each case different ectopic expression conditions and constructs were used, and it is unclear if common pathways are involved in pathology [87, 132]. Experiments in this chapter use the *Drosophila* model previously established in our group to examine the hypothesis that disease associated CAG and CUG trinucleotide repeat RNA contributes to pathology through common pathways. Initially this involved examining whether the expression of different trinucleotide repeat RNA sequences is able to induce common phenotypes in *Drosophila*.

Constructs were available to ectopically express each of the hairpin-forming CUG or CAG repeats, or a non-hairpin forming CAA repeat [37, 38]. Common transcriptional changes were previously identified when CUG or CAG repeats were ectopically expressed in neurons, however no significant morphological phenotype was observed

with neuronal expression, or ectopic expression in the eye [38, 166]. Therefore initial experiments described in this chapter aimed to identify quantitative morphological phenotypes in *Drosophila* by expressing constructs in different tissues using the GAL4 / UAS ectopic expression system.

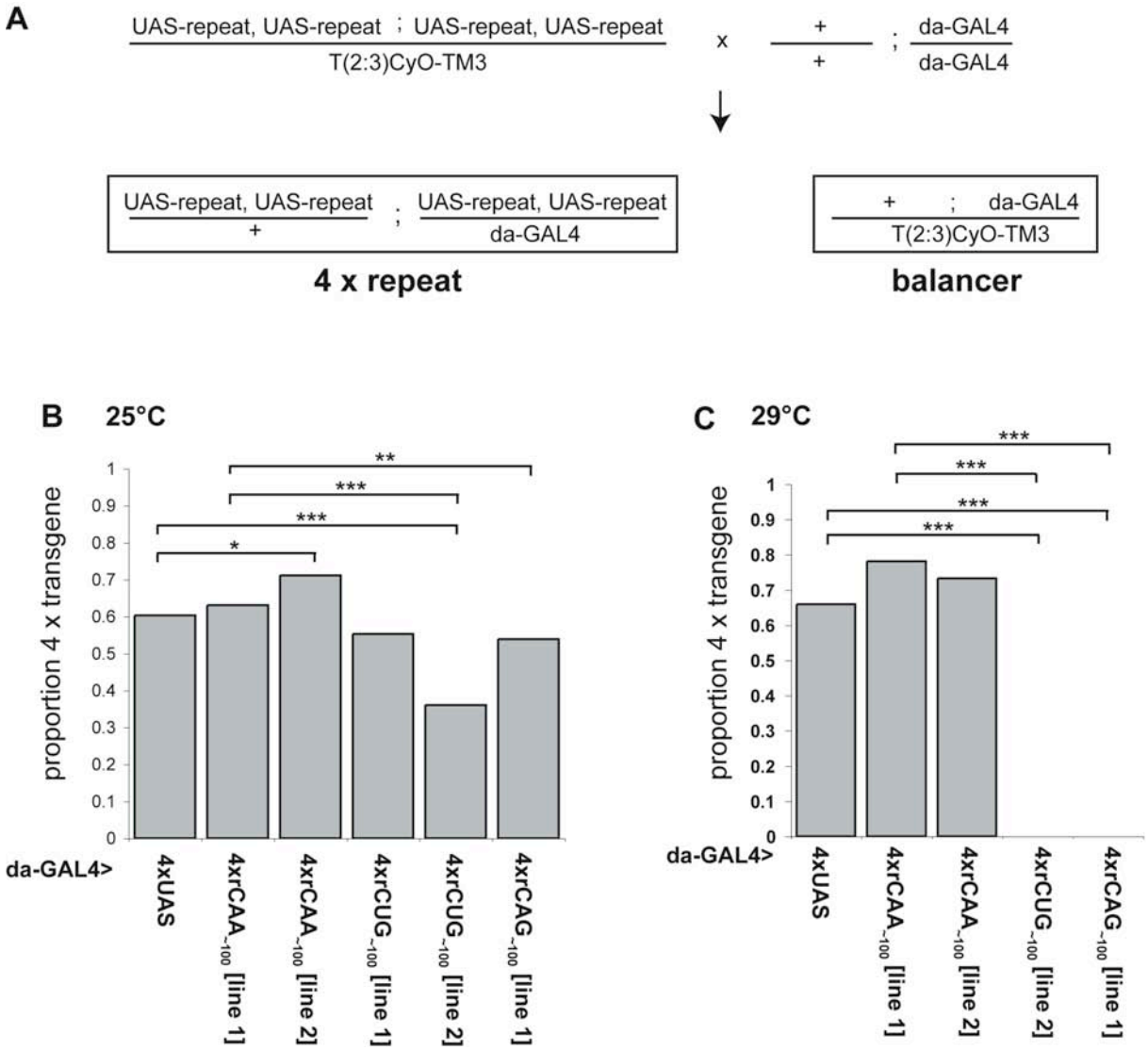
3.1 Ubiquitous expression of CUG or CAG repeat RNA causes reduced viability in *Drosophila*

Ectopic expression in the eye, or neurons does not lead to repeat RNA phenotypes in this *Drosophila* model and therefore in this study repeats were initially expressed ubiquitously under the control of the *da-GAL4* driver [180]. It was hypothesised that ubiquitous expression would give a greater chance of observing effects by perturbing multiple cell types. Initially viability was measured to determine if repeat RNA expression perturbs one or more pathways necessary for survival to adulthood. To maximise repeat expression levels, lines carrying four copies of a *rCUG*$_{\sim 100}$ (‘*4xrCUG*$_{\sim 100}$’) or *rCAG*$_{\sim 100}$ (‘*4xrCAG*$_{\sim 100}$’) transgene were used [38, 166]. Comparisons were made to flies carrying four copies of an *rCAA*$_{\sim 100}$ (‘*4xrCAA*$_{\sim 100}$’), construct that is not predicted to form a hairpin structure and to flies carrying four copies of a UAS-empty vector control construct (4xUAS). The UAS-empty vector construct was created from the same vector used to create repeat lines without any insert and since UAS sites are still present, provides a control for titration of free GAL4 that occurs when repeats constructs are expressed [166].

Viability was determined by comparing the number of adult flies expressing four copies of the transgene as a proportion of the total populaton, including siblings that inherited the balancer chromosome only. In the case that repeat expression is completely lethal, only balancer progeny will be obtained. Lines were created such that the four copies were present in sets of two copies, with each recombinant pair on a separate chromosome (Materials and Methods, Appendix 4). To avoid analysis being complicated by the presence of progeny carrying only two copies, four copy lines were balanced over a compound 2nd/3rd chromosome balancer (T(2:3)CyO-TM3, hsp-GAL4, UAS-GFP) [191]. Using this approach progeny inherit either four copies of the transgene, or the compound balancer, enabling less complex statistical analysis (Figure 3.1 A). For each experiment, all progeny were scored as either having the balancer, or expressing four repeat copies. Multiple crosses set up under identical conditions were examined, and data was tallied to get a total for each group per genotype (Appendix 1). For each genotype, the proportion of four transgene expressing progeny was determined as well as the 95% confidence interval for that proportion (Figure 3.1 B, Appendix 1). Comparisons were made between genotypes

using Fisher's exact test to determine if the two populations differ significantly in their distribution between groups (Figure 3.1 B, Appendix 1).

Figure 3.1 Ubiquitous expression of hairpin forming repeat RNA leads to reduced viability.



A, Crosses were set up so that progeny inherited the *da-GAL4* driver and either all four copies of the repeat transgene, or the balancer. **B**, Graphs show the proportion of four copy repeat progeny in each population. Statistical significance is indicated based on Fisher's exact test comparing the distribution of progeny for each population (Appendix 1). *<0.05, **<0.01, ***<0.001. **C**, Graphs show the proportion of four copy repeat progeny when raised at the higher temperature of 29°C, in this case no rCUG or rCAG progeny are observed. Significance is as for **B**.

For the control 4xUAS line, four copy progeny gave a proportion of 0.603 (Figure 3.1 B). This was higher than the expected mendelian proportion of 0.5 and may indicate that compound chromosome balancer flies have reduced viability compared to siblings with *da-GAL4* and *4xUAS*. This effect is expected to be constant for all lines and therefore should not effect comparisons. Progeny expressing *4xrCAA$_{\sim 100}$ [line 1]* gave a proportion of 0.631 (n=474), that was not significantly different to *4xUAS* (p=0.431) (Figure 3.1 B). However, *4xrCAA$_{\sim 100}$ [line 2]* progeny gave a proportion of 0.711 (n=201), indicating a significant increase compared to *4xUAS* (p = 0.01) (Figure 3.1 B). It is unclear why expression of the *rCAA$_{\sim 100}$* construct would lead to a higher viability, however these results confirm that expression of a non-hairpin forming RNA repeat does not cause a reduction in adult viability.

To examine the effect of CUG repeat RNA on viability, two independent *rCUG$_{\sim 100}$* four copy lines were crossed to *da-GAL4*. Progeny expressing *4xrCUG$_{\sim 100}$ [line 1]* gave a proportion of 0.553 (n=152), with no significant difference compared to *4xUAS* (p = 0.327), or *4xrCAA$_{\sim 100}$ [line 1]* (p = 0.104) (Figure 3.1 B). However, *rCUG$_{\sim 100}$ [line 2]* viability was reduced, with a proportion of 0.360 (n=247), a significant reduction compared to *4xUAS* (p<0.001) and *4xrCAA$_{\sim 100}$ [line 1]* (p < 0.001) (Figure 3.1 B).

Only a single four copy *rCAG$_{\sim 100}$* line, *4xrCAG$_{\sim 100}$ [line 1]*, was tested as *4xrCAG$_{\sim 100}$ [line 2]* contains two copies of the transgene on the X chromosome, which therefore cannot be balanced by the compound chromosome, preventing analysis in this way. Progeny expressing *4xrCAG$_{\sim 100}$ [line 1]* represented a proportion of 0.539 (n=201), indicating a small reduction in viability. This was not significant compared to *4xUAS* (p = 0.087), however there was a significant reduction compared to *4xrCAA$_{\sim 100}$[line 1]* (p = 0.008) (Figure 3.1 B).

To further examine the reduced viability seen in these lines, experiments were repeated at a higher culture temperature of 29°C, rather than the standard temperature of 25°C. Increased temperature results in increased GAL4 activity and therefore a higher level of repeat expression. Under these conditions similar viabilities were seen with *da-GAL4* expression of *4xUAS* (0.659, n=85), 4x*rCAA$_{\sim 100}$ [line 1]* (0.781, n=73)

and *4xrCAA$_{\sim 100}$ [line 2]* (0.733, n=101), (Figure 3.1 C). However, no repeat expressing progeny were obtained with *rCUG$_{\sim 100}$ [line 2]* (n=17) or *4xrCAG$_{\sim 100}$ [line 1]* (n=45) (Figure 3.1 C) indicating that that expression from these lines gives no viable adult progeny under these conditions.

In this set of experiments ubiquitous expression of one of two *4xrCUG$_{\sim 100}$* lines led to a significant reduction in viability compared to the control and *4xrCAA$_{\sim 100}$* repeat, while 4xrCAG$_{\sim 100}$ led to a reduction in viability that was only significant compared to *4xrCAA$_{\sim 100}$* expression. An increase of expression level however led to complete lethality with both CAG and CUG repeats, indicating that expression of hairpin forming RNA can perturb pathways essential for survival to adulthood in *Drosophila*. High levels of expression appear to be necessary to achieve this, with expression of four transgene copies at a higher growth temperature in this case.

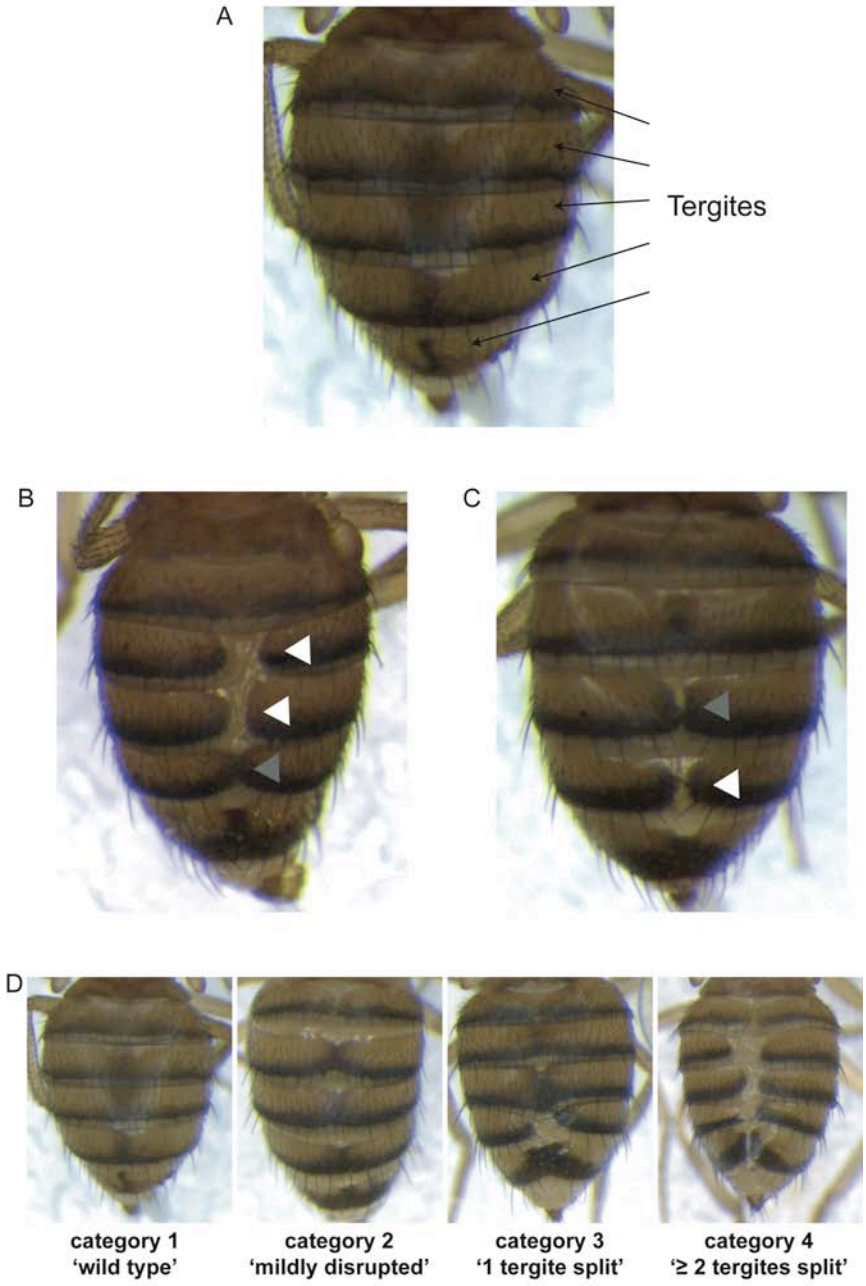
3.2 Ubiquitous expression of CUG or CAG repeat RNA causes disruption to adult *Drosophila* tergite patterning

The observed viability results may be explained by relatively subtle, or specific, repeat RNA effects such that a very high dosage is required to cause lethality. In support of this, examination of *4xrCUG*$_{\sim 100}$ or *4xrCAG*$_{\sim 100}$ adult progeny from these experiments identified a phenotype, involving disruption to the morphology of the adult dorsal abdomen, that was not present in *4xrCAA*$_{\sim 100}$ progeny or *4xUAS* controls. This phenotype may indicate perturbation of a specific pathway or cell type, and was therefore examined in more detail.

In wild-type progeny the abdomen contains a series of regularly arranged bands called tergites (Figure 3.2 A). In repeat expressing progeny one or more of the tergite bands were often split down the midline so that the two sides of the tergite did not meet at all, or met only partially (Figure 3.2 B, C). This phenotype was not observed with *da-GAL4* expression of two independent *4xrCAA*$_{\sim 100}$ lines, or the *4xUAS* control. Experiments were set up to examine and quantify the phenotype in independent transgenic lines expressing repeats ubiquitously via *da-GAL4*. As previously, lines containing four copies of the repeat transgene were used to ensure high expression levels, with two independent lines tested per repeat sequence. Comparisons were made to control progeny expressing *4xUAS* and also progeny with *da-GAL4* alone. To ensure that any phenotypes observed were not due to insertional effects, each line was also crossed to the w^{1118} wild-type line to produce progeny that contain all four transgene copies in the absence of GAL4 driven expression.

To enable comparison between lines, a method was established to categorise the severity of tergite disruption. Individual female flies were scored and placed into one of four categories: category 1, like wild-type; category 2, tergites mildly disrupted but not split; category 3, at least one split tergite; category 4, greater than two tergites split (Figure 3.2 D). Comparisons were then made by determining the proportion of the population within each category (indicated by [1] = proportion in category 1, and so on), (Figure 3.3, for full data set see Appendix 2.1).

Figure 3.2 Tergite disruption is observed in *rCUG~100* and *rCAG~100* expressing flies



A, Wild-type female flies show a regular arrangement of bands on the dorsal abdomen called tergites. **B**, Example of a disrupted abdomen. Some tergites are not fully formed such that they do not meet in the middle and appear split (white arrows), while others have met but are not properly formed (grey arrow). **C**, Shows another example of the disruption phenotype showing one split tergite (white arrow) and one that is only partially split (grey arrow). **D**, Disruption was quantified by scoring individual flies based on the severity of disruption where category 1 was like wild-type; category 2, tergites mildly disrupted but not split; category 3, at least one tergite split ; category 4, two or more tergites split.

Tergite disruption was observed at a negligible frequency in wild-type progeny, (n=401) with almost all progeny scored as category 1 ([1] = 0.998, n=401) (Figure 3.3 A'). Likewise progeny with *da-GAL4* but no repeat construct were almost all like wild-type ([1] = 0.99, n=506), (Figure 3.3 A). A similar result was observed for expression of the *4xUAS* control ([1] = 0.994, n=161), and all progeny with *4xUAS* but no *da-GAL4* driver fell within category 1 ([1]=1.00, n=203), (Figure 3.3 B, B').

When crossed to w^{1118} to generate four copy repeat progeny with no driver, all repeat lines showed no phenotype, with proportion in category 1 greater than 0.99 for all (n>127 for all, Appendix 2.1, Figure 3.3 A'-H'). This indicates that any phenotypes observed are likely caused by repeat expression rather than a transgene insertion effect. Progeny expressing *4xrCAA*$_{\sim 100}$ appeared like controls with, *4xrCAA*$_{\sim 100}$*[line 1]*, [1] = 0.986, (n=158) and *4xrCAA*$_{\sim 100}$*[line 2]*, [1] = 0.987 (n=148), confirming that *rCAA*$_{\sim 100}$ expression does not lead to tergite disruption (Figure 3.3 C, D).

Progeny ubiquitously expressing *4xrCUG*$_{\sim 100}$ *[line 1]* (n=271) showed a moderate phenotype. Category 1 contained few progeny ([1] = 0.07), the majority were scored as category 2 ([2] = 0.782) indicating mild tergite disruption, and a small number as category 3 ([3] = 0.125) and 4 ([4] = 0.022) (Figure 3.3 E). The second *rCUG*$_{\sim 100}$ line showed a slightly milder phenotype, with *4xrCUG*$_{\sim 100}$ *[line 2]* progeny (n=63) being scored as either like wild-type, ([1] = 0.222), or having a mild phenotype, ([2] = 0.746) with few in category 3 ([3] = 0.032) and none in category 4 (Figure 3.3 F). This less severe phenotype in *4xrCUG*$_{\sim 100}$ *[line 2]* indicates that tergite phenotype may not directly correlate with reduced viability, as this line showed a greater reduction in viability than *4xrCUG*$_{\sim 100}$ *[line 1]*(Figure 3.1 B).

Expression of *4xrCAG*$_{\sim 100}$ appeared to result in stronger phenotypes than *4xrCUG*$_{\sim 100}$, with expression of *4xrCAG*$_{\sim 100}$ *[line 1]* giving no progeny in category 1, few in category 2 ([2] = 0.125), none in category 3 and the majority in category 4 ([4] = 0.875), (Figure 3.3 G). This line gave the strongest phenotype of all the lines tested. However, under these conditions adult viability was severely reduced in this line, with total n=8 over multiple crosses (compared the next lowest population of n=63), and therefore the distribution among categories is likely not accurate. Previously this line showed only a modest reduction in viability (Figure 3.1), however only female

progeny, which may be less viable, were scored for tergite disruption.

Figure 3.3 Comparison of tergite disruption in independent repeat lines and controls.

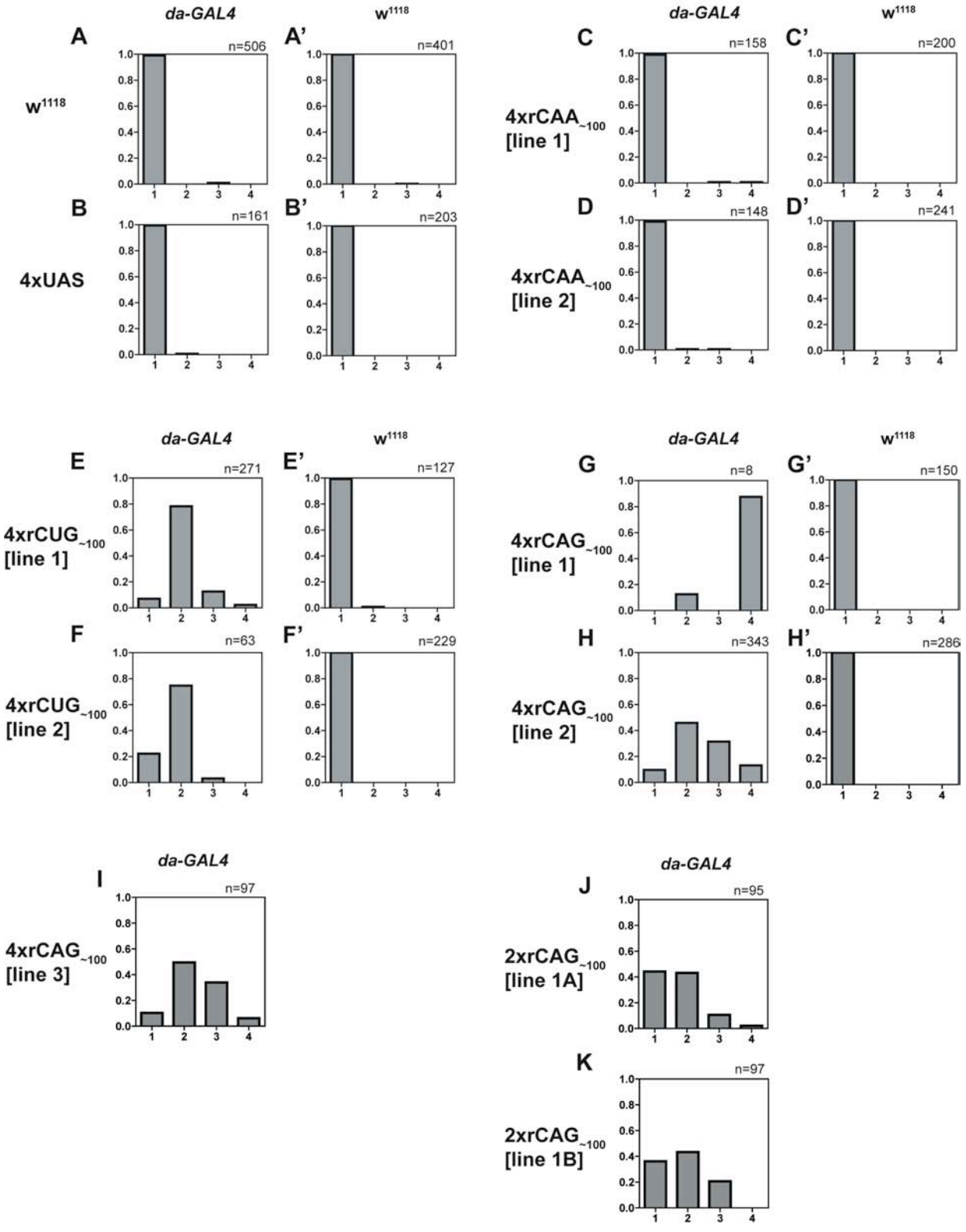


Figure 3.3
Graphs show the proportion of progeny within each category for all genotypes (full data in Appendix 2.1). Proportion (0.0 to 1.0) is shown on the y-axis while each category (1-4) is shown on the x axis. Total population size, n, is indicated above each graph. **A – K**, phenotype when each line is ubiquitously expressed with *da-GAL4*. **A' – K'**, phenotype when the same lines are crossed to w^{1118} to give progeny with all four repeat transgenes, in the absence of GAL4 driven expression. **A, A'** w^{1118} wild-type lines. **B, B'** *4xUAS* control line. **C, C'** *4xrCAA$_{\sim 100}$ [line 1]*, and **D, D'** *4xrCAA$_{\sim 100}$ [line 2]*. **E, E'** *4xrCUG$_{\sim 100}$ [line 1*], and **F, F'** *4xrCUG$_{\sim 100}$ [line 2*]. **G, G'** *4xrCAG$_{\sim 100}$ [line 1]*, and **H, H'** *4xrCAG$_{\sim 100}$ [line 2]*. I, *4xrCAG$_{\sim 100}$ [line 3]*. **J**, *2xrCAG$_{\sim 100}$ [line 1A]*, and **K**, *2xrCAG$_{\sim 100}$ [line 1B]*, the two transgene copy lines that were used to create *4xrCAG$_{\sim 100}$ [line 1]*. When each of the two copies is expressed via *da-GAL4*, **J**, **K**, the resultant phenotype is weaker than in the 4 copy line, **G**.

Progeny expressing *4xrCAG$_{\sim 100}$ [line 2]* progeny showed a phenotype milder than 4x*rCAG$_{\sim 100}$ [line 1]*, but still stronger than both *4xrCUG$_{\sim 100}$* lines. A small number of *4xrCAG$_{\sim 100}$ [line 2]* progeny (n=343) fell into category 1 ([1] = 0.096), with most having a mild phenotype ([2] = 0.458), or at least 1 tergite split ([3] = 0.315) and few with more than one tergite split ([4] = 0.131), (Figure 3.3 H).

Since numbers available for *4xrCAG$_{\sim 100}$ [line 1]* were limited by poor viability, a third *rCAG$_{\sim 100}$* line was tested. *4xrCAG$_{\sim 100}$ [line 3]* has some of its four transgene insertion sites in common with the other four copy lines, however the influence of insertion site on the phenotype seems minimal given the observation of phenotypes in multiple lines. Progeny expressing *4xrCAG$_{\sim 100}$ [line 3]* showed a phenotype similar to *4xrCAG$_{\sim 100}$ [line 2]*, with few progeny in category 1 ([1] = 0.133), most in category 2 ([2] = 0.517) or 3 ([3] = 0.267) and a small number in category 4 ([4] = 0.083), (Figure 3.3 I).

To examine whether the severity of the phenotype is altered by repeat RNA dosage, two independent lines carrying two transgene copies (*2xrCAG$_{\sim 100}$*) were expressed with *da-GAL4*. Since *4xrCAG$_{\sim 100}$ [line 1]* appeared to give the strongest phenotype, the two *2xrCAG$_{\sim 100}$* lines (*2xrCAG$_{\sim 100}$[line 1A]* and *2xrCAG$_{\sim 100}$ [line 1B]*) that make up *4xrCAG$_{\sim 100}$ [line 1]* were tested to determine if they gave milder phenotypes. This appeared to be the case with both *2xrCAG$_{\sim 100}$ [line 1A]* ([1] = 0.442, [2]=0.432, [3]=0.105, [4]=0.021) and *2xrCAG$_{\sim 100}$ [line 1B]* ([1] = 0.361, [2]=0.433, [3]=0.206,

[4]=0.000) giving similarly mild phenotypes (Figure 3.3 J, K). This result suggests that the tergite phenotype is dependent on repeat RNA dosage, with expression of a four transgene copy line giving a stronger phenotype than each pair of two transgene copies.

These results show that ubiquitous expression of either CUG or CAG hairpin-forming RNA repeats, but not a CAA repeat that is not predicted form a stable structure, can cause disruption to the development of the abdominal tergites in *Drosophila*. This observation suggests that hairpin-repeat RNA may perturb specific cell types or pathways. The phenotype is present in both CUG and CAG hairpin RNA expressing lines, indicating that both repeat sequences can cause cellular perturbation and that this may occur via a common pathway.

The *4xrCAG*$_{\sim 100}$ lines tested appeared to give a stronger phenotype than *4xrCUG*$_{\sim 100}$ lines, and reducing the dosage of CAG RNA to two transgene copies still caused a distinct phenotype. Furthermore, preliminary experiments using another ubiquitous driver, *Act5c-GAL4*, led to tergite disruption in hairpin forming CAG and CUG lines but not CAA or controls indicating that the phenotype is not an artefact of unusually high expression in certain tissues with *da-GAL4* (Appendix 2.2).

3.3 Repeat RNA expression in developing histoblast cells is sufficient to cause tergite disruption.

Following from the identification of tergite disruption in CUG and CAG repeat RNA expressing flies, experiments were undertaken to examine the basis for the phenotype. The *Drosophila* tergites are formed from the histoblast cells, a small population that arise during development and are symmetrically located either side of the larvae [192, 193]. During pupation these cells divide and proliferate moving from either side towards the midline to form the tergite bands, at the same time displacing larval epithelial cells (LECs) that undergo apoptosis [193] (Figure 3.4 A). Disruption of this process can lead to cleft tergites, resembling the phenotype observed with repeat expression [181]. It was therefore hypothesised that the repeat mediated tergite phenotype may be due to an effect on either of these cell populations. In repeat expressing flies the tergites do not appear to meet in the midline, indicating that the histoblast cells may have a reduced ability to proliferate and migrate, or that the mechanism by which larval epithelial cells are removed is disrupted. Experiments were therefore set up to examine the effect of hairpin RNA expression specifically on these cells.

The *T155-GAL4* driver was used to express repeat RNA within the developing adult epithelia that includes histoblast cells, but not LECs [185, 194]. A second driver, *Eip71CD-GAL4*, was used that drives expression within LECs, but not histoblasts [181]. In both cases crosses were initially performed at 29°C to ensure a high level of GAL4 activity and hence repeat expression. At 29°C *T155-GAL4* driven expression of the *4xUAS* control gave wild-type tergites in all progeny (Figure 3.4 B (i)). Likewise, with expression of $4xrCUG_{\sim 100}$ and $4xrCAA_{\sim 100}$, no phenotype was observed (Figure 3.4 B (ii – v)). Expression of $4xrCAG_{\sim 100}$ *[line 1]* led to a mild phenotype, while $4xrCAG_{\sim 100}$ *[line 2]* showed a very minor disruption (Figure 3.4 B (vi, vii)). Statistical analysis comparing the distributions between those with any phenotype (category 2, 3 and 4) and those without a phenotype (category 1), indicated a significant difference compared to *4xUAS* for $4xrCAG_{\sim 100}$ *[line 1]* ($p<0.0001$) and $4xrCAG_{\sim 100}$ *[line 2]* ($p=0.0145$) but none of the other lines. When driven ubiquitously (with *da-GAL4*) $4xrCAG_{\sim 100}$ *[line 1]* gave the strongest tergite phenotype (Figure 3.3), while expression in the developing histoblasts with *T155-GAL4*, led to a

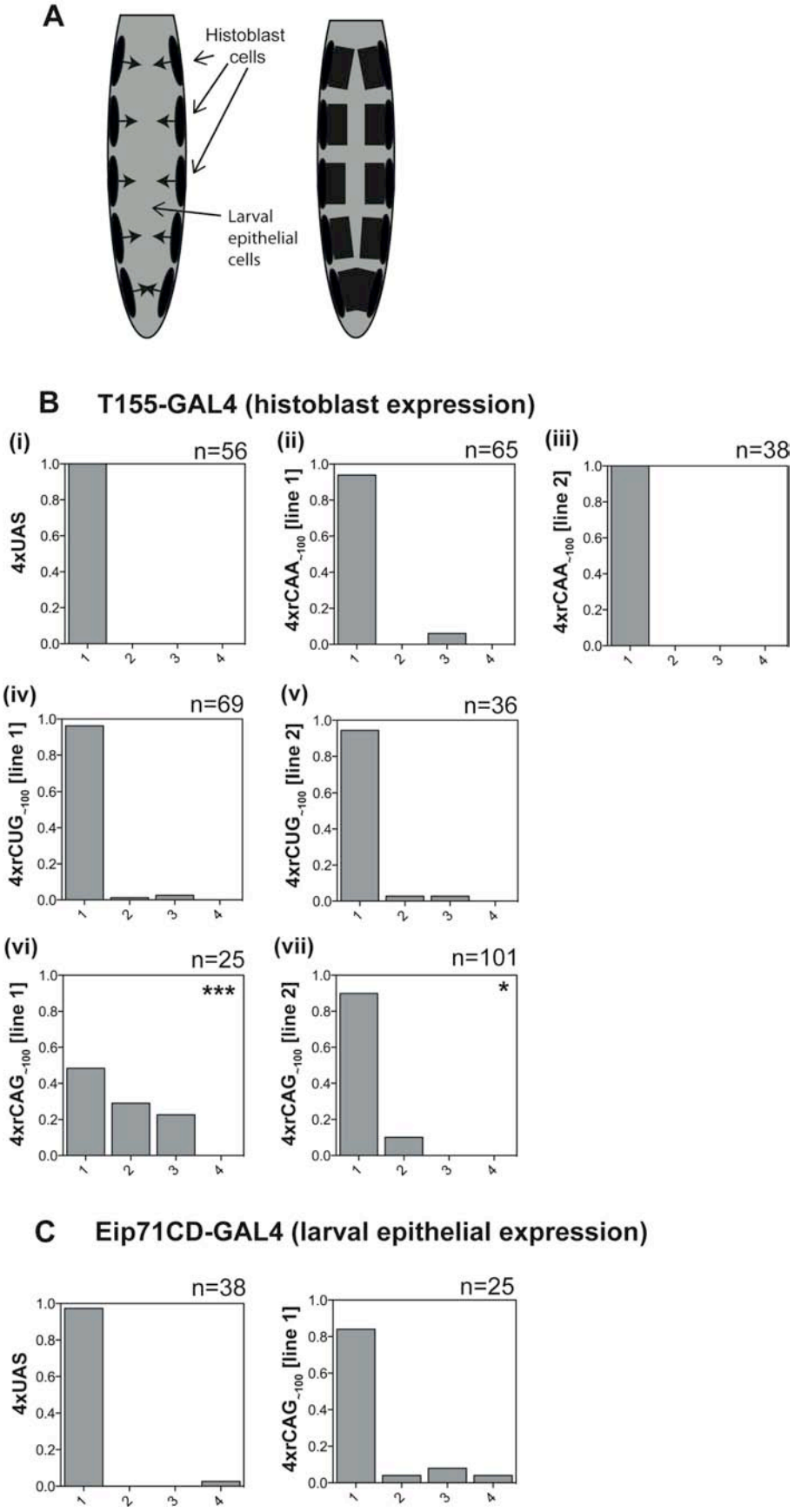
mild phenotype (Figure 3.4 B). In each case this was the strongest phenotype of all lines tested, such that the relative severity was consistent between drivers. Expression of *4xrCAG$_{\sim 100}$ [line 1]* in LECs using the *Eip71CD-GAL4* driver did not lead to a phenotypic distribution that was significantly different to that with *4xUAS* (Figure 3.4 C).

These results indicate that *rCAG$_{\sim 100}$* expression in histoblast cells is sufficient to induce tergite disruption in *Drosophila*. This provides preliminary evidence that repeat induced disruption of the adult tergites may occur through a cell autonomous effect on histoblasts. However, while expression in LECs was not sufficient to give a phenotype, it is possible that expression in both cell types may be necessary to give the strong phenotype observed with *da-GAL4*. Alternatively, the milder phenotype may be due to lower transgene expression from the *T155-GAL4* driver. Similarly for CUG RNA, which gives a milder *da-GAL4* phenotype, expression from *T155-GAL4* may not be sufficient to observe an effect. Further investigation will be required to determine the exact cellular mechanism by which disruption occurs. Nonetheless, these results support the observation that expression of repeat RNA in *Drosophila* leads to perturbation of specific cells to give the tergite disruption phenotype.

Figure 3.4

A, Histoblast cells (black) located either side of the larval body proliferate and migrate during pupation to form the tergite bands. Larval epithelial cells are displaced and undergo apoptosis [193]. **B**, Proportion of progeny within each tergite phenotype category when repeat constructs were expressed in histoblasts with *T155-GAL4*. Population size, n, is shown above each graph. Significance indicated is based on comparing the distribution of each population between those with any phenotype (category 2, 3 and 4) and those like wild-type (category 1), using Fisher's exact test. *P<0.05, **p<0.01, ***p<0.001. **C**, Graphs are as in **B**, but for expression in LECs using *Eip71CD-GAL4*, indicating that *rCAG$_{\sim 100}$ [line 1]* does not cause a significant phenotype in this case.

Figure 3.4 Expression of *rCAG~100* in histoblast cells leads to mild tergite disruption



3.4 Examining the effect of reduced *muscleblind* levels on RNA-mediated tergite disruption.

Results thus far suggest that tergite disruption is caused by specific cellular perturbation resulting from expression of CUG or CAG repeat RNA. The phenotype can be quantified and appears to be reproducible, and dosage dependent, and therefore provides an *in vivo* experimental system in which to examine common pathways that contribute to repeat RNA pathology. *Drosophila* has been well established as a system in which to examine pathways contributing to pathology by examining the ability of specific genetic alterations to modify a phenotype [173, 174]. Previous work investigating expanded repeat pathology in *Drosophila* has examined genetic modifiers of phenotypes cased by ectopic expression in the eye [37, 38, 174]. The same approach was taken using the tergite phenotype, thereby providing a unique system in which to examine pathways of cellular perturbation common to CUG and CAG repeat RNA.

Attempts were made to examine whether previously identified pathways of repeat RNA-mediated pathology are responsible for the tergite phenotype. One proposed pathway involves the sequestration of MBNL-1 by repeat RNA, leading to dysregulation of splicing and pathology. The *Drosophila* ortholog of MBNL-1, *muscleblind,* was therefore chosen as an initial candidate modifier. Previous work shows that loss of one copy of *muscleblind* leads to enhancement of CUG RNA toxicity in the developing *Drosophila* eye, while over-expression of MBNL-1 leads to suppression [87, 89, 116]. However, in a *Drosophlia* model of CAG RNA induced neurodegeneration, overexpression of *Drosophila muscleblind isoform A* enhanced toxicity, possibly via stabilising repeat RNA transcripts [132]. These contrasting results may indicate different involvement of *muscleblind* in the toxicity of each repeat sequence, however studies were done in different tissues and with different repeat constructs therefore preventing direct comparison. The tergite phenotype identified in this study enabled comparison of the effect of altering *muscleblind* levels on a common CUG or CAG repeat RNA-mediated phenotype. Experiments were undertaken with the aim of examining whether altering *muscleblind* levels is sufficient to modify the tergite disruption phenotype. It is expected that, if tergite disruption is caused by sequestration of muscleblind, a further reduction of

muscleblind levels would lead to an enhanced phenotype. In this case *da-GAL4* mediated phenotypes were examined as this gave a robust phenotype in multiple independent lines.

Preliminary observations revealed that over-expression of human MBNL-1, used in other studies, was lethal when expressed ubiquitously via *da-GAL4*, therefore it was not possible to determine its effect on the tergite disruption. However, a *muscleblind* allele, mbl^{E27}, was available with a deletion of exon 1 and 2, resulting in a non-functional protein [80]. mbl^{E27} is homozygous lethal and therefore tests were performed in a heterozygous mbl^{E27} / + background. Previous studies have shown that this level of *muscleblind* reduction is sufficient to modify CUG repeat RNA induced eye phenotypes [89, 116].

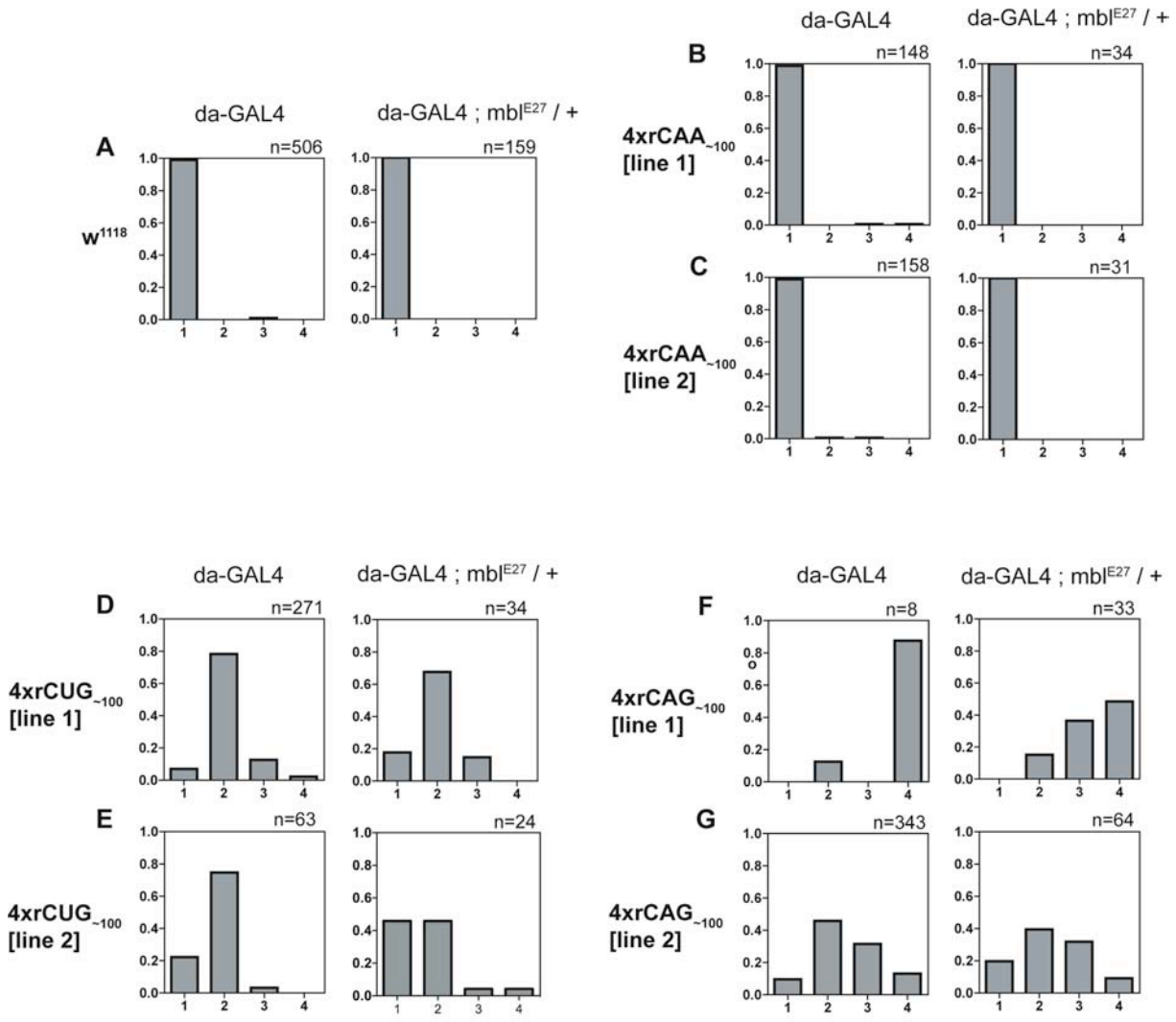
Repeat lines were crossed to a *da-GAL4* stock carrying the mbl^{E27} allele and the tergite phenotype in progeny was scored as described previously so that the distribution of phenotype strengths could be compared with and without mbl^{E27} (Appendix 3.1). Control progeny carrying *da-GAL4* alone previously did not show a phenotype, and this was not altered by the introduction of the mbl^{E27} allele ([1] = 1.000, n = 159) (Figure 3.5 A). Likewise, the presence of mbl^{E27} did not cause a phenotype in progeny expressing $4xrCAA_{\sim 100}$ *[line 1]* ([1] = 1.000, n = 34) and $4xrCAA_{\sim 100}$ *[line 2]* ([1] = 1.000, n = 31) (Figure 3.5 B and C).

Expression of $4xrCUG_{\sim 100}$ *[line 1]* with mbl^{E27} did not lead to a striking difference in phenotype distribution, however there was a slight increase in the proportion with wild-type tergites ([1] = 0.07, n=271 to [1] = 0.176, n=34) (Figure 3.5 D). The second line, $4xrCUG_{\sim 100}$ *[line 2]*, also showed an increase in the proportion of progeny like wild-type ([1] = 0.222, n=63 to [1]=0.458, n=24), and corresponding decrease in proportion with a phenotype when mbl^{E27} was present, indicating a mild trend towards suppression of the phenotype (Figure 3.5 E).

Expression of $4xrCAG_{\sim 100}$ *[line 1]* alone led to a strong phenotype and a similar phenotype was observed with mbl^{E27} with all progeny showing disruption ([1] = 0, n=33) (Figure 3.4 F). In this case accurate comparison was not possible due to the small population size (Figure 3.5 F). Expression of $4xrCAG_{\sim 100}$ *[line 2]* with mbl^{E27}

led to no obvious change in the phenotype severity compared to expression of *4xrCAG~100 [line 2]* alone (Figure 3.5 G).

Figure 3.5 Effect of reducing *muscleblind* levels on tergite disruption



Ubiquitous expression of repeat lines via *da-GAL4* alone (left, as indicated), or with one copy of the *mbl*^{E27} mutant allele (right, as indicated). Graphs show the proportion of progeny within each phenotype scoring category. Proportion (0.0 to 1.0) is shown on the y-axis while each category (1-4) is shown on the x-axis. Population size, n, is shown above each graph. Genotypes shown are **A**, w^{1118}, **B**, *rCAA~100 [line 1]*,**C**, *rCAA~100 [line 2]*, **D**, *rCUG~100 [line 1]*, **E**, *rCUG~100 [line 2]*, **F**, *rCAG~100 [line 1]*, **G**, *rCAG~100 [line 2]*.

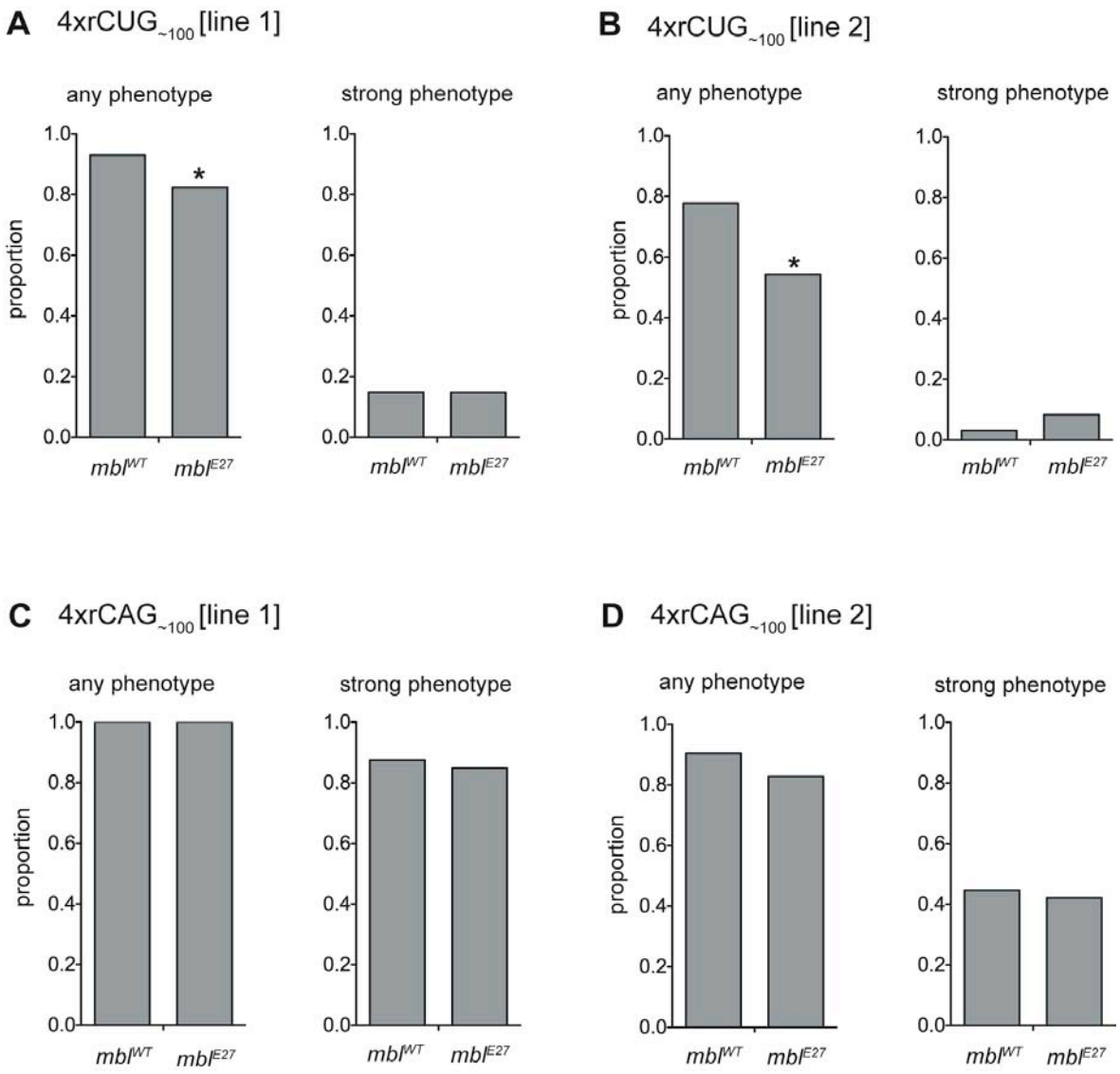
To examine if these effects were statistically significant data was combined to determine the proportion of the population with any phenotype ('any phenotype' – total in category 2, 3 and 4), and the proportion with at least one tergite completely split ('strong phenotype' – total in category 3 and 4). Fisher's exact test was used to compare the distribution between each of these measures and the remaining progeny, to determine if the differences observed with mbl^{E27} were significant (Figure 3.6, Appendix 3.4).

For *4xrCUG$_{\sim 100}$ [line 1]* this indicated a decrease in the proportion of progeny showing any phenotype (0.930, n=271 to 0.824, n=34) when mbl^{E27} was present (Figure 3.6 A) that was just within the p=0.05 threshold for significance (p = 0.045). The same analysis for *4xrCUG~100 [line 2]* showed a decrease in the proportion of progeny with any phenotype (0.778, n=63 to 0.542, n=24), that was also just within the threshold of significance (p=0.037) (Figure 3.6 B). When the distribution between the 'strong' phenotype category compared to the remaining population was examined no significant differences were found for either rCUG$_{\sim 100}$ line (Figure 3.6 A and B).

Statistical analysis of *4xrCAG$_{\sim 100}$ [line 1]* could not be accurately performed given the small number of progeny (Figure 3.6 C). Progeny expressing *4xrCAG$_{\sim 100}$ [line 2]* showed a slight decrease in the proportion of progeny with any phenotype (0.904, n=343 to 0.828 , n=64) (Figure 3.6 D). Comparison of the populations indicated that this change was not statistically significant (p=0.081). Likewise comparison of the number with a strong phenotype showed no significant difference in this population.

Although, in the presence of mbl^{E27} all lines appeared to show a mild decrease in the proportion with any phenotype, this was only weakly significant in the case of both *rCUG$_{\sim 100}$* lines. No significant change was observed when comparing progeny with a strong phenotype, suggesting that reducing muscleblind levels via the mbl^{E27} allele does not have a major effect on the tergite disruption phenotype. Furthermore, the mild effects observed involved a suppression, rather than the enhancement that would be expected based on the proposed mechanism involving sequestration of muscleblind [89, 142]. Together this data is not supportive of muscleblind sequestration as a rate limiting step in the common pathway leading to the tergite phenotype.

Figure 3.6 Statistical analysis of phenotypic changes due to reduced *muscleblind* levels.



The proportion of progeny showing any level of disruption ('any phenotype' - total for category 2, 3 and 4), and the proportion showing a strong phenotype ('strong' phenotype - total for category 3 and 4) when repeats are expressed alone (mbl^{WT}) or in the presence of one copy of mbl^{E27} **A**, *4xrCUG~100 [line 1]*, **B**, *4xrCUG~100 [line 2]*, **C**, *4xrCAG~100 [line 1]* and **D**, *4xrCAG~100 [line 2]*. Significance is indicated based on analysis comparing the distribution of progeny between groups using Fisher's exact test (Appendix 3.4). *p<0.05, **p<0.01, ***p<0.001.

3.5 Chapter discussion

Experiments described in this chapter aimed to examine the hypothesis that CAG and CUG repeat RNA contributes to pathology through common pathways. Experiments aimed to determine whether expression of different trinucleotide repeat RNA sequences leads to common phenotypes in *Drosophila*. Ubiquitous expression resulted in a reduction in adult viability and disruption to the abdominal tergites in survivors expressing either CUG or CAG hairpin-forming repeat RNA. The identification of common phenotypes therefore supports the hypothesis that repeat RNA may induce cellular perturbation through common pathways.

In the lines examined, tergite phenotypes were stronger in CAG than CUG repeat expressing lines. Previous work indicates that the transcripts in each case are expressed at approximately equivalent levels, indicating that the greater severity of the CAG repeat phenotype may be due to a property of the repeat itself (Appendix 4, Supplementary Figure S2) [166]. Conversely, when examining viability, one of the CUG lines showed the greatest reduction, yet only a mild tergite phenotype, suggesting that tergite disruption does not correlate with reduced viability in these lines. While this situation may be complicated if those with the most severe tergite defect are not viable, this does not appear to fit the observed data since most CUG expressing progeny are distributed about the mildest category, and very few show a strong phenotype. Expression of CAA RNA gave no phenotype, supporting the possibility that perturbation is specific to hairpin-forming RNAs. However, real time qRT-PCR results from our lab suggest that CAA RNA is present at a lower steady state level than CUG or CAG, and it is therefore not possible to make an accurate comparison (S.Samaraweera, unpublished) [38]. Low steady state levels may be expected if CAA RNA is degraded, as it is not predicated to form a stable secondary structure, however further investigation of this is required.

The tergite phenotype appears to involve an effect on specific cells despite ubiquitous expression from *da-GAL4*. Preliminary results indicate that phenotypes are also observed in these repeat expressing lines using the ubiquitous *Act5c-GAL4* driver, confirming that the effect is not due to a specific effect of the *da-GAL4* driver (Appendix 2.2). The formation of tergites has been characterised in *Drosophila* and

involves the proliferation and migration of histoblast cells, leading to invasion and subsequent apoptosis of larval epithelial cells [192, 193]. Ectopic expression in histoblasts led to a mild phenotype in the *rCAG*$_{\sim 100}$ line which gave the strongest *da-GAL4* phenotype, while expression of the same line within larval epithelial cells had no significant effect. This suggests that repeat expression within histoblast cells is sufficient to cause tergite disruption, and may indicate a direct effect on proliferation or migration of these cells. Tergite development has been previously identified as a useful system to study the mechanisms involved in proliferation, migration and invasion of epithelial cells in *Drosophila* [193, 195]. Therefore, the tergite phenotype may potentially be a useful system in which to examine the effect of repeat RNA on a number of basic cellular processes. Further analysis of the mechanism by which tergites are disrupted may provide insights into this aspect of disease.

The tergite phenotype provides a reproducible, quantitative phenotype that enables examination of common CUG or CAG repeat RNA mediated cellular perturbation. Using tergite disruption as a biological read-out of repeat mediated pathology, attempts were made to examine whether altering levels of the candidate protein muscleblind could modify the phenotype. Consistent modification was not observed, with a mild suppression in CUG expressing progeny only, and only those with a milder starting phenotype. These results are not consistent with the proposed mechanism whereby muscleblind is sequestered, such that further reduction of muscleblind is able to enhance CUG mediated phenotypes [142]. The mild suppression observed in the tergite phenotype may be attributed to other factors such as muscleblind regulation of RNA stability, where reduced muscleblind levels may result in a general reduction in RNA stability and subsequent reduction of repeat mediated effects [132, 139].

A limitation of the approach used is that muscleblind levels are only reduced by approximately 50% by introducing one copy of the *mbl*^{E27} allele. While homozygosity for *mbl*^{E27} is lethal, using an RNAi construct may enable levels to be reduced below 50% but allow viable adults, thus allowing the effect on tergite disruption to be examined. An RNAi line (VDRC #28731) was available at the time of these experiments, however results indicate that expression of this line does not significantly reduce *muscleblind* transcript levels, and may not target all isoforms (C.

van Eyk, personal communication). Using an alternative RNAi line that effectively targets all transcripts would enable this approach to be taken in the future. Nonetheless, these experiments provide evidence that the tergite disruption phenotype can be successfully used to examine whether specific pathways are rate-limiting for cellular perturbation. This provides a system in which to identify pathways that will provide candidates for further studies in vertebrate systems.

CHAPTER 4 : Repeat RNA nuclear localisation in *Drosophila*

Introduction

A mechanism proposed to account for repeat RNA pathology is the sequestration of RNA binding proteins through interactions mediated by the repeat RNA secondary structure. This involves the formation of nuclear RNA foci that co-localise with MBNL-1 and are hence thought to be sites at which binding proteins are sequestered [75, 76, 85]. RNA foci are observed in a number of diseases, suggesting that this may be a hallmark of repeat RNA-mediated pathology (Table 1.2). As the formation of RNA foci is proposed to play a role in the sequestration and dysregulation of MBNL-1, the identification of RNA foci in the *Drosophila* model described in this thesis could provide evidence for the involvement of this pathway in pathology. In *Drosophila*, expanded CUG repeat RNA has been shown to form foci that co-localise with muscleblind, and induce changes in splicing, suggesting that key components of the pathway are conserved [89, 142].

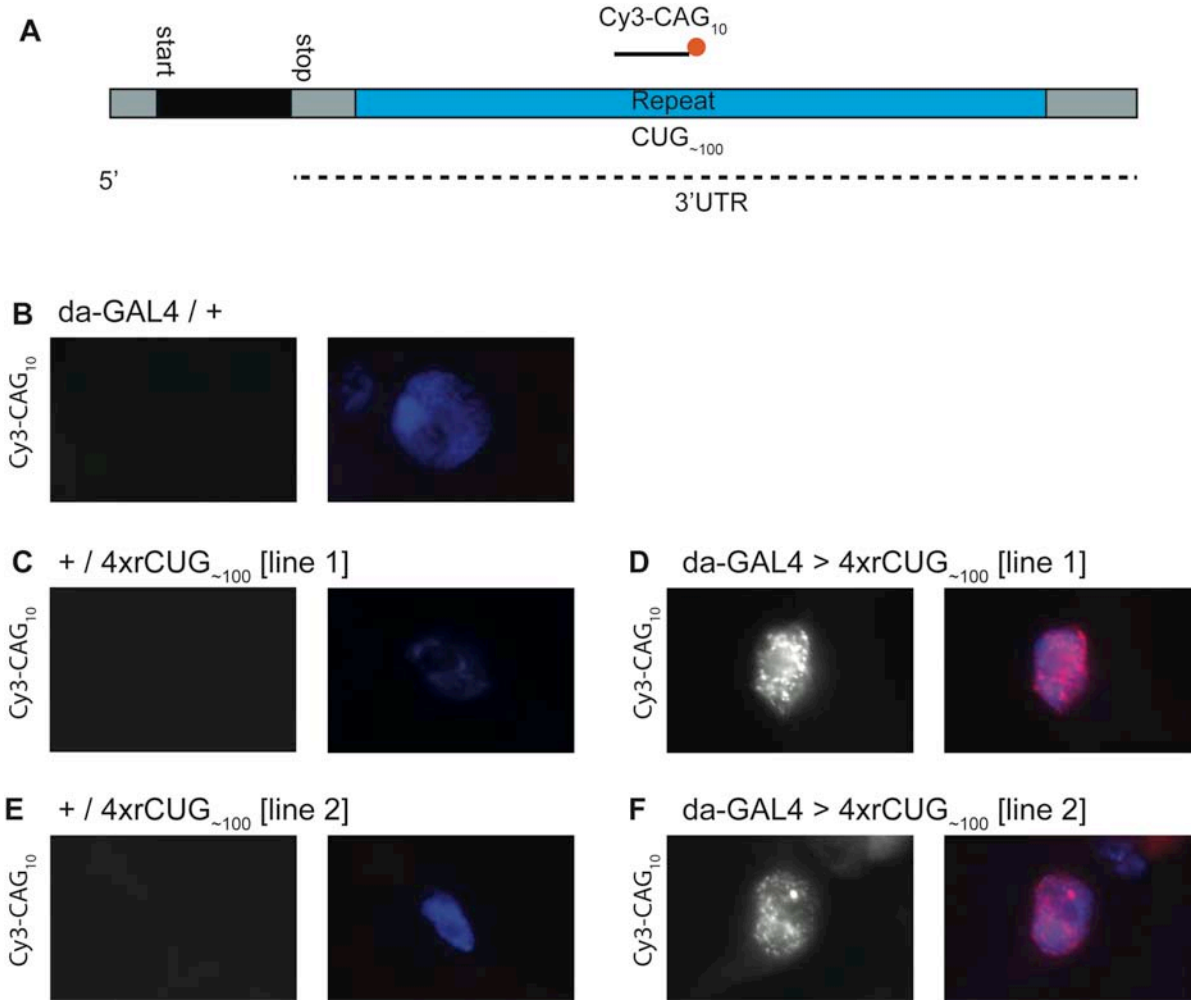
Experiments described in Chapter 3 identified a phenotype caused by either CUG, or CAG repeat RNA, but not CAA repeat RNA expression, involving disruption to adult tergite patterning due to the perturbation of specific cells. The phenotype was not consistently modified by reducing levels of *muscleblind*, suggesting that this pathway is not a rate-limiting contributor to cellular perturbation. To further examine a role for the pathway involving muscleblind sequestration, experiments in Chapter 4 aimed to determine whether repeat RNA in this model forms foci that correlate with pathology. A lack of correlation would provide further evidence that muscleblind sequestration is not responsible for the phenotype. As this model allows comparison of the localisation of hairpin-forming CAG or CUG repeat RNA with non-hairpin forming CAA RNA, a secondary aim was to examine and compare the intrinsic ability of each repeat to form RNA foci in *Drosophila*.

4.1 CUG repeat RNA forms specific nuclear foci within *Drosophila* larval muscles

Drosophila larvae expressing repeats ubiquitously via *da-GAL4* were used to examine repeat RNA nuclear localisation. In the context of the transgene, the repeat is within the 3'UTR of a short peptide sequence such that the repeat forms a majority of the transcript (Figure 1.5, 4.1, 4.3, 4.4) [37, 38]. Therefore, to examine localisation *in-situ*, tissue was hybridised with a fluorescent Cy3-CAG_{10}, Cy3-CTG_{10} or Cy3-TTG_{10} probe complementary to the CUG, CAG and CAA repeat sequence respectively (Figure 4.1, 4.3 ,4.4). Cryosections of whole third instar larvae were used to examine localisation in multiple tissues. As done previously, two transgenic lines for each repeat were tested, each carrying four independent insertions of the repeat transgene.

Initially *4xrCUG$_{\sim 100}$* expressing larvae were examined to determine whether foci were formed, as previously reported in *Drosophila* [139, 142]. To confirm binding specificity, the probe was hybridised to sections of *da-GAL4 /* + larvae that express *da-GAL4*, but do not carry a repeat transgene. No signal was observed with the Cy3-CAG_{10} probe in these larvae (Figure 4.1 B). As a further control, each line was crossed to the w^{1118} wild-type line to generate + */ 4xrCUG$_{\sim 100}$* progeny that carry all four repeat constructs, but do not express any CUG repeat RNA as *da-GAL4* is absent. For two independent lines control + */ 4xrCUG$_{\sim 100}$* larvae did not show any staining, indicating that the probe does not bind to the repeat DNA under these conditions (Figure 4.1 C and E).

Figure 4.1 Localisation of *rCUG~100* in *Drosophila* muscle nuclei

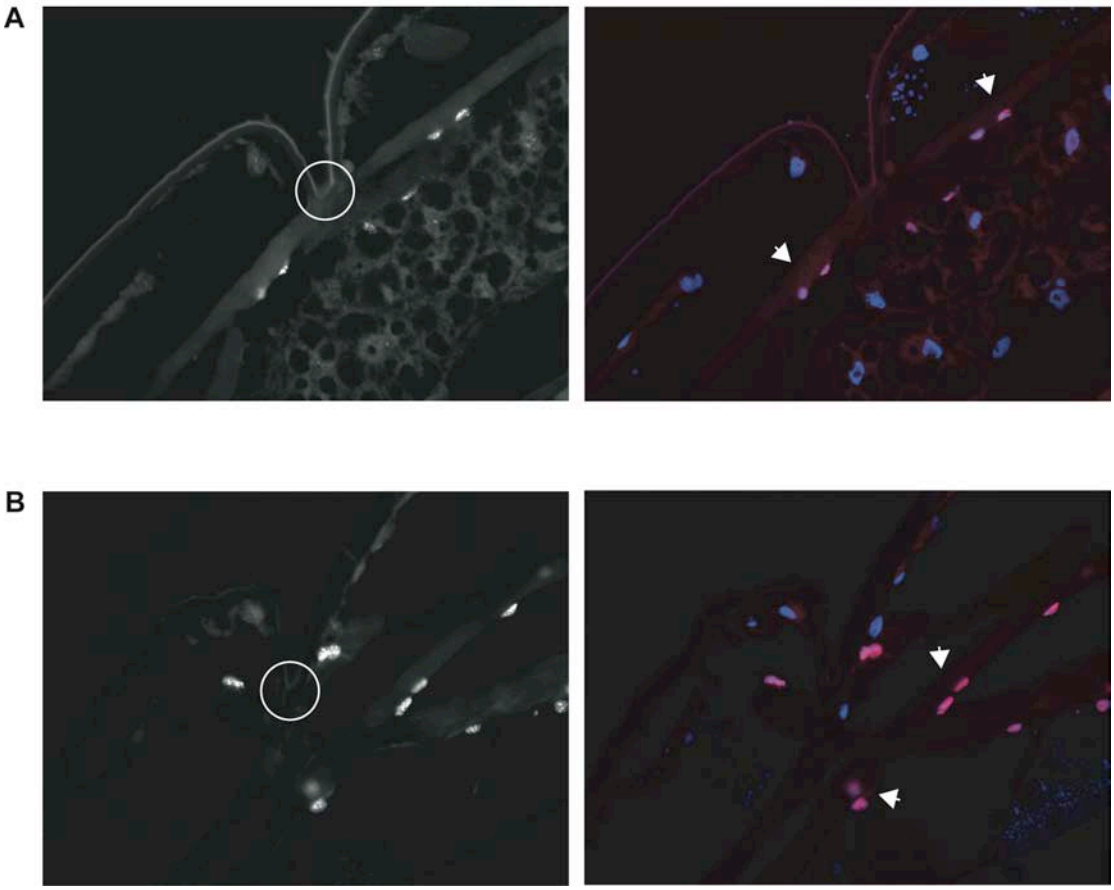


A, Schematic of the *rCUG~100* transcript (not to scale). A short non-functional peptide (black) is encoded upstream of the repeat (blue) which is within the 3'UTR (dotted line). Probes were designed to be complementary to the repeat, in this case a Cy3-CAG_{10} probe targets the $CUG_{\sim100}$ repeat. **B-F**, Microscope images (63x) of larval muscle cells probed with the Cy3-CAG_{10} probe. Left panel shows the Cy3 signal alone, right panel shows a merge of the Cy3 signal (red) and DAPI (blue) to label nuclei. **B**, *da-GAL4 /* + larvae show no Cy3 signal. **C**, *+ / 4xrCUG~100 [line 1]* progeny with four transgenes but no GAL4 driver show no Cy3 signal. **D**, *da-GAL4* driven expression of *4xrCUG~100 [line 1]* leads to many foci throughout the nucleus. **E**, *+ / 4xrCUG~100 [line 2]* progeny with no GAL4 driven expression show no signal, while, **F**, expression of *4xrCUG~100 [line 2]* via *da-GAL4* leads to multiple nuclear foci.

Larvae ubiquitously expressing *4xrCUG$_{\sim 100}$* via *da-GAL4* showed specific staining throughout the body, with similar patterns seen in two independent lines. Strongest staining was seen within nuclei, with weaker staining present in the cytoplasm of some cells. Many cells showed one to four nuclear foci, however a particular population was clearly identified with distinct staining consisting of numerous nuclear foci. In these cells many small foci were observed, which were distributed throughout the nucleus (Figure 4.1 D and F). These foci were identified in all larvae examined with multiple cells of this type present, and located within the same region of the body in each case. Nuclei were attached to large, clearly identifiable structures, just within the body wall, and sites of attachment to the body wall indicative of muscle fibres could be identified (Figure 4.2). Based on these observations, nuclei were identified as those belonging to muscle cells.

These results indicate that CUG RNA forms foci in muscle nuclei in this *Drosophila* model. This is consistent with previous studies and suggests that the repeat transcript used may undergo similar interactions [87, 139]. Non-muscle cells tended to show between one and four foci, indicating concentration of repeat RNA at particular sites. Given the presence of four transgenes this may correlate with a site of transcription, or processing, where between one and four sites would be visible depending on the focal plane. Since the *da-GAL4* driver leads to expression in all cells, these results indicate that cell-specific factors may be necessary to induce the numerous nuclear foci observed in only a subset of *Drosophila* cells.

Figure 4.2 Muscle specific *rCUG~100* localisation



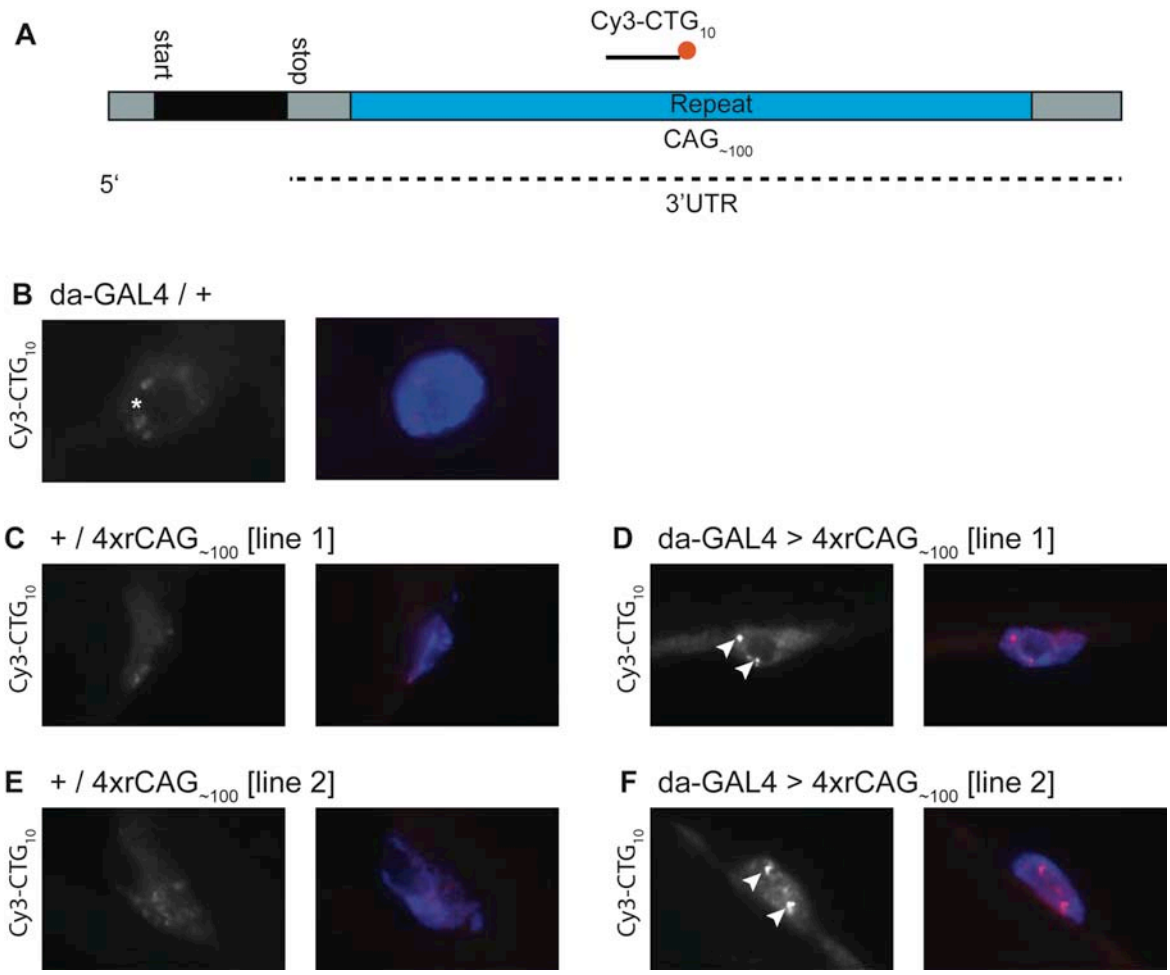
A, B, Examples from two different larvae expressing *4xrCUG~100* ubiquitously, hybridised with a Cy3-CAG_{10} complementary probe. All images at 20x magnification (compare to 63x in Figure 4.1). Left panel shows Cy3 signal and right panel shows a merge of Cy3 signal (red) showing repeat localisation, and DAPI (blue) showing nuclei. Background Cy3 staining highlights long muscle fibres that contain multiple foci containing nuclei (white arrows). Attachment to the cell wall indicative of muscle fibres is indicated by a white circle in each case.

4.2 CAG repeat RNA does not form muscle-specific nuclear foci

Repeat RNA localisation in larvae expressing *4xrCAG$_{\sim 100}$* was examined in the same manner as *rCUG$_{\sim 100}$*, using a complementary Cy3-CTG$_{10}$ probe (Figure 4.3 A). When hybridised to sections of larvae expressing *da-GAL4* but no repeat construct, the Cy3-CTG$_{10}$ probe showed a higher level of background than Cy3-CAG$_{10}$, with some cells showing a weak but specific pattern within the nucleus (Figure 4.3 B). This may be due to probe binding to endogenous transcripts as studies indicate a higher level of endogenous CAG containing transcripts, than CUG containing transcripts, in the *Drosophila* genome [196]. Since this background signal was weak, it was predicted that signal from specific CAG RNA binding would be clearly identifiable. Control + */ 4xrCAG$_{\sim 100}$* progeny for each line showed staining that was no stronger than the background signal observed with *da-GAL4 /* + (Figure 4.3 C, E).

When crossed to *da-GAL4,* progeny for both *4xrCAG$_{\sim 100}$* lines showed strong staining clearly distinguishable from background (Figure 4.3 D, F). Muscle nuclei showed only between one and four foci with no nuclei observed like *rCUG$_{\sim 100}$* expressing muscle nuclei that showed numerous nuclear foci (Figure 4.1 D, F). Examination of other cells throughout larvae showed a similar pattern, with none showing a distinct pattern of localisation. These results show that CAG repeat RNA is concentrated at sites within the nuclei of *Drosophila* cells, but does not undergo the muscle specific localisation observed with CUG repeat RNA. One possible explanation for the observations of one to four foci is that in muscle cells with CAG, and in non-muscle cells with both repeats, RNA is concentrated at sites related to transgene transcription, or normal processing. In this case expression of four transgenes would result in one to four visible sites depending on the focal plane examined.

Figure 4.3 Localisation of *rCAG~100* in *Drosophila* muscle nuclei



A, Schematic of the *rCAG~100* transcript (not to scale). A short non-functional peptide (black) is encoded upstream of the repeat (blue) which is within the 3'UTR (dotted line). Probes were designed to be complementary to the repeat, in this case a Cy3-CTG_{10} probe targets the *rCAG~100* repeat. **B-F**, Microscope images (63x) of larval muscle cells probed with the Cy3-CTG_{10} probe. Left panel shows the Cy3 signal alone, right panel shows a merge of the Cy3 signal (red) and DAPI (blue) to label nuclei. **B**, da-GAL4 / + larvae show a weak Cy3 signal due to background staining (asterisk). **C**, *+ / 4xrCAG~100 [line 1]* progeny with four transgenes but no GAL4 driver show only weak background staining. **D**, *da-GAL4* driven expression of *4xrCAG~100 [line 1]* leads to only one to four foci (arrowheads) throughout the nucleus. **E**, *+ / 4xrCAG~100 [line 2]* progeny with no GAL4 driven expression show only weak background staining, while, **F**, expression of *4xrCAG~100 [line 2]* via *da-GAL4* leads to only a small number of foci (arrowheads).

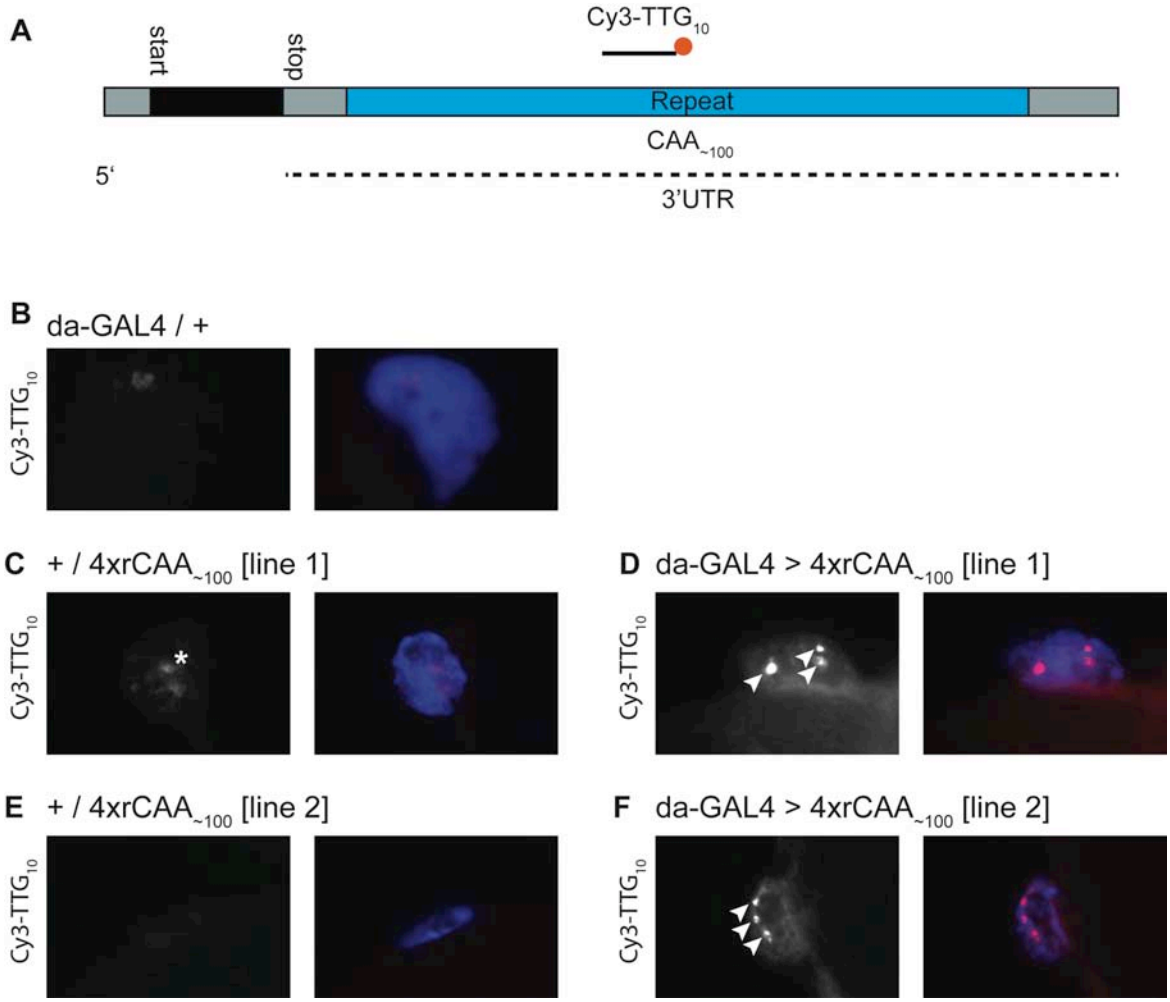
4.3 Non hairpin-forming CAA repeat RNA shows similar localisation to CAG repeat RNA

Localisation was next examined in *4xrCAA~100* expressing larvae. The CAA RNA repeat is not predicted to form a hairpin secondary structure, or bind to muscleblind [81, 197]. Therefore, by comparing to CUG and CAG repeat localisation it may be possible to determine whether the distinct localisation seen in each is dependent on hairpin-forming RNA, or whether over-expression of any expanded repeat results in concentrated sites of RNA within the nucleus.

Two independent sets of four copies of the *4xrCAA~100* transgene were expressed via *da-GAL4* and hybridised with a complementary Cy3-TTG_{10} probe (Figure 4.4 A). Control *da-GAL4 /* + sections showed a similar level of weak specific background staining as the Cy3-CTG_{10} probe (Figure 4.3 B). As for the Cy3-CTG_{10} probe, this may be due to a relatively high number of endogenous CAA repeat containing transcripts within the genome [196]. Each set of four independent *rCAA~100* transgenes in the absence of GAL4 showed a signal similar to that in *da-GAL4* / + (Figure 4.4 C and E). Progeny expressing *4xrCAA~100* via *da-GAL4* showed staining almost identical to that observed with *4xrCAG~100*. Muscle nuclei showed only one to four sites of RNA accumulation in two independent lines (Figure 4.4 D and F). Similar staining was observed in non-muscle cells. Therefore in this system non-hairpin forming CAA repeat RNA undergoes localisation similar to that seen with CAG repeat expression.

These similar findings for CAA and CAG RNA suggests that hairpin formation is not necessary for this type of nuclear accumulation to occur. This is consistent with the interpretation that *rCAG~100* and *rCAA~100* foci are the result of RNA accumulated during normal processing, and are unrelated to repeat specific effects. Relatively high levels of ectopic expression occurs in this case, which may lead to inefficient processing or transport, and therefore the high local concentration of RNA observed in these experiments. In contrast, CUG RNA expression gave multiple nuclear foci in muscle cells, in addition to the type of staining observed with CAG or CAA RNA in non-muscle cells.

Figure 4.4 Localisation of *rCAA~100* in *Drosophila* muscle nuclei



A, Schematic of the *rCAA~100* transcript (not to scale). A short non-functional peptide (black) is encoded upstream of the repeat (blue) which is within the 3'UTR (dotted line). Probes were designed to be complementary to the repeat, in this case a Cy3-TTG_{10} probe targets the $CAA_{\sim 100}$ repeat. **B-F**, Microscope images (63x) of larval muscle cells probed with the Cy3-TTG_{10} probe. Left panel shows the Cy3 signal alone, right panel shows a merge of the Cy3 signal (red) and DAPI (blue) to label nuclei. **B**, da-GAL4 / + larvae show only weak background staining. **C**, *+ / 4xrCAA~100 [line 1]* progeny carrying four transgenes but no GAL4 driver show only weak background staining (asterisk). **D**, *da-GAL4* driven expression of *4xrCAA~100 [line 1]* leads to one to four foci (arrowheads) throughout the nucleus. **E**, *+ / 4xrCAA~100 [line 2]* progeny with no GAL4 driven expression show only weak background staining, while, **F**, expression of *4xrCAA~100 [line 2]* via *da-GAL4* leads to one to four foci (arrowheads).

4.4 Repeat sequence specific localisation patterns are independent of transcript context

Experiments described thus far have involved the expression of repeat RNA within the 3'UTR of a short, non-functional peptide sequence. While *rCUG$_{\sim 100}$* expressing lines show distinct, muscle specific nuclear localisation, *rCAG$_{\sim 100}$* and *rCAA$_{\sim 100}$* expressing lines show only non-specific concentration of RNA in nuclei (Figure 4.1 - 4.4). To examine the possibility that these effects are related to the repeat expression construct, localisation of repeats expressed within an independent context was examined.

Repeats were expressed within the 5'UTR of GFP (named *4xrCUG$_{\sim 100}$-GFP* etc), so that a functional GFP protein was encoded by the transgene (Figure 4.5 A). GFP can be easily detected in live larvae prior to sectioning and provides an indicator that the repeat containing mRNA is properly processed, exported from the nucleus and translated. This approach also enables *in situ* probes complementary to the GFP sequence to be used (Figure 4.5 A), therefore overcoming the necessity to use different complementary repeat probes that led to background staining in some cases (Figure 4.3 B, 4.4 B).

As previously, transgenic lines carrying four copies of each expression construct were used. In this case, transgenic *Drosophila* were created using the *attB* targeted insertion system so that for each repeat line the same four insertion sites are used, leading to a comparable level of expression [188] (Materials and Methods). Each four copy line was expressed in wandering third instar larvae using *da-GAL4*, and localisation examined in separate experiments using a Cy3-GFP probe (Figure 4.5). In all cases GFP fluorescence was detected in larvae prior to sectioning, indicating that transcripts are processed and translated.

Expression of four insertions of the UAS-empty vector construct (*4xUAS*) did not show any signal when probed with Cy3-GFP. This line does not express the GFP transcript and therefore indicates that there is no non-specific binding by this probe (Figure 4.5 B). Expression of *4xrCUG$_{\sim 100}$-GFP* with *da-GAL4* resulted in numerous foci specifically within muscle cells when probed with the Cy3-CAG_{10} probe used

previously (Figure 4.5 D, Figure 4.1). Foci were almost identical to those observed in the two *4xrCUG$_{\sim 100}$* lines indicating that CUG repeats are able to form muscle specific foci in different transcript contexts. Muscle specific foci were also identified using the GFP specific Cy3-GFP probe (Figure 4.5 C). In this case, less foci were observed, possibly because only one Cy3-GFP probe can bind each transcript, while multiple Cy3-CAG$_{10}$ probes can bind the long repeat, resulting in greater sensitivity. Interestingly, with the GFP specific probe one to four brighter foci, surrounded by numerous smaller foci, were observed. These brighter foci may represent RNA concentrated at sites of transcription or processing, as observed with CAG or CAA expression.

Expression of *4xrCAG$_{\sim 100}$-GFP* with *da-GAL4* did not lead to numerous foci as seen in CUG lines. Instead, a Cy3-GFP probe revealed between one and four foci in muscle nuclei, as seen previously with *4xrCAG$_{\sim 100}$* lines (Figure 4.5 E). This supports the observations that CAG and CUG repeats undergo different localisation in muscle nuclei, and suggest that this can occur independent of transcript context. Results for *da-GAL4* driven expression of *4xrCAA$_{\sim 100}$-GFP* were similar to those for CAG lines. Using a Cy3-GFP probe, between one and four larger foci were observed in muscle nuclei (Figure 4.5 F). This is consistent with previous results and supports that CAA and CAG repeat expression leads to a similar pattern of localisation. Results using GFP lines therefore confirm findings that muscle specific foci are formed by CUG repeat RNA, but not CAG or CAA repeats in this system. Furthermore, since the GFP repeat lines are expressed from the same set of four insertion sites, these results suggest that the CUG specific effect is not simply due to a lower level of expression in other repeat lines.

If the foci seen in CAG and CAA lines are a result of the site of transgene transcription or processing, it is expected that expression of a GFP transcript in the absence of any repeat sequence may lead to a similar signal when using a Cy3-GFP probe. To test this, localisation of 4x*rCAG$_{\sim 100}$-GFP* and *4xrCAA$_{\sim 100}$-GFP* transcripts was compared to a GFP transcript not containing any repeat. Experiments showed that when a single copy of GFP was expressed using *da-GAL4*, muscle cells showed a single site of localisation within the nucleus similar to those observed with repeat expression (Figure 4.5 G). These results support that localisation observed in the

CAG and CAA lines may represent accumulation due to normal RNA processing. Further experiments would be required to determine with certainty the nature of nuclear RNA accumulations, however these results suggest that they are not repeat specific.

Figure 4.5 Nuclear localisation of repeats expressed within a GFP transcript.

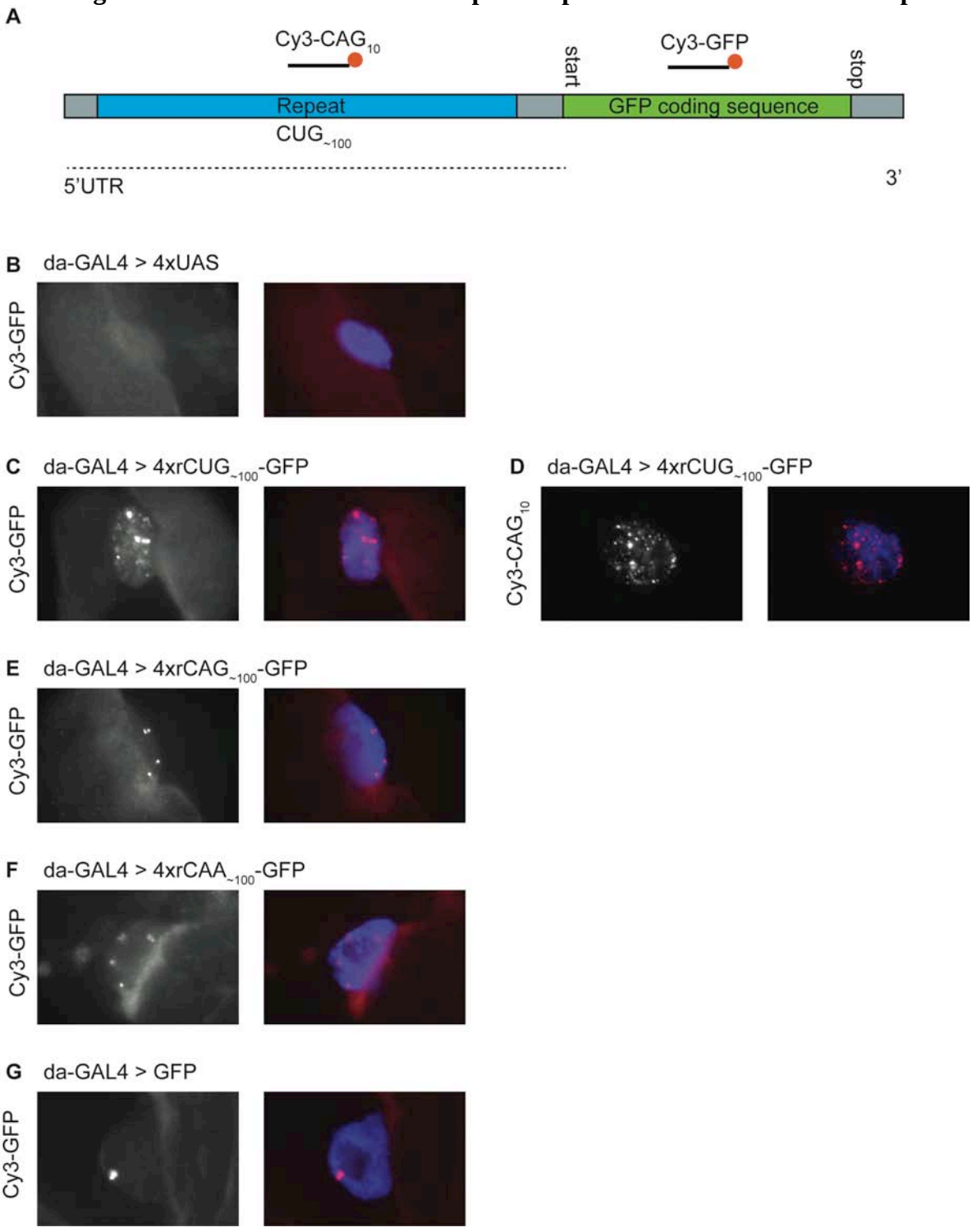


Figure 4.5
A, Schematic of the GFP repeat expression transcript (not to scale). The repeat sequence (blue) is within the 5'UTR of GFP (green). The transcript can be detected using either a complementary repeat probe, or a probe complementary to the GFP sequence (Cy3-GFP). **B-H**, Microscope images (63x) of larval sections that have been hybridised to detect repeat-GFP transcripts. Left panel shows Cy3 staining for repeat RNA, right shows merge of Cy3 signal (red) and DAPI staining of nuclei (blue). **B**, Expression of the empty vector construct produces no staining with a Cy3-GFP probe. **C**, Expression of *4xrCUG$_{\sim 100}$-GFP* leads to multiple foci in muscle nuclei when detected using a Cy3-GFP probe, or **D**, Cy3-CAG_{10} probe. **E**, Expression of *4xrCAG$_{\sim 100}$-GFP* leads to one to four single sites of RNA accumulation within muscle nuclei when detected with a Cy3-GFP probe. **F**, Expression of *4xrCAA$_{\sim 100}$-GFP* leads to one to four single sites of RNA accumulation when detected using a Cy3-GFP probe. **G**, Expression of a single copy of a non-repeat containing GFP transcript gives a single site of RNA accumulation similar to those seen in CAG and CAA expressing larvae.

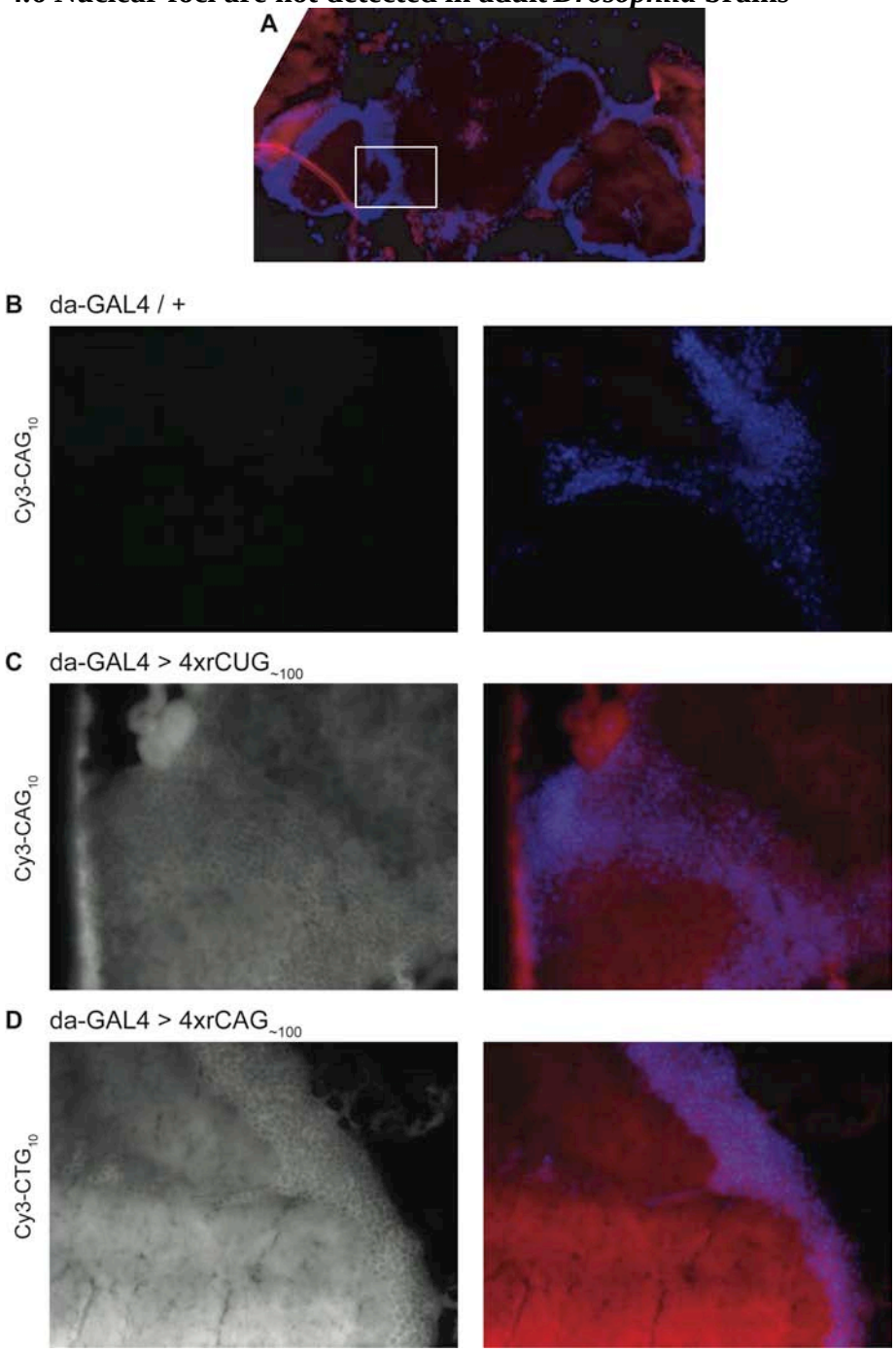
4.5 Repeat RNA foci are not observed in adult *Drosophila* brains

In previous experiments, specific repeat RNA nuclear localisation was observed in muscle cells of whole third instar larvae expressing CUG, but not CAG or CAA repeat RNA ubiquitously. However, CAG RNA may be able to form foci in other tissues or at other stages of development. Specific neuronal pathology is a major feature of dominant expanded repeat disease, and therefore RNA repeat localisation was also examined within the adult *Drosophila* brain to determine whether either repeat sequence forms specific foci in these cells.

Cryosections of whole adult brains from flies ubiquitously expressing repeats via *da-GAL4* were hybridised as for larval experiments. In this case using a Cy3-CAG_{10} probe did not reveal any specific nuclear signal within neurons expressing four copies of the *rCUG*$_{\sim 100}$ construct (Figure 4.6 A, C). A similar result was obtained when hybridising a Cy3-CTG_{10} probe to brain sections expressing four copies of the *rCAG*$_{\sim 100}$ construct, with no specific signal obtained within the nucleus (Figure 4.6 D). While previously non muscle cells showed sites of RNA accumulation in all repeat lines, these were not observed in adult brains with either repeat. In each case in brains, strong staining was observed outside the nucleus compared to the *da-GAL4*/+ control, indicating that RNA is being expressed (Figure 4.6 B - D).

While these experiments were not able to detect repeat RNA foci in adult *Drosophila* neurons, this approach is limited by the smaller size of neurons such that any foci present may be too small to detect. Examining larval brains, which have larger neurons, may provide an approach to overcome this limitation in the future. Alternatively, neurons may have a different sensitivity to both repeat, and non-repeat specific accumulation of RNA observed in muscle cells. This is consistent with previous work in *Drosophila* indicating that CUG RNA does not form foci in neurons [139]. However, neuronal foci are observed in human tissue, and therefore this may represent a limitation of the *Drosophila* system [99, 123].

Figure 4.6 Nuclear foci are not detected in adult *Drosophila* brains



A-D, Microscope images of sections of adult *Drosophila* brain, hybridised with probe to detect repeat transcripts. **A**, Example of a section at lower magnification showing the structure of the brain with the central tissue surrounded by nuclei (DAPI in blue). Repeat RNA is shown as Cy3 signal (red). White box indicates an example of the are shown at higher magnification below. **B-D**, Higher magnification (63x) images. Left image shows Cy3 signal from complementary repeat probe, right shows merge of Cy3 signal (red) and DAPI staining of nuclei (blue). **B**, *da-GAL4 /* + control brain shows no staining with a Cy3-CAG_{10} probe. **C**, Expression of *rCUG~100*, or **D**, *rCAG~100* leads to strong staining but no nuclear foci.

4.6 Chapter discussion

Results in this chapter indicate that CUG repeat RNA forms specific foci within muscle cell nuclei, while CAG or CAA repeat RNA does not show muscle specific localisation. In non-muscle cells, all three repeat sequences showed similar localisation with between one and four larger sites of RNA concentration observed in a repeat sequence-independent manner. Both CUG and CAG repeat RNA expression leads to disruption of adult tergites, and therefore results in this chapter support that pathways associated with the formation of specific nuclear foci do not contribute to this phenotype. One explanation for this may be that hairpin repeat RNA is able to induce pathology through an alternative mechanism to that involving formation of foci and protein sequestration. This is consistent with observations that reducing *muscleblind* levels did not enhance tergite disruption. Together these results support the possibility that the tergite phenotype is caused by a common hairpin RNA-mediated pathway that is independent of muscleblind sequestration to nuclear RNA foci.

These results also suggest that the ability to form specific RNA foci is not shared by all hairpin-repeat sequences in *Drosophila* such that CUG and CAG repeat RNA is able to undergo distinct localisation. Further studies will be required to determine if this is due to a property of our *Drosophila* system, or has wider implications. Several observations support that formation of CUG specific foci is due to an intrinsic property of the CUG repeat sequence in this case. Results were replicated in multiple independent transgenic lines as well as lines with each repeat sequence expressed from the same set of transgene insertion sites. This indicates that differences are not likely due to a higher level of expression in CUG lines. Likewise, CUG specific localisation was observed in two different transcript contexts, suggesting that the presence of the CUG repeat is sufficient to induce muscle specific localisation. While both CUG and CAG repeats are predicted to form a similar hairpin structure, structural studies indicate some differences between the two, and these differences may mediate interactions that underlie the distinct localisation observed in this current study [137].

While other have previously observed CAG specific foci [132, 133, 170, 171],

experiments presented in this chapter find no evidence for this, suggesting that CAG repeat expression alone is not sufficient to induce specific foci in this model. Other factors, or specific properties of the transcript, may be required. Interestingly, the results presented in this chapter appear to be the first direct comparison of hairpin-forming CAG and CUG localisation with a non hairpin-forming CAA repeat, and suggest that RNA may concentrate at sites in the nucleus in a hairpin-independent manner. Further work will be required to determine the basis for these observations, and whether this has any relevance to previous observations in *Drosophila*.

CUG specific localisation was only observed in muscle nuclei, suggesting the involvement of a muscle specific factor. This is in agreement with previous work in *Drosophila* examining the formation of CUG RNA foci [139]. Houseley et al, observe muscle specific nuclear foci with no pathology, while we observe tergite disruption, but this does not appear to correlate with the formation of muscle specific nuclear foci. Results in this chapter therefore appear to be consistent with the possibility of the tergite phenotype being independent of muscleblind sequestration. Similarly, expression of CAG RNA in *Drosophila* and mouse has previously been shown to form specific foci that correlate with pathology but are not associated with changes in splicing, suggesting that an alternative pathway may be involved [132, 171]. Further studies will be necessary to examine whether CAG RNA pathology in other models, and the tergite phenotype, involve common muscleblind-independent pathways.

Although there was no apparent correlation between the formation of specific nuclear foci and the tergite phenotype, attempts were made to examine the localisation of muscleblind when each repeat was expressed. Despite repeated attempts, the available antibody did not detect endogenous *muscleblind*, and expression of a tagged human MBNL-1 protein resulted in lethality when using the *da-GAL4* driver. In our model, co-localisation of muscleblind with both CUG and CAG repeat RNA would indicate that muscleblind is sequestered by both repeats, but that other factors are required to induce distinct CUG specific muscle foci. In this case, co-localisation with CAA RNA would not be expected. However, given the similarity between hairpin-repeat CAG localisation, and that of non-hairpin forming CAA repeat, it is possible that muscleblind would still be present due to a non-repeat mediated, and

possibly non-pathogenic role for muscleblind. This would be interesting given previous studies suggesting that CAG RNA is pathogenic and forms foci that co-localise with *muscleblind*, but does not lead to detectable changes in splicing [88, 132, 170]. In this case, muscleblind co-localisation as part of a non-pathogenic process would further support a role for repeat RNA in disease that is independent of sequestration of muscleblind in RNA foci.

CHAPTER 5 : Characterisation of dominant phenotypes from expression of a specific *rCAG$_{\sim 100}$* transgene insertion.

Note: A substantial portion of the work described in this chapter has been published as a component of the publication, Lawlor et al, Hum. Mol. Genet. 2011 [166]. (Included for reference as Appendix 4).

Introduction

Experiments described in this chapter aimed to characterise the basis for phenotypes in a specific transgenic line, *rCAG$_{\sim 100}$ [line C + D]*, carrying two copies of the CAG repeat expression transgene. Initial results presented here show that the observed dominant phenotypes were dependent on expression of *rCAG$_{\sim 100}$ [line C]*. It was therefore hypothesised that phenotypes are associated with a repeat mediated mechanism specific to the insertion site of this particular transgene. Subsequent work revealed that the repeat is inserted such that it is bi-directionally transcribed from the transgenic UAS sequence on one strand, and an endogenous promoter on the other. This leads to dominant phenotypes in the eye and in neurons and, with other results from our group, supports a novel mechanism of repeat RNA-mediated pathogenesis.

The *Drosophila* eye, a tissue containing both neuronal and non-neuronal cells, has frequently been used to model human neurodegenerative disease [173]. Therefore, previous work in our group examined the effects of RNA repeat expression specifically in the eye [37]. Initial results found no effect due to CAG RNA expression in the eye, and subsequently increasing dosage by expression of four transgene insertions of either *rCUG$_{\sim 100}$* or *rCAG$_{\sim 100}$* did not lead to disruption [37, 166]. However, a single line with two transgene insertions, line *rCAG$_{\sim 100}$ [line C +D]*, was identified that gave a mild disorganisation in the eye, and age-dependent degeneration of underlying cells when expressed with *GMR-GAL4 [177]*. Preliminary work indicated that pan-neuronal expression of this line also led to locomotion defects and a reduction in lifespan (K. Lawlor, unpublished) and therefore experiments were initially undertaken with the aim of determining the basis for these phenotypes.

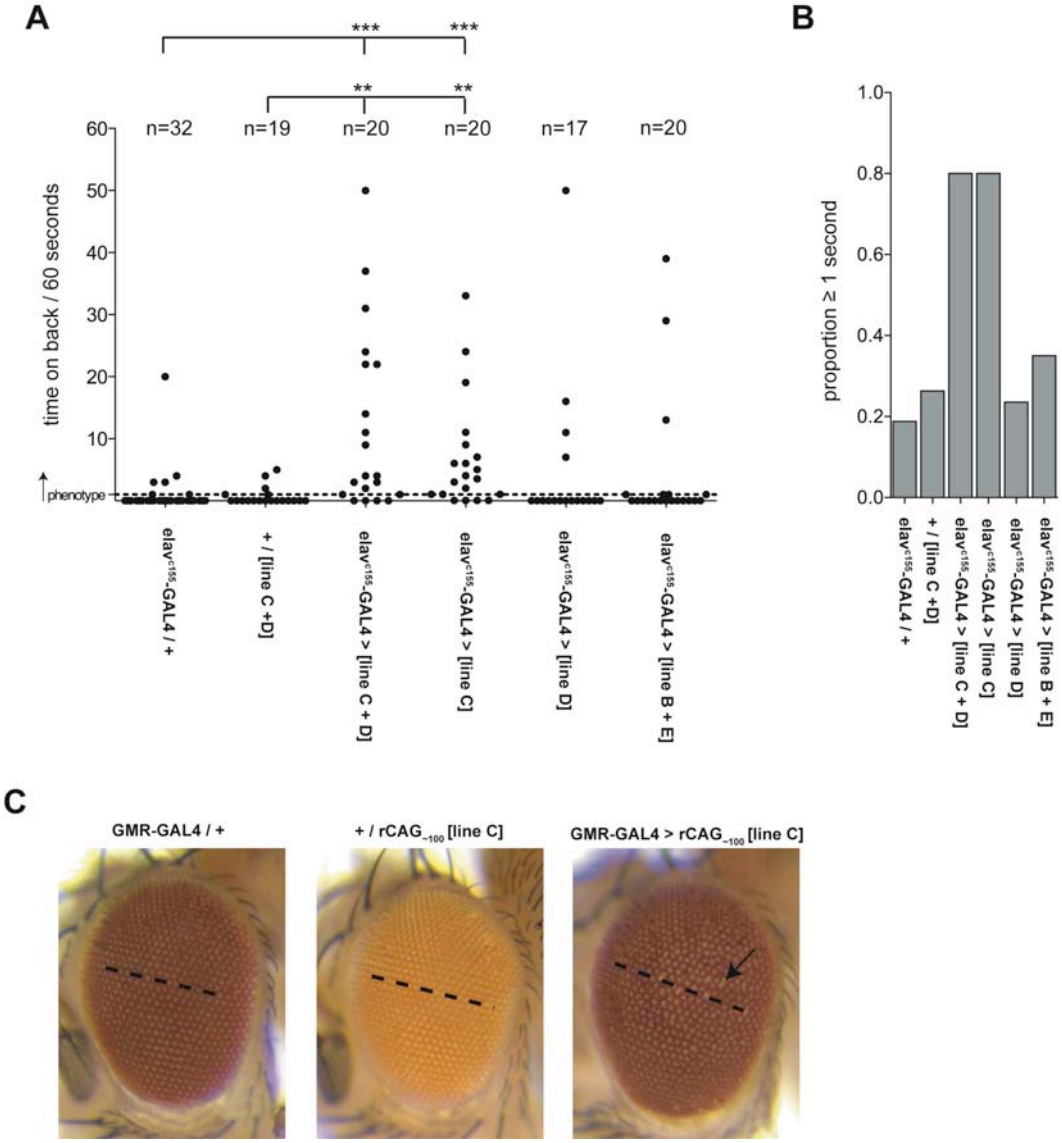
5.1 Expression of *rCAG$_{\sim 100}$ [line C]* is sufficient to cause dominant phenotypes in *Drosophila.*

In order to characterise the effects due to *rCAG$_{\sim 100}$ [line C + D]* expression in neuronal cells, flies were crossed to the pan-neuronal *elav^{c155}-GAL4* driver. Progeny were raised at 29°C, rather than the standard 25°C, to increase GAL4 activity and hence transgene expression level since this was shown to be necessary to reliably observe an effect in preliminary work. At 29°C, progeny expressing *rCAG$_{\sim 100}$ [line C + D]* showed locomotion defects that involved flies suddenly losing coordination and falling on their backs. This was followed by a brief period of uncoordinated movement in which flies could not right themselves. In some cases flies would fall on their backs and then become paralysed for several seconds, before righting and continuing normal walking (Appendix 4, Supplementary movie : hmg.oxfordjournals.org/content/20/19/3757/suppl/DC1). It appeared that in some cases the behaviour could be induced by knocking flies to the bottom of the vial.

Preliminary experiments indicated that the locomotion phenotype is enhanced by exposing flies to heat stress at 37°C. This effect has been reported previously for *Drosophila* mutants involving disrupted neuronal function [198]. In the present study, it is unclear whether this effect is due to a specific enhancement of the phenotype, or a general effect due to increased physiological stress, however this method provided a convenient means to observe the phenotype more readily. Male flies were separated into individual vials, allowed to recover 30 minutes following anaesthesia then placed at 37°C for approximately 60 minutes. Vials were knocked to induce the phenotype, then video taken of the 60 second period starting from when the vial was knocked on the bench. Video was later reviewed at a slower rate, or frame-by-frame, to score the phenotype. The time that each fly spent on its back during a 60 second period after being knocked was recorded, as this was easily identified and appeared to be a reliable indirect measure of the phenotype. The total time each fly spent on its back during 60 seconds was tallied to enable comparison between each population (Figure 5.1 A). As the total time appeared to be highly variable, the proportion of the population showing any phenotype was calculated and used to compared between groups (Figure 5.1 B). For this purpose flies spending more then 1 second on their backs during a 60 second period were deemed to show a

phenotype (Figure 5.1 A). Although this appeared to include a small proportion of control progeny, this low threshold was used to ensure that any mild, but high incidence effects, were not excluded.

Figure 5.1 Ectopic expression of *rCAG$_{\sim 100}$ [line C]* is sufficient to cause locomotion defects and disruption to the patterning of the eye.



A, Results from individual locomotion phenotype assays at 37°C. Total time spent on back was recorded for each fly over a 60 second period, and is plotted where 1 point represents a single fly. Those showing 1 second or greater (indicated by dotted line) were deemed to have a phenotype. Statistical comparisons were made between groups using Fisher's exact test to determine whether the distribution of progeny with/without a phenotype differed significantly between genotypes. * < 0.05, ** < 0.01, *** < 0.001. **B,** Proportion of progeny scoring 1 second or greater. **C,** Light microscope photographs of the *Drosophila* eye, anterior facing left in all. *GMR-GAL4* / + control eyes show wild-type patterning with a regular array of ommatidia forming straight lines across the eye (dotted line). *+ / rCAG$_{\sim 100}$ [line C]* eyes show a similar wild-type eye structure. The lighter colour is due to the *white*$^{-}$ genetic background where eye colour is determined by the transgene insertion. **C**, *GMR-GAL4 > rCAG$_{\sim 100}$ [line C]* eyes show mild disruption so that the ommatidia are disorganised (arrow) and do not always form straight lines across the eye (dotted line).

Control *elav^{c155}-GAL4 /* + progeny gave a phenotype proportion of 0.26 (n=32), while + / *rCAG$_{\sim 100}$ [line C + D]* progeny with both transgene insertions but no driver showed a proportion of 0.19 (n=19) (Figure 5.1 B). In both cases most individual flies that showed times greater than the 1 second phenotype threshold nonetheless showed only a very low total time (Figure 5.1 A). However, expression of *rCAG$_{\sim 100}$ [line C + D]* gave a much higher phenotype proportion of 0.80 (n=20), with the most severely affected flies showing much higher times than controls (Figure 5.1 A, B). Comparing the numbers of progeny with and without a phenotype revealed a significant difference in the distribution between populations compared to *elav^{c155}-GAL4 /* + ($p < 0.0001$, Fisher's exact test) and + / *rCAG$_{\sim 100}$ [line C + D]* ($p = 0.0012$), suggesting that this assay provides a reliable measure of the locomotion phenotype.

To examine the possibility of an insertion specific effect, each of the two transgene insertion lines that make up *rCAG$_{\sim 100}$ [line C + D]* were tested individually. Expression of *rCAG$_{\sim 100}$ [line C]* gave a similarly strong phenotype with a proportion of 0.80 (n=20) that was also significant compared to *elav^{c155}-GAL4 /* + ($p < 0.0001$) and + / *rCAG$_{\sim 100}$ [line C + D]* ($p = 0.0012$). Although individual progeny do not show as greater total times as *rCAG$_{\sim 100}$ [line C + D]* progeny in these populations, this may simply be a product of the large variation observed between individuals in both groups (Figure 5.1 B). Progeny expressing *rCAG$_{\sim 100}$ [line D]* showed only a low proportion with a phenotype (0.24, n=17) that was not significantly different to controls. Therefore, expression of *rCAG$_{\sim 100}$ [line C]* alone appears to be sufficient to cause the phenotype. Additionally expression of another two copy line, *rCAG$_{\sim 100}$ [line B + E]* gave a phenotype proportion of 0.35 (n=20) that was not significantly different to controls, suggesting that CAG RNA expression is insufficient to cause the effect. Together these results suggest that the locomotion phenotype is due to a specific property of the *rCAG$_{\sim 100}$ [line C]* insertion.

To determine if this effect is the same in the eye, *rCAG$_{\sim 100}$ [line C]* transgenic flies were crossed to *GMR-GAL4* to ectopically express the construct within all cells of the developing adult eye [184]. As a control, *GMR-GAL4 /* + progeny with the driver but no repeat construct were examined and showed a regular array of ommatidial units

characteristic of the wild-type *Drosophila* eye (Figure 5.1 C). + / *rCAG$_{\sim 100}$ [line C]* progeny with the insertion but no GAL4 driver also showed a regular array of ommatidia similar to *GMR-GAL4*/ +, indicating that the presence of the insertion alone is insufficient to cause a phenotype (Figure 5.1 C). In progeny expressing *rCAG$_{\sim 100}$ [line C]* the eye was mildly disrupted so that ommatidia were not arranged in regular lines across the eye, and appeared disordered (Figure 5.1 C). These results suggest that, as for the locomotion phenotype, expression of *rCAG$_{\sim 100}$ [line C]* is sufficient to cause disruption to the eye.

Together, these results indicate that expression of *rCAG$_{\sim 100}$ [line C]* is sufficient to cause effects in both the *Drosophila* neurons and eye. Results do not support that this is simply due to a higher level of repeat RNA expression in this line as expression of *rCAG$_{\sim 100}$ [line C]* alone led to a similar locomotion phenotype in the presence of increased dosage from a second transgene insertion, and the phenotype was not observed in independent two transgene lines. Likewise, subsequent work has shown that expression of independent four copy *rCAG$_{\sim 100}$* lines does not reproduce the locomotion phenotype [166]. In the eye, expression of *rCAG$_{\sim 100}$ [line C]* led to mild disruption while expression of four copies of *rCAG$_{\sim 100}$* has no significant effect [166].

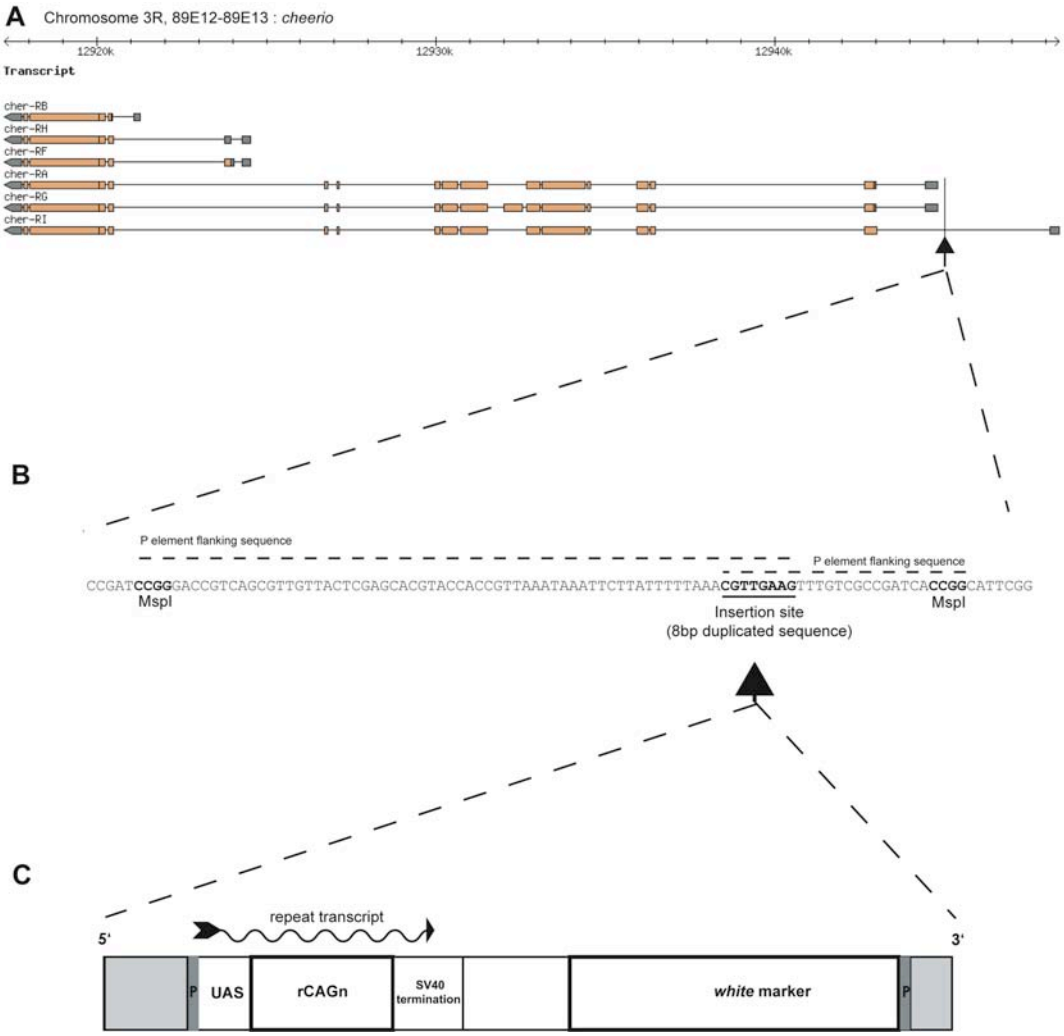
An alternate explanation may be that the *rCAG$_{\sim 100}$ [line C]* insertion causes an effect by disrupting an endogenous gene through the process of random integration used to generate these stocks. However, in the case of both the locomotion and eye phenotypes, progeny are heterozygous for the transgene and thus a loss of function insertional mutation could only cause a phenotype in the case of haploinsufficiency. In this case, insertional mutations disrupting the coding sequence of an endogenous gene would be expected to cause disruption regardless of the presence of a GAL4 driver. In contrast, phenotypes appear to be dependent on expression of the transgene, and are specific to the cells in which expression occurs, indicating a dominant effect. These results therefore indicate that rather then the insertion directly disrupting an endogenous gene, the phenotype occurs via a mechanism that requires transcription of the transgene. This may involve alterations to transcription of an endogenous gene, or a specific repeat mediated effect that is dependent on the insertion site. Given the possibility of a repeat mediated effect further investigation of *rCAG$_{\sim 100}$ [line C]* was undertaken.

5.2 *rCAG$_{\sim 100}$ [line C]* is inserted at the *cheerio* locus

To further examine the basis for the *rCAG$_{\sim 100}$ [line C]* phenotype, inverse PCR and sequencing were used to determine the genomic site of transgene insertion. Genomic sequence flanking the 5' and 3' P elements of the transgene confirmed that the insertion is on chromosome 3, located within the *cheerio* gene (Figure 5.2 A, B). At the time the experiment was initially conducted, the site of insertion was predicted to be within the 5'UTR of the *cheerio-A* isoform, 34bp downstream from the start of the transcript (NM_079659.2). However, subsequent revisions of the annotation (NM_079659.3) place the start of the *cheerio-A* transcript further downstream such that the insertion now maps to a site within intron 1 of the *cheerio-I* isoform (Figure 5.2 A). In either case, the insertion does not appear to disrupt the coding sequence. Furthermore, previous studies indicate that homozygous *cheerio* mutants are female sterile, but otherwise develop to viable adults, and there is no evidence of locomotion or eye disruption phenotypes in *cheerio* loss of function mutants [199-201].

The *rCAG$_{\sim 100}$ [line C]* transgene is inserted such that UAS driven transcription is in the opposite direction to the endogenous cheerio promoter (Figure 5.2 C). Therefore it was predicted that expression of the transgene may result in reduced *cheerio* levels through a mechanism such as transcriptional interference [202]. To examine the effect on *cheerio* transcript levels when driving the *rCAG$_{\sim 100}$ [line C]* insertion, real time quantitative PCR (qPCR) was performed using primers that detect all full length *cheerio* isoforms. Template cDNA was obtained from + / *rCAG$_{\sim 100}$ [line C]*, *GMR-GAL4* / + and *GMR-GAL4 > rCAG$_{\sim 100}$ [line C]* adult heads. Across three biological replicates, results indicated that there was no significant difference in *cheerio* levels between + / *rCAG$_{\sim 100}$ [line C]* and *GMR-GAL4* / + (Figure 5.3 A). This suggests that the presence of the insertion does not disrupt cheerio expression. In *GMR-GAL4 > rCAG$_{\sim 100}$ [line C]* samples, driving transcription also did not lead to a decrease in *cheerio* levels, instead there was an unexpected increase in transcript levels (Figure 5.3 A). Therefore, driving transgene transcription does not lead to a reduction in *cheerio* transcript levels. Previous studies report that over-expression of *cheerio* does not lead to disruption of the eye, suggesting that this is not the likely cause of the *rCAG$_{\sim 100}$ [line C]* eye phenotype [203-205].

Figure 5.2 The *rCAG~100 [line C]* insertion is within the *cheerio* gene



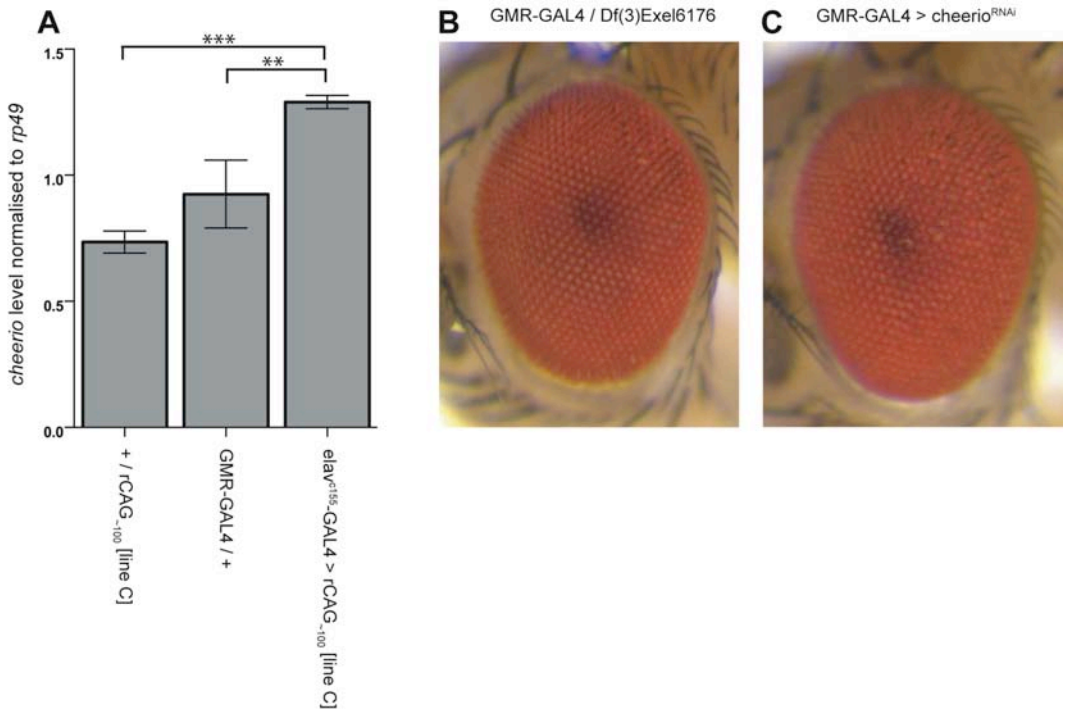
A, Approximate position of the insertion in relation to *cheerio* transcripts, adapted from flybase.org, Gbrowse feature [206]. **B**, Specific details of the insertion site. 8bp duplicated region flanking the insertion is indicated in bold and underlined. *Msp1* sites used in inverse PCR analysis are indicated and recovered genomic flanking sequence is indicated by dotted lines. **C**, Schematic (not to scale) of the insertion with orientation shown relative to the *cheerio* transcript. Transcription of *rCAG~100* occurs from the UAS sequence and produces a transcript that is terminated by the SV40 sequence. The white marker gene is present to enable the selection of transgenic insertion whilst lines are being established, but is not under UAS control.

To confirm these results genetically, attempts were made to examine whether reduced *cheerio* levels can replicate the *rCAG$_{\sim 100}$ [line C]* eye phenotype (Figure 5.1 C). The deficiency line *Df(3)Exel6176* removes a region of chromosome 3 from 89E11-89F1, which includes *cheerio* and approximately 20 other genes. While flies homozygous for this deficiency are not viable, heterozygotes do not show any disruption to the eye. Flies containing the deficiency in combination with *GMR-GAL4* also showed no disruption to the eye (Figure 5.3 B). Likewise, *GMR-GAL4* driven expression of an RNAi line targeting all *cheerio* isoforms (VDRC KK107451) did not result in any significant disruption to the eye (Figure 5.3 C).

Examining the effect of alterations in *cheerio* in the nervous system is more complex than the eye, however, some evidence suggests that the *rCAG$_{\sim 100}$ [line C]* locomotion phenotype is not caused by an alteration of *cheerio* levels. In *Drosophila,* both increasing and decreasing cheerio levels in the brain have been shown to lead to learning defects, however no locomotion defects like those observed with *rCAG$_{\sim 100}$ [line C]* have been reported [207, 208]. Furthermore, some evidence indicates that *cheerio* mediated effects in the brain may be dependent on the shorter *cheerio-B* isoform (Figure 5.2) that appears less likely to be effected by the *rCAG$_{\sim 100}$ [line C]* insertion, since its transcription start site is approximately 15kb away [208]. Additionally, reducing *cheerio* levels via the *Df(3)Exel6176* deficiency allele, did not lead to locomotion defects. Together these results support the conclusion that alterations in cheerio levels are not sufficient to explain the observed phenotypes in *rCAG$_{\sim 100}$ [line C]*.

In further support of this, flies homozygous for *rCAG$_{\sim 100}$ [line C],* but lacking a GAL4 driver, did not show any disruption to the eye, or locomotion phenotype. However, adults showed a phenotype whereby the wings were permanently ‘held out’. This was never observed in heterozygotes alone, or with any of the GAL4 lines tested. Trans-heterozygotes for *rCAG$_{\sim 100}$ [line C]* and *Df(3)Exel6176* did not have ‘held out wings’, suggesting that this effect is not due to loss of *cheerio* function.

Figure 5.3 *rCAG~100 [line C]* insertion phenotypes are not caused by a decrease in *cheerio* levels



A, Real time qPCR results showing the level of *cheerio* transcript normalised to housekeeping gene *rp49*. Bars show the mean ± standard deviation for three biological replicates of each genotype. *GMR-GAL4 > rCAG~100 [line C] cheerio* levels are significantly increased compared to each of the other genotypes, while + / *rCAG~100 [line C] cheerio* levels are not reduced compared to the GMR-GAL4 / + control. Student's t-test, ** $p<0.01$, *** $p<0.001$. **B, C**, Comparison of light microscope images of the external structure of the adult eye, anterior to the left in all cases. **B,** Reduction of *cheerio* via the *Df(3)Exel6176* deficiency allele in the presence of GMR-GAL4, and **C**, expression of an RNAi line targeting *cheerio* do not lead to disruption of ommatidial patterning.

5.3 *rCAG$_{\sim 100}$ [line C]* enables bi-directional expression of an expanded repeat

Results suggest that the *rCAG$_{\sim 100}$ [line C]* phenotypes are not due to alterations in *cheerio* levels, and therefore an alternative explanation may be a dominant effect of the repeat RNA that is mediated by transcription from this specific locus. Some basis exists for this in a study using *Drosophila* to examine the effect of CUG and CAG RNA toxicity where a phenotype was identified in only a single transgenic line [209]. In this case the transgene was inserted in the intron of an endogenous gene [209]. Based on results showing CUG RNA co-localisation with the endogenous transcript, the authors conclude that the phenotype was due to the formation of a fusion transcript placing the repeat within a new context that enabled it to mediate toxicity [209].

The *rCAG$_{\sim 100}$ [line C]* insertion site is such that expression of the transgenes from the UAS site occurs in the opposite direction to that of the endogenous *cheerio* promoter (Figure 5.2). Therefore if a fusion transcript were formed, it would involve sequences from the 5' region of *cheerio* that were transcribed from the opposite strand to that of the endogenous *cheerio* transcript. At the time of analysis, the insertion was predicted to be 34bp downstream of the start of the *cheerio-A* transcript (NM_079659.2). Therefore forming a fusion transcript in this case would only add a short amount of sequence. In the next revision of the genome (NM_079659.3) the insertion site is predicted to be within the first intron of the *cheerio-I* transcript. In this case it may be possible to form a fusion transcript with exon 1 of *cheerio-I*, however in either case this process would require UAS driven transcription to override the SV40 termination sequence within the construct and transcribe across the majority of the insertion construct sequence before including any endogenous genomic sequence (Figure 5.2). While it is not possible to rule this out, this has not been previously reported, and does not appear to clearly explain how the effect may be unique to *rCAG$_{\sim 100}$ [line C]*.

Another possible mechanism by which a repeat containing fusion transcript may be formed is if the repeat transgene is transcribed from an endogenous promoter. In this case, transcription from the *cheerio* promoter may continue into the transgene insertion producing a transcript containing part of the *cheerio* transcript, and part of the transgene insertion. As the *cheerio* promoter is on the opposite strand to the

inserted UAS promoter, this could result in the transcription of a *cheerio* transcript containing a complementary CUG repeat (Figure 5.2).

To test this hypothesis, primers were designed to amplify products flanking the transgene insertion boundaries (Figure 5.4 A). cDNA was created using RNA from *rCAG$_{\sim 100}$ [line C] / Df(3)Exel6176* trans-heterozygotes to ensure that any products obtained were from the allele with the insertion, rather than the wild-type *cheerio* allele. Control primers spanning the *cheerio* exon 2/3 boundary produced a product of expected size for a w^{1118} control, and *rCAG$_{\sim 100}$ [line C] / Df(3)Exel6176* cDNA, indicating that cheerio transcription is not terminated by the insertion (Figure 5.4A (i), B). For primers spanning the 5' transgene insertion boundary, the correct size product was obtained and confirmed by sequencing, while no product was obtained from control w^{1118} cDNA as expected (Figure 5.4 A (ii), B). This indicates that transcription from the *cheerio* promoter continues into the transgene insertion. Using a second set of primers spanning the repeat and 3' insertion boundary, a product was obtained and sequenced to confirm its correct identity, and the presence of a CUG repeat (Figure 5.4A(iii), C). This indicates that the repeat is present in a transcript containing a fusion of the complement of the repeat construct as well as part of the *cheerio* sequence.

To determine whether transcription spans the entire insertion, primers either side of the insertion were used (Figure 5.4 A (iv), C). In this case a product that was shorter than expected was obtained (Figure 5.4 C). Sequencing revealed that this product represented an alternatively spliced transcript spanning the 5' insertion boundary ending within the P element, fused with *cheerio-A* exon 2, suggesting that the repeat was spliced out (Figure 5.4 D). In this case, part of the cheerio 5'UTR is lost, but the coding region is not affected. This smaller transcript would likely by favoured in PCR, and therefore it is not possible to exclude the possibility that transcription occurs across the entire insertion. More extensive analysis via northern blot would be necessary to determine this.

Figure 5.4 The *rCAG~100 [line C]* insertion is transcribed to produce a complementary rCUG repeat transcript.

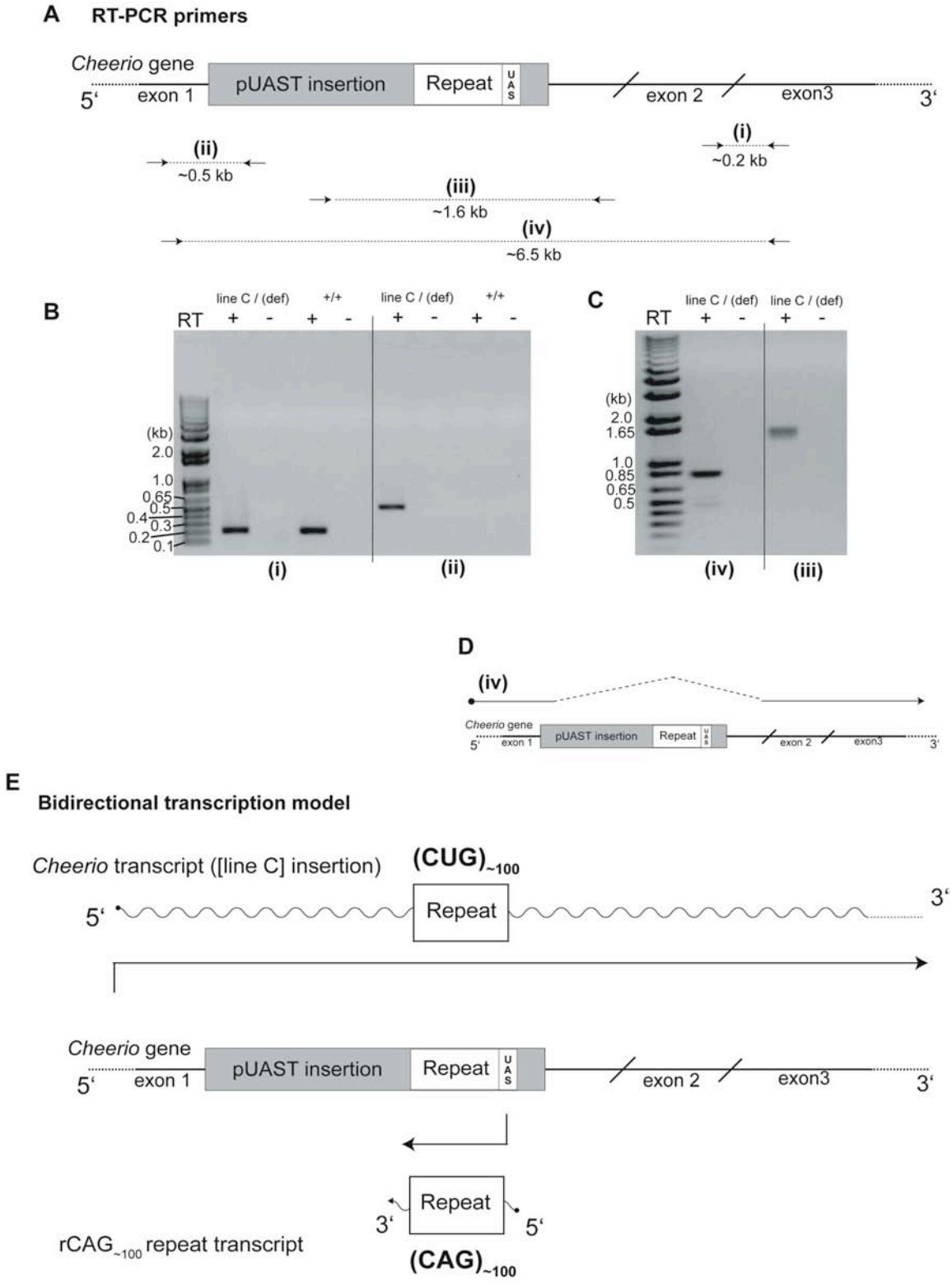


Figure 5.4
A, Schematic (not to scale) showing primers used to test transcription of the *rCAG$_{\sim 100}$ [line C]* insertion locus. (i): spanning the exon2/3 boundary, (ii): spanning the 5'insertion boundary, (iii) spanning the 3' insertion boundary including the repeat sequence (iv): spanning the entire insertion site. **B**, **C** RT-PCR products using primer pairs indicated in A. cDNA template is indicated above each lane as either *rCAG$_{\sim 100}$ [line C] / Df(3)Exel6176* deficiency, 'line C / (def)', or wild-type w^{1118}, '+/+'. Presence of reverse transcriptase to control for DNA contamination is indicated by +/- RT above lanes. **B**, (i) and (ii) both show correct size products that were confirmed by sequencing. **C**, (iii) produced a product of predicted size that was confirmed by sequencing to contain the repeat as well as part of the *cheerio* sequence, (iv) produces a smaller product than expected, revealed by sequencing to be a novel splice event. **D**, Schematic of the splice event identified with primer pair (iv). **E**, Model for bi-directional transcription of the repeat whereby GAL4 driven expression produces an rCAG transcript from the transgene insertion while the *cheerio* promoter produces a complementary rCUG transcript.

Together these results confirm that a CUG repeat-containing RNA is transcribed from the *cheerio* endogenous promoter. While the exact nature of the transcript may be complex, with the repeat either part of a large 5' UTR or an intron, in each case the *cheerio* protein coding sequence is conserved. Human disease causing repeats (DM2, SCA10) are found within introns and therefore the latter case does not preclude a role in pathogenesis [74, 126].

Endogenous transcription of the repeat transgene may provide an explanation for the 'held-out-wings' phenotype in progeny lacking a driver but homozygous for *rCAG$_{\sim 100}$ [line C]*, where two insertion alleles may give a greater level of *CUG-cheerio* fusion transcript (Figure 5.4 E). However, this mechanism does not fully explain the observed GAL4 mediated phenotypes. It appears that the phenotypes are dependent on both the presence of the *rCAG$_{\sim 100}$ [line C]* insertion, producing a *CUG-cheerio* transcript, and also GAL4-driven expression of a CAG transcript from the same locus (Figure 5.4 E). Expression of either repeat alone does not appear to account for the *rCAG$_{\sim 100}$ [line C]* locomotion or eye phenotypes and it was therefore hypothesised that the presence of both complementary repeat transcripts leads to a specific effect in this case.

Based on the available evidence, this mechanism explains both the dependence on GAL4 mediated transcription of the *rCAG$_{\sim 100}$* repeat, and the unique nature of the

rCAG$_{\sim 100}$ [line C] insertion site in its ability to produce a complementary rCUG transcript. While further examination may be required to be certain of this mechanism, studies conducted in parallel within our group provide further evidence for this effect [166]. In this case expression of complementary repeat RNA in *trans* led to phenotypes in the eye much stronger than expression of either repeat alone. Together this evidence supports a novel pathogenic mechanism involving complementary repeat RNA. Given increasing evidence for bi-directional transcription of repeats in disease, findings in this *Drosophila* model may be relevant to understanding pathways that lead to human pathology.

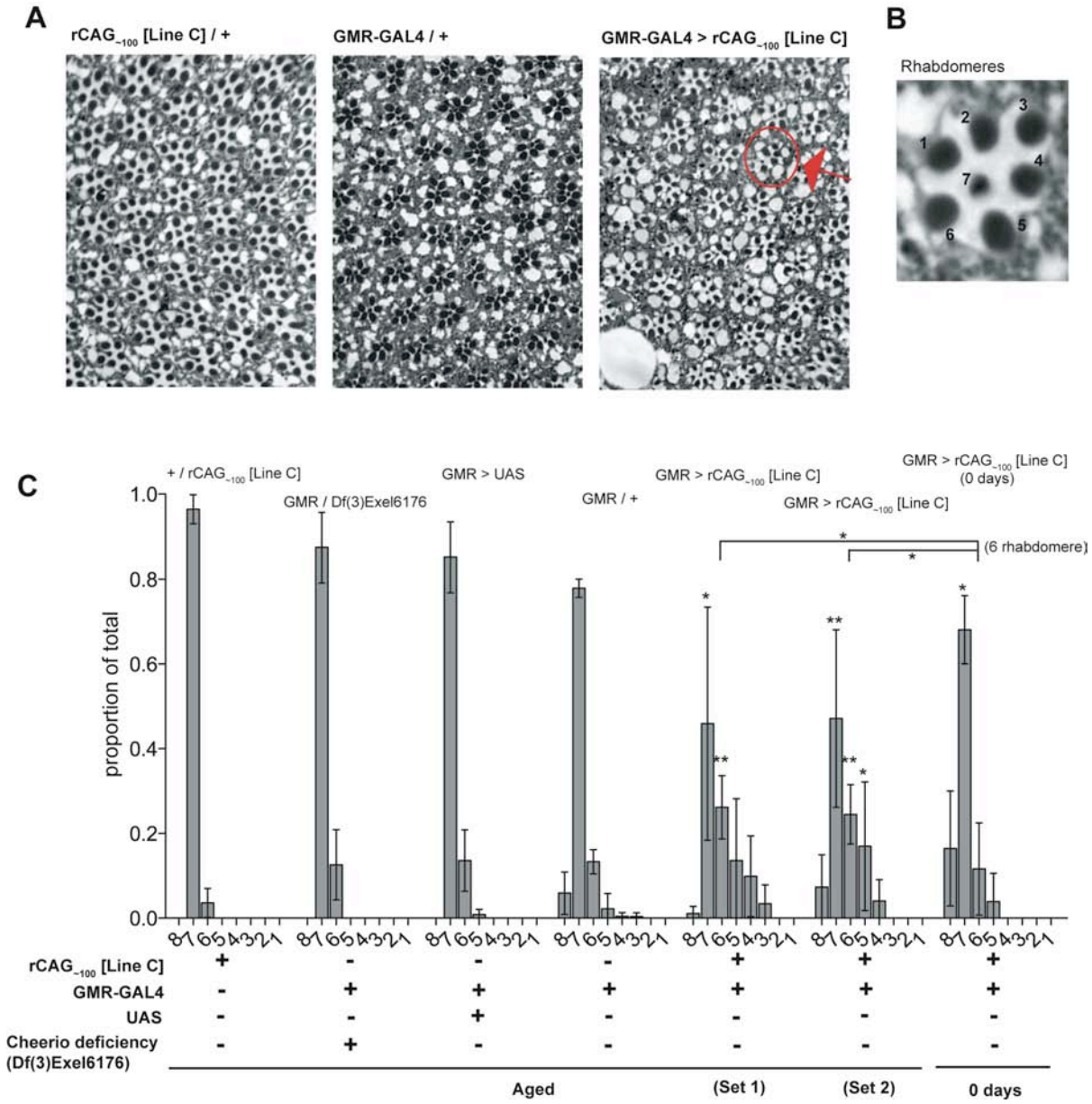
5.4 Ectopic expression of *rCAG$_{\sim 100}$ [line C]* leads to loss of photoreceptors

Based on examination of the *rCAG$_{\sim 100}$ [line C]* insertion locus, it was hypothesised that in this case bi-directional expression of the repeat leads to the production of complementary repeat RNA that is necessary to cause a phenotype. The basis for this effect is unclear and therefore further characterisation was undertaken. Experiments were aimed to examine these effects in more detail and attempt to identify assays that may be used to determine the pathways involved.

Initially the *rCAG$_{\sim 100}$ [line C]* eye phenotype was examined to determine whether the underlying cellular structure was disrupted, and whether this represented a degenerative process. Sections of aged flies expressing *rCAG$_{\sim 100}$ [line C]* were compared to controls consisting of *GMR-GAL4* > + and + / *rCAG$_{\sim 100}$ [line C]*. *GMR* > *rCAG$_{\sim 100}$ [line C]* eyes appeared more disrupted than either control, with one or more rhabdomeres missing from some ommatidia (Figure 5.5 A).

To quantify the level of disruption, sections were scored to determine the number of rhabdomeres in each ommatidium. Wild-type eyes have eight rhabdomeres that make up the ommatidial structure, however due to the arrangement of these in a tangential section, only seven are observed in one plane (Figure 5.5 B). Degeneration of the cells results in a lower number of rhabdomeres per ommatidium, which can be observed in sections. In some cases eight rhabdomeres can be observed when their normal arrangement is disrupted. To determine the extent of degeneration in *GMR* > *rCAG$_{\sim 100}$ [line C]* eyes each section was scored to determine the number of rhabdomeres per ommatidium. The number of rhabdomeres (1-8) was counted for each ommatidia and used to generate a tally of the number of ommatidia within each category (determined by rhabdomere count). The proportion within each category was determined for each section (minimum n=20 ommatidia) and a mean calculated across all sections (minimum n=4 images per genotype). Further control sections from *GMR-GAL4 > UAS* and GMR-*GAL4 / Df(3)Exel6176* aged eyes, as well as an independent set of aged *GMR-GAL4 > rCAG$_{\sim 100}$ [line C]* sections were included in this analysis to ensure that the method used was reproducible.

Figure 5.5 Ectopic expression of *rCAG~100 [line C]* in the eye leads to photoreceptor degeneration.



A, Tangential sections stained with toluidine blue from *+ / rCAG~100 [line C]* control eyes, *GMR-GAL4 /* + control eyes, and eyes expressing *rCAG~100 [line C]*. All are after ageing (see methods). Loss of rhabdomeres is observed in some ommatidia (circled, black arrow) in aged *rCAG~100 [line C]* expressing eyes. **B**, Enlarged image indicating seven rhabdomeres that is observed in control ommatidia. **C**, Results from scoring the number of rhabdomeres per ommatidium. The number of ommatidia with each rhabdomere count was tallied to obtain a proportion, then a mean proportion was obtained across a number of images (minimum n = 4 images, 20 ommatidia per image). Error bars show standard deviation. Student's t-tests were used to compare proportions for the same rhabdomere count category across genotypes. Significance is indicated where *<0.05, **<0.01, ***<0.001. Comparisons were made to *GMR-GAL4/+* unless otherwise indicated by bars.

Comparisons were made between genotypes using Student's t-test to determine whether there was a significant difference in the proportion of ommatidia within each category. Control lines showed few ommatidia with less than 7 rhabdomeres in aged + / *rCAG$_{\sim 100}$ [line C]* (mean proportion with 7 = 0.964) and GMR-GAL4 > UAS (mean proportion with 7 = 0.852) sections, indicative of little or no disruption, as expected (Figure 5.5 C). *GMR-GAL4 / Df(3)Exel6176* aged eyes (mean proportion with 7 = 0.874) were similar to controls, confirming that the phenotype is not caused by loss of *cheerio* function (Figure 5.5 C). *GMR-GAL4 /* + aged eyes (mean proportion with 7 = 0.779) showed significantly less ommatidia with 7 rhabdomeres than + / *rCAG$_{\sim 100}$ [line C]* controls ($p<0.001$), possibly due to a mild effect of GAL4 protein in this case [210] (Figure 5.6 C). The *GMR-GAL4/+* line was therefore used to make comparisons to *rCAG$_{\sim 100}$ [line C]* expressing eyes to ensure that any effect was not overestimated due to UAS-independent GAL4 mediated effects.

Compared to *GMR-GAL4 / +*, aged eyes expressing *rCAG$_{\sim 100}$ [line C]* showed a significant decrease in the proportion with 7 rhadomeres in aged set 1 (mean proportion = 0.459, p=0.0177) and set 2 (mean proportion = 0.471, p=0.004). Consistent with a decrease in the 7 rhabdomere category, comparison to *GMR-GAL4 / +* amongst the 6 rhabdomere category (mean proportion = 0.133) revealed a significant increase in aged set 1 (mean proportion = 0.262, p=0.0028) and set 2 (mean proportion = 0.245, p=0.0033). The mean values for the 5 rhabdomere category also appeared to be increased in eyes expressing *rCAG$_{\sim 100}$ [line C]*, however this was only significant in the case of *GMR-GAL4 > rCAG$_{\sim 100}$ [line C]* (set 2) (mean proportion = 0.170) compared to *GMR-GAL4/+* (mean proportion = 0.022, p=0.039) possibly due to the amount of variation within this category. These results indicate that ectopic expression of *rCAG$_{\sim 100}$ [line C]* in the eye leads to loss of photoreceptor cells such that a lower proportion of ommatidia have the normal 7 rhabdomeres and subsequently a greater number have 6, or 5.

To examine whether this process involves loss of adult cells, rather then disrupted development of cells, 0 day old flies expressing *rCAG$_{\sim 100}$ [line C]* were also compared to aged *rCAG$_{\sim 100}$ [line C]* expressing eyes. Although rhabdomere loss did not appear as strong in this case as in aged sets, there was no statistically significant

difference within the 7 rhabdomere category (mean proportion = 0.681) when compared to both aged *GMR-GAL4 > rCAG$_{\sim 100}$ [line C]* sets. However, the 6 rhabdomere category for 0 day old flies (mean proportion = 0.116) was significantly reduced compared to aged set 1 (mean proportion = 0.262, p=0.0274) and set 2 (mean proportion = 0.245, p=0.0238), indicating a reduced level of degeneration. These results therefore provide evidence that degeneration in *rCAG$_{\sim 100}$ [line C]* expressing eyes occurs over the adult lifespan. Further studies using larger numbers, or an alternate method such as pseudopupil analysis [211] may be necessary to examine this effect in more detail. 0 day old *rCAG$_{\sim 100}$ [line C]* expressing eyes did appear to show some degeneration, with the 7 rhabdomere category reduced compared to the aged *GMR-GAL4* / + control (p=0.018), suggesting that some effects leading to reduced rhabdomere count may occur prior to adulthood.

Together these results indicate that bi-directional expression of the repeat transgene in *rCAG$_{\sim 100}$ [line C]* expressing flies is associated with photoreceptor degeneration in the eye. Compared to the mild external eye phenotype this provides a more robust and quantifiable measure of *rCAG$_{\sim 100}$ [line C]* effects.

5.5 Ubiquitous expression of *rCAG$_{\sim 100}$ [line C]* leads to reduced lifespan

To further examine whether bi-directional transcription from the *rCAG$_{\sim 100}$ [line C]* insertion leads to degeneration, adult lifespan was examined in *Drosophila* ubiquitously expressing the transgene. *rCAG$_{\sim 100}$ [line C]* was expressed using the ubiquitous *da-GAL4* driver, and compared to controls consisting of flies with *da-GAL4* alone or *rCAG$_{\sim 100}$ [line C]* alone. In this case no locomotion defects were observed.

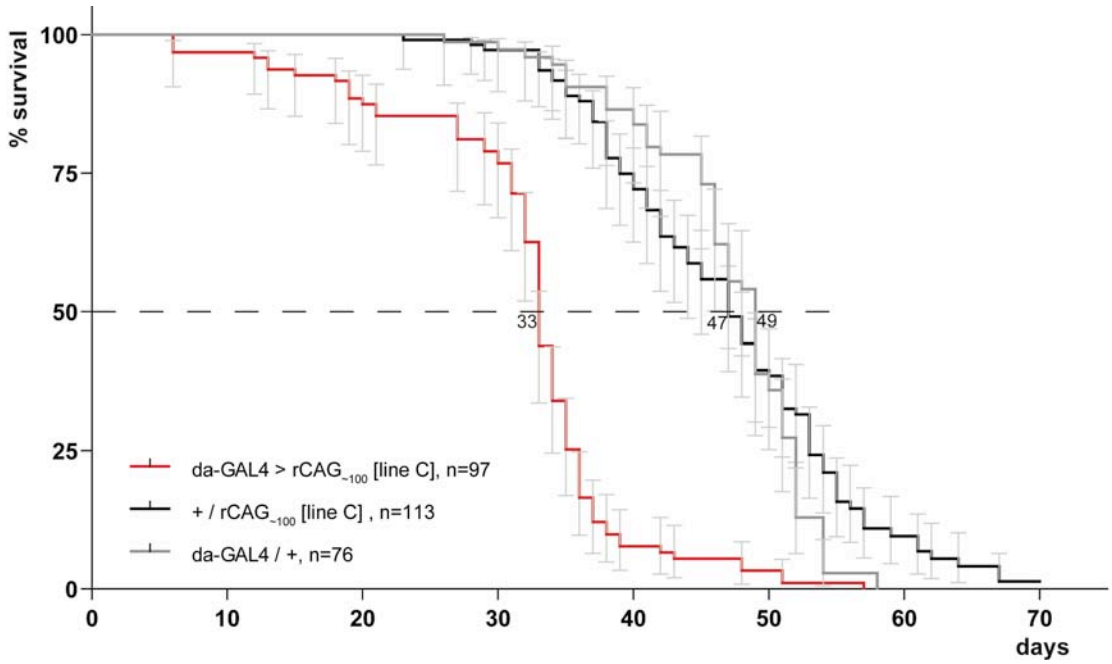
Progeny were aged at 25ºC, scoring deaths or escapees every 2 days. Data was used to plot survival curves for each genotype, which were compared to determine any statistical differences (Figure 5.6). Flies expressing *rCAG$_{\sim 100}$ [line C]* via da-GAL4 showed a significant reduction in lifespan (median age 33 days) compared to da-GAL4 / + controls (median age, 49 days, $p<0.0001$, log-rank test) and + / *rCAG$_{\sim 100}$ [line C]* controls (47 days, $p<0.0001$) (Figure 5.6). No significant difference was observed when comparing controls ($p=0.2082$).

Over-expression of *cheerio* using a tetracycline inducible promotor has previously been shown to increase lifespan in *Drosophila*, although the basis for this is unknown [212]. In contrast, the current study indicates a significant decrease in lifespan with *rCAG$_{\sim 100}$ [line C]*, further supporting that this effect is not due to the previously observed increase in *cheerio* levels. Together these results confirm, using an independent driver (*da-GAL4*), that bi-directional expression from *rCAG$_{\sim 100}$ [line C]* leads to a degenerative effect in adult flies.

One limitation when interpreting this result is that the effect of ubiquitous CAG RNA expression on adult lifespan has not been extensively characterised. In Chapter 4, experiments report that ubiquitously expressing much higher levels of each RNA repeat (four copies) results in reduced adult viability and disruption to tergite development. While expressing four copies of the *rCAG$_{\sim 100}$* construct still resulted in viable adult progeny, lifespan analysis was not undertaken in this case due to technical limitations in obtaining sufficient progeny, and therefore it is not possible to rule out a contribution of CAG RNA mediated effects in the reduction of lifespan observed with *rCAG$_{\sim 100}$ [line C]*. However, no tergite disruption was observed in *da-*

GAL4 > rCAG$_{\sim 100}$ [line C] progeny suggesting that different effects may be involved. In further support of this, *rCAG$_{\sim 100}$ [line C]* expression leads to mild disorganisation when expressed in the eye, and locomotion defects when expressed in neurons (Figure 5.1) while expression of four copies of independent *rCAG$_{\sim 100}$* lines gives no phenotype in these tissues [166].

Figure 5.6 Ubiquitous expression of *rCAG$_{\sim 100}$ [line C]* leads to a reduction in lifespan.



Survival curves comparing percentage survival over time for *da-GAL4 > rCAG$_{\sim 100}$ [line C]* (red), *+ / rCAG$_{\sim 100}$ [line C]* (black), and *da-GAL4 / +* flies (grey). Error bars indicate 95% confidence interval. Median age (50% survival) is indicated for each genotype alongside graph. Population size for each is indicated on graph.

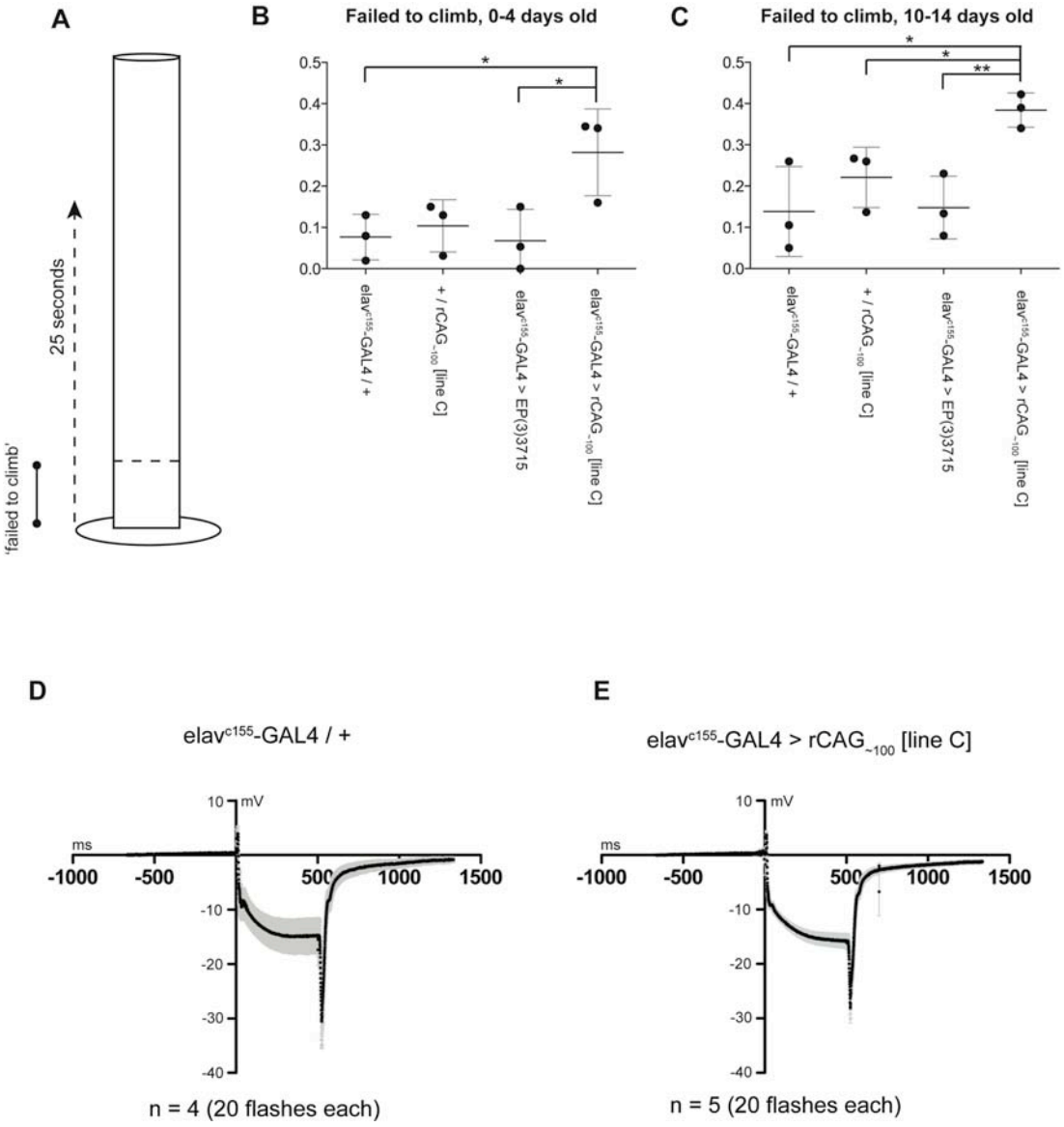
5.6 Pan-neuronal expression of *rCAG$_{\sim 100}$ [line C]* leads to neuronal defects

Previous experiments indicate that pan-neuronal expression of *rCAG$_{\sim 100}$ [line C]* led to locomotion defects indicative of neuronal dysfunction. This effect was quantified under heat stress, however this assay is time consuming and not ideal for making comparisons at different ages. Therefore, alternative methods were investigated to examine the effect of *rCAG$_{\sim 100}$ [line C]* on *Drosophila* neurons..

Previous preliminary work indicated that neuronal expression of *rCAG$_{\sim 100}$ [line C + D]* led to a reduction in adult lifespan [166]. Subsequent attempts in this study were unable to reproduce this finding. Progeny expressing *rCAG$_{\sim 100}$ [line C]* with *elav^{C155}-GAL4* showed no reduction in lifespan compared to controls in this case (data not shown). Lifespan analysis may be complicated by the locomotion defect in these flies, where an increased risk of falling and getting stuck in food may lead to an apparent reduction in lifespan. Experiments in this study were undertaken such that vials were stored horizontally to reduce this effect, and this may provide a simple explanation for the difference.

Climbing assays are an established method used to measure the propensity of *Drosophila* to climb upwards (negative geotaxis) when knocked to the bottom of a vessel [213]. A reduction in this ability has been used previously as a measure of neuronal dysfunction [132], and allows comparison of multiple groups at different ages as flies are not harmed during the assay. Multiple age matched populations (n=3 groups, n=20 each initial population) were tested for climbing ability at 0-4 days and 10-14 days old, by scoring flies as ‘failed to climb’, based on whether a certain height was reached within 25 seconds of being knocked to the bottom of the vessel (Figure 5.7 A). Expression of *rCAG$_{\sim 100}$ [line C]* using the pan-neuronal *elav^{c155}-GAL4* driver resulted in a mean proportion of flies that failed to climb of 0.384 at 0-4 days of age (Figure 5.7 B). This was significantly greater than *elav^{c155}-GAL4* / + (mean proportion 0.0767, p=0.0405) and greater than + / *rCAG$_{\sim 100}$ [line C]*, however this was not within the significance cut-off of 0.05 (mean proportion 0.104, p=0.0665) (Figure 5.7 B).

Figure 5.7 Pan-neuronal expression of *rCAG~100 [line C]* leads to a reduction in climbing ability.



A, Schematic of the assay used to determine climbing ability. After being knocked to the bottom of a standard measuring cylinder flies that failed to reach the 27mm point within 25 seconds were scored as 'failed to climb'. **B, C** Results for genotypes *elav^{c155}-GAL4 / +* , *+ / rCAG~100 [line C]*, *elav^{c155}-GAL4 > EP(3)3715* (*cheerio* overexpression) or *elav^{c155}-GAL4 > rCAG~100 [line C]*. Results from 3 biological replicates are shown with each represented by a black dot. Mean and standard deviation for each population is shown in grey. Significance based on Student's t-test is indicated where *p<0.05, **p<0.01. **B**, Populations at 0-4 days old. **C**, The same populations at 10-14 days old. **D, E**, Electroretinograms (ERGs) measuring photoreceptor response to light for 7-10 day old flies of genotype **D**, *elav^{c155}-GAL4 / +* and **E**, *elav^{c155}-GAL4 > rCAG~100 [line C]*. Each chart shows the mean and SEM based on the response for 4 different flies per genotype, where a single flies response is based on the average of 20 light flashes. Light flash occurs from 0-500ms in each case.

Climbing assays were also used to examine the possibility that the previously observed increase in *cheerio* transcript levels (Figure 5.3) leads to locomotion phenotypes, by comparing flies expressing *rCAG$_{\sim 100}$ [line C]* to those expressing *cheerio* from the the EP(3)3715 insertion that is also within the *cheerio* 5'UTR and enables GAL4-dependent overexpression of cheerio [203]. Flies expressing *rCAG$_{\sim 100}$ [line C]* showed a significantly greater proportion that failed to climb compared to those over-expressing *cheerio* from the EP(3)3715 insertion (mean proportion 0.0678, p=0.0463), confirming that the phenotype is not caused by the increase in *cheerio* levels (Figure 5.7 B).

After aging at 29ºC for 10-14 days the same populations showed a significant difference between the proportion that failed to climb for *elav^{c155}-GAL4 > rCAG$_{\sim 100}$ [line C]* (mean 0.384) compared to *elav^{c155}-GAL4/+* (mean 0.138, p=0.0217), *rCAG$_{\sim 100}$ [line C] / +* (mean 0.221, p=0.0284) and *elav^{c155}-GAL4>EP(3)3715* (mean 0.148, p=0.0091) (Figure 5.8 C). Although a greater proportion of aged flies with expression of *rCAG$_{\sim 100}$ [line C]* failed to climb, this was also observed in controls. The difference between *elav^{c155}-GAL >rCAG$_{\sim 100}$ [line C]* flies at 0-4 and 10-14 days was not significant (p=0.1912).

Therefore expression of *rCAG$_{\sim 100}$ [line C]* leads to a greater proportion of flies that fail to climb. This is not a strong effect, but is statistically significant compared to controls. A likely explanation for this is that 'knocking' the vessel at the start of each climbing assay induces the locomotion defect in some flies, and subsequently leads to a delay before they begin climbing. The proportion of flies exhibiting this effect appears to be fairly consistent across replicates, and this method can therefore be considered a reliable way to quantify the locomotion phenotype.

This experiment did not indicate a significant increase in the severity of the phenotype with age. This method may be limited in that an increase in the severity of the phenotype can only be measured as a greater proportion of the population that is affected. Alternatively, increased severity may involve a more severe phenotype in already affected individuals. This is not measured in this case, and would require a different methodology.

These results indicate that the locomotion phenotype is clearly present in newly eclosed adults. One explanation for this observation may be that the locomotion phenotype represents some form of dysfunction that therefore may be related to a cause rather then an effect of degeneration. A range of assays to measure neuronal function have been developed for *Drosophila*. One well established method is the electroretinogram (ERG) that involves measuring the electrical potential at a site within the eye to measure the activation of photoreceptor neurons in response to light. In this way the photoreceptors can be used as a model system to examine the effect of ectopic expression of *rCAG$_{\sim 100}$ [line C]* on neuronal function [214, 215].

Attempts were therefore made to examine photoreceptor function when *rCAG$_{\sim 100}$ [line C]* was expressed in all neurons with *elav^{C155}-GAL4*. ERGs were performed on a number of flies expressing *rCAG$_{\sim 100}$ [line C]* and *elav^{c155}-GAL4 /* + control flies. In both cases, flies gave a typical ERG response with a small peak at light on, followed by a slower depolarisation, then a quick peak at light off, followed by a return to normal potential [214, 215]. This indicates that there is no major perturbation of photoreceptor function in flies expressing *rCAG$_{\sim 100}$ [line C]* in all neurons. It may be possible that more subtle alterations occur that are not detected with this method, or that photoreceptor neurons are less susceptible than neurons in the central brain. However, the ability to directly examine function in the central brain in *Drosophila* is technically limited and therefore this was not investigated further in this study.

5.7 Chapter discussion

Results in this chapter identify a specific insertion, *rCAG$_{\sim 100}$ [line C],* that produces dominant phenotypes when ectopically expressed in different cell types. Phenotypes include neuronal dysfunction and reduced lifespan and thus are relevant to studies human disease. Expression of multiple copies of *rCAG$_{\sim 100}$* from independent insertion sites did not replicate the phenotype seen with *rCAG$_{\sim 100}$ [line C],* indicating an insertion specific effect. However, the presence of the insertion alone was also insufficient to cause a phenotype, with all phenotypes being dependent on GAL4 driven expression of *rCAG$_{\sim 100}$ [line C]*. The insertion was identified at a site near the *cheerio* gene, however further experiments concluded that expression of the transgene does not reduce *cheerio* levels. Genetically altering *cheerio* levels did not replicate the eye or neuronal phenotypes, providing evidence that phenotypes are caused by a repeat-mediated effect rather than disruption of *cheerio* function.

RT-PCR experiments identified a transcript containing part of the *cheerio* sequence and the complementary CUG repeat, indicating that the repeat transgene may be transcribed in the opposite direction from the *cheerio* promoter. This led to the hypothesis that in this case bi-directional transcription produces complementary repeat RNAs that contribute to the observed phenotypes. Recent findings suggest that bi-directional transcription may play a role in human disease, with antisense transcripts identified for several disease causing repeats [119, 124, 151, 153-155]. Hence, the observations in *rCAG$_{\sim 100}$ [line C]* progeny may be relevant to repeat mediated pathology in human disease, as this line provides a model for bi-directional transcription of an expanded repeat. The observed phenotypes may be used to examine the effect of altering candidate modifier genes to investigate the pathways involved. This approach is utilised in the following chapter.

While results suggest that the phenotypes observed are dependent on repeat expression rather than altered *cheerio* levels, it remains a possibility that the presence of the insertion within the *cheerio* locus contributes to the phenotypes. Interestingly, mutations in *Filamin-A*, the human ortholog of *cheerio*, result in periventricular heterotopia, a condition involving disruption to neuronal development and subsequent clinical symptoms including seizures [216, 217]. *Drosophila cheerio* mutants do not

appear to replicate this [203, 207, 218], however it is not possible to entirely rule out a contribution from *cheerio* to the findings in this thesis. One likely contribution may be that in the case of the *rCAG$_{\sim 100}$ [line C]* insertion, the complementary CUG transcript is controlled by the endogenous *cheerio* promoter rather than the UAS sequences. Thus, bi-directional transcription only occurs in cells that have overlapping *cheerio* and GAL4 driven expression, and at times when both promoters are active. For example, while GAL4 expression from *elav^{C155}-GAL4* occurs pan-neuronally only a subset of neurons may express both transcripts. Furthermore, the relative abundance of each transcript may differ between cells based on regulation of the *cheerio* promotor. Another possibility is that the presence of the CUG repeat within a *cheerio* transcript has some effect on *cheerio* function, through mechanisms that are as yet unclear.

Further work from our group has examined the effect of complementary repeat expression, where each repeat is expressed from UAS sites at different loci. While not examining bi-directional expression from the same locus directly, this system enables GAL4 driven control of each transcript and avoids the limitations associated with the *cheerio* insertion line. These experiments indicate that expression of complementary repeat RNA leads to strong toxicity in the eye that is dependent on Dicer activity, indicating the involvement of a dsRNA mediated mechanism [166]. Together with the results observed in *rCAG$_{\sim 100}$ [line C],* this supports the hypothesis that a novel mechanism involving bi-directional transcription, leading to the expression of complementary repeat sequences, may be involved in the pathogenesis of human expanded repeat disease. *rCAG$_{\sim 100}$ [line C]* and expression of complementary repeat RNA from different loci therefore provide two systems in which to examine this novel mechanism, and determine its relevance to human disease.

CHAPTER 6 : Comparison of pathways responsible for double-stranded and hairpin-forming repeat RNA-mediated pathology

Introduction

Experiments described in Chapter 5 identified a specific transgene insertion, *rCAG$_{\sim 100}$ [line C],* that is inserted at the *cheerio* locus. When CAG RNA was expressed from *rCAG$_{\sim 100}$ [line C],* flies showed dominant phenotypes suggestive of degeneration and neuronal dysfunction. The presence of the insertion without a driver, or alterations in *cheerio* levels were not sufficient to cause phenotypes. Analysis of the insertion site showed that the repeat was transcribed in the opposite direction from the endogenous *cheerio* promoter, leading to the production of a *cheerio* transcript containing a CUG repeat. Each phenotype was dependent on the presence of both the rCUG containing cheerio transcript, and GAL4/UAS driven rCAG expression, leading to the hypothesis that a mechanism involving expression of complementary expanded repeat transcripts leads to dominant effects in this line.

Further support for this hypothesis is provided by independent results from our lab showing that GAL4/UAS mediated ectopic expression of complementary *rCUG$_{\sim 100}$* and *rCAG$_{\sim 100}$* constructs from different loci strongly disrupts the *Drosophila* eye [166]. Expression of complementary repeats within neurons leads to degeneration, as indicated by an age dependent failure to climb [166]. Further experiments showed that altering Dicer-2 levels modifies eye phenotypes, while deep-sequencing of small RNAs identified CAG 21-mers in the adult *Drosophila* brain. These results suggest that toxicity is mediated by the formation of double-stranded repeat RNA that is processed by Dicer pathways [166]. These experiments showed a change in the abundance of specific miRNAs, suggesting that alterations to the miRNA profile may play a role in complementary repeat RNA-mediated phenotypes. In addition, alterations in Dicer-1 levels were also shown to modify complementary repeat RNA mediated eye phenotypes, to a lesser extent than Dicer-2 (S. Samaraweera, unpublished). Dicer-1 processing is proposed to primarily regulate miRNA

processing, while Dicer-2 is proposed to regulate siRNA processing [187]. Therefore, modification of the complementary repeat expression phenotype by altering either Dicer protein indicates that multiple aspects of double-stranded RNA processing may play a role in pathology. These results support the hypothesis that complementary repeat RNA is able to form dsRNA that is pathogenic, and indicates a mechanism involving processing by Dicer enzymes.

Evidence therefore exists that expanded repeat RNA in *Drosophila* can lead to pathogenesis as either single-stranded hairpin-forming RNA (tergite disruption – Chapter 4), bi-directional complementary transcripts from a single locus (*rCAG$_{\sim 100}$ [line C]*), or as complementary repeats expressed from different loci [166]. Understanding the relative contributions of each of these toxic species to pathology will be important to determining the basis for human disease. Using the *Drosophila* models described in this thesis will enable examination of the pathways involved in each case.

The experiments within this chapter were designed to compare different repeat mediated effects, and examine whether common pathways may be involved in each case. Initially comparisons were made between *rCAG$_{\sim 100}$ [line C]* bi-directional repeat-mediated effects and those caused by expression of complementary repeats from different loci. Following from this, experiments were undertaken to examine whether double-stranded RNA processing pathways can genetically modify single stranded RNA mediated phenotypes. Evidence for an effect in this case may indicate that common pathways are involved in each type of pathogenesis.

6.1 Comparison of neuronal bi-directional repeat expression from *rCAG$_{\sim 100}$ [line C],* and complementary repeat expression from different loci

Previous experiments indicate that pan-neuronal expression of *rCAG$_{\sim 100}$ [line C]* using *elav^{c155}-GAL4* leads to locomotion defects that can be measured as a reduction in climbing ability. Analysis indicated that this phenotype is dependent on bi-directional expression of complementary repeats from the *cheerio* insertion locus (Figure 5.7 B, C, Figure 6.1 A). Experiments were undertaken to determine whether this phenotype also occurs when complementary rCAG and rCUG RNA transcripts are expressed from different loci, both under UAS control (Figure 6.1 B).

Three independent lines, each carrying a single copy of each of the *rCAG$_{\sim 100}$* and *rCUG$_{\sim 100}$* transgenes (‘*rCAG$_{\sim 100}$.rCUG$_{\sim 100}$*’) at different genomic locations (Materials and Methods), were crossed to *elav^{c155}-GAL4* to give neuronal expression of complementary repeat RNA. Crosses were performed at 29ºC to replicate conditions in which the *rCAG$_{\sim 100}$ [line C]* climbing defect was previously observed. Two complementary repeat expression lines, each expressing transgenes from independent insertion sites, gave either complete male lethality, or severely reduced viability (each total population n > 50, including balancer progeny). A third line produced viable male progeny and was therefore used to compare to *elav^{c155}-GAL4 > rCAG$_{\sim 100}$ [line C]* males.

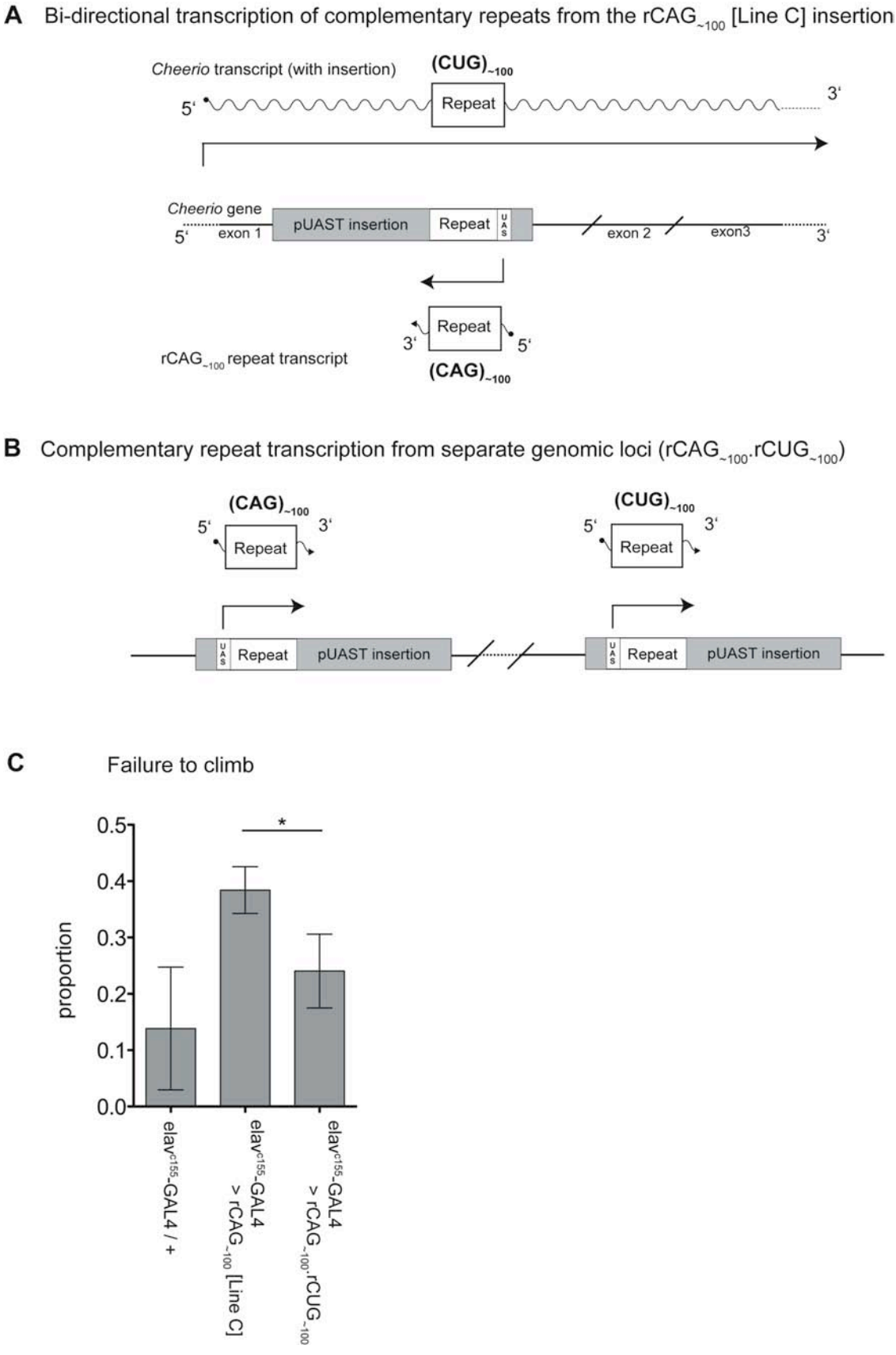
Initial observations indicated that *elav^{c155}-GAL4 > rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* progeny did not show a locomotion defect. This was confirmed in assays measuring failure to climb, where results indicate normal climbing ability in the complementary repeat expression line, with no significant difference between *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* expressing progeny and the *elav^{c155}-GAL4 /* + control at 10-14 days old. Consistent with this a significant difference was observed when *elav^{c155}-GAL4 > rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* progeny were compared to *elav^{c155}-GAL4 > rCAG$_{\sim 100}$ [line C]* progeny that do show an increased failure to climb ($p=0.0326$) (Figure 6.1 C). Therefore, while two out of three complementary repeat expressing lines gave lethality, the third was viable but showed no locomotion defect. In contrast bi-directional expression from *rCAG$_{\sim 100}$ [line C]* gave viable progeny that showed locomotion defects as measured by an increased failure to climb.

These observations may be explained by a situation where bi-directional expression from *rCAG$_{\sim 100}$ [line C]* leads to distinct cellular perturbation compared to complementary expression of *rCUG$_{\sim 100}$* and *rCAG$_{\sim 100}$* RNA, resulting in specific locomotion defects. This could be caused by cellular specificity caused by the endogenous *cheerio* promoter. Alternatively, lethality in complementary repeat expressing lines may represent a more severe perturbation of the same pathway that leads to locomotion defects in *rCAG$_{\sim 100}$ [line C]*. In this case, the *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* complementary repeat expression line that is viable may not be perturbed sufficiently to show a locomotion phenotype, possibly due to a lower level of transgene expression. However, in support of a specific *rCAG$_{\sim 100}$ [line C]* effect, expression of two copies of each complementary repeat ('2x*rCAG$_{\sim 100}$*.2x*rCUG$_{\sim 100}$*') using an alternate neuronal driver (*elavII-GAL4*) still gives viable progeny [166]. In this case, flies show an age-dependent decrease in climbing ability, but do not show early locomotion defects like those observed with *rCAG$_{\sim 100}$ [line C]* [166]. Nonetheless, further examination will be required to determine the basis for distinct observations in each case.

Figure 6.1

A, Schematic of bi-directional repeat expression from *rCAG$_{\sim 100}$ [line C]*, where CAG RNA is under UAS control while complementary CUG RNA is transcribed as part of the endogenous *cheerio* transcript. **B**, Schematic of complementary repeat expression from different loci. In this case each repeat is transcribed from a separate location, under UAS control. **C**, Climbing assay measuring the proportion of flies failing to climb with each complementary repeat expression system using the pan-neuronal *elav^{c155}-GAL4* driver. Graph shows mean and standard deviation for n=3 populations per genotype, n=20 flies initially per population. Genotype is indicated under the graph. Significance was determined using Student's t-test, * = $p<0.05$.

Figure 6.1 Comparison of bi-directional and complementary repeat expression in neurons.



6.2 Altering Dicer-2 levels does not significantly alter *rCAG$_{\sim 100}$ [line C]* photoreceptor degeneration

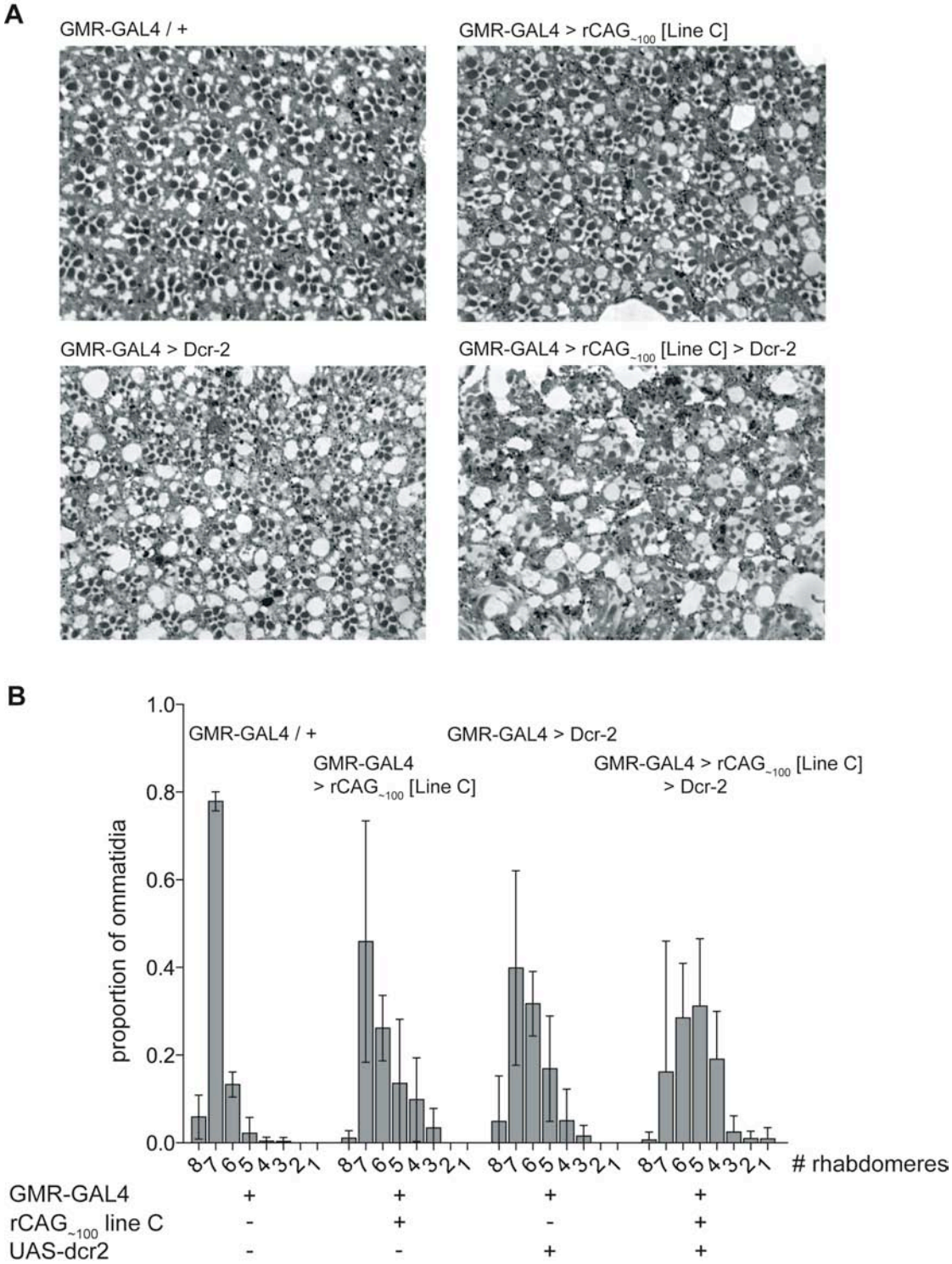
The *rCAG$_{\sim 100}$ [line C]* locomotion phenotype was not observed when complementary repeats were expressed neuronally, from different genomic loci. However, this situation did lead to lethality and shared pathogenic pathway may still be involved in each case. Ectopic expression of complementary repeats from different loci leads to disruption when expressed in the eye that is modified by altering levels of Dicer-2 [166]. Therefore, experiments were undertaken to examine if this is also the case with *rCAG$_{\sim 100}$ [line C]*. Previously the underlying structure of the eye was shown to be disrupted in aged flies expressing *rCAG$_{\sim 100}$ [line C]* (Chapter 5). Given the mild external eye phenotype with *rCAG$_{\sim 100}$ [line C]* examining underlying degeneration may offer a more informative read out of the effects of altering *Dicer-2* (*Dcr-2*) levels. Therefore, sections of aged eyes co-expressing *rCAG$_{\sim 100}$ [line C]* and *UAS-Dcr-2*, or either transgene alone were examined. It was predicted that an enhancement of this phenotype could be detected as a greater loss of rhabdomeres in tangential sections of the eye compared to controls. The approach of examining overexpression was taken rather then examining loss of *Dcr-2* function as the expected suppression in this case may be too subtle to detect.

Sections of eyes from flies co-expressing *rCAG$_{\sim 100}$ [line C]* and *UAS-Dcr2* (Figure 6.2 A) appeared more disrupted than controls, with more missing rhabdomeres and the presence of larger vacuoles in the tissue, than in sections of those expressing *rCAG$_{\sim 100}$ [line C]* alone (Figure 6.2 A). However, compared to the *GMR-GAL4 /* + control, expression of *UAS-Dcr2* alone also appeared to cause disruption, comparable to that observed with expression of *rCAG$_{\sim 100}$ [line C]* alone (Figure 6.2 A). This observation was confirmed when the number of rhabdomeres in each ommatidium was scored, indicating that a significant reduction of the proportion with 7 rhabdomeres ($p=0.0016$) and a subsequent increase in the proportion with 6 ($p=0.0001$) was observed with expression of *UAS-Dcr-2* compared to the *GMR-GAL4 /* + control (Figure 6.2 B). This was a similar reduction to that previously observed with expression of *rCAG$_{\sim 100}$ [line C],* with no significant difference between the proportions in each category when comparing expression of *rCAG$_{\sim 100}$ [line C]* to

expression of *UAS-Dcr-2* alone (Figure 6.2 B). Compared to expression of either construct alone, co-expression of *rCAG$_{\sim 100}$ [line C]* and *UAS-Dcr-2* resulted in a lower proportion of ommatidia with 7 rhabdomeres, a similar proportion with 6 rhabdomeres and a slightly increased proportion with 5 rhabdomeres (Figure 6.2 B). However, these differences were not statistically significant, with all $p>0.1$ except in the 5 rhabdomere category where results were closer to the significance threshold ($p<0.05$) compared to expression of *rCAG$_{\sim 100}$ [line C]* ($p=0.0589$) or *UAS-Dcr-2* alone ($p=0.0765$).

While larger scale analysis may determine the true relevance of this effect, this would be complicated by disruption due to the *UAS-Dcr-2* construct alone, and thus the possibility that an enhanced phenotype is the result of an additive effect rather than a specific interaction. These results therefore provide no significant evidence for modification of the *rCAG$_{\sim 100}$ [line C]* eye phenotype by altering *Dicer-2* levels. This may indicate that the mechanism causing disruption in this case is not reliant on Dicer-2 processing. Alternatively, the difference may be explained by a relatively low level of available double-stranded repeat RNA with bi-directional expression from *rCAG$_{\sim 100}$ [line C]* compared to expression with *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$*.

Figure 6.2 Increased Dcr-2 levels does not significantly modify *rCAG~100 [line C]* photoreceptor degeneration



A, Tangential sections through aged eyes showing the internal structure of the ommatidia. Expression of *rCAG~100 [line C]*, or *UAS-Dcr-2* causes disruption and loss of rhabdomeres, that appears to be worse when both are co-expressed. **B**, Quantification of rhabdomeres in tangential sections showing the proportion of ommatidia with a particular number of rhabdomeres. A reduction in the proportion with the normal seven indicates disruption. Each bar represents mean ± standard deviation. n≥4 images per genotype, 20 ommatidia per image. All flies were aged before sectioning (Materials and Methods).

6.3 Examining the role of Dicer processing pathways in hairpin RNA-mediated tergite phenotypes

Experiments described in Chapter 4 identified a phenotype caused by the expression of either CAG or CUG single stranded repeat RNA ubiquitously in adult *Drosophila*. In this case adult viability was reduced and surviving adults had a specific phenotype whereby the development of the tergite bands is disrupted, suggesting specific cells may be more susceptible to the effect. Previously, hairpin-forming single-stranded repeat RNA has been shown to be a substrate for processing by the Dicer proteins [150]. This raises the possibility that single stranded repeat-mediated pathogenesis may involve a similar mechanism to that identified for complementary repeat expression [166].

Initially, attempts were made to examine the effect of ubiquitous complementary repeat expression in adult flies. When two copies of each repeat construct were co-expressed (2x*rCAG*$_{\sim 100}$.2x*rCUG*$_{\sim 100}$), to give a total of four repeat transgene copies, complete adult lethality was observed in all of three independent lines tested. Ubiquitous co-expression of only one copy of each of rCAG$_{\sim 100}$ and rCUG$_{\sim 100}$ (*rCAG*$_{\sim 100}$.*rCUG*$_{\sim 100}$), and hence a reduced repeat dosage, still led to lethality in all independent lines tested (n=3). In comparison expression of four copies of either *rCAG*$_{\sim 100}$ or *rCUG*$_{\sim 100}$ leads to only some reduction in viability (Chapter 4).

These results suggest that ubiquitous expression of complementary repeat RNA is more toxic than expression of each RNA repeat sequence alone. This also prevents the direct comparison of effects on the adult abdominal tergites. However, it may be possible that hairpin RNA effects are mediated through a similar mechanism, albeit less efficiently, resulting in a milder effect. Given that complementary repeat expression eye phenotypes have been shown to be dependent on Dicer activity, experiments were set up to examine whether hairpin RNA-mediated tergite phenotypes could also be modified by altering Dicer levels.

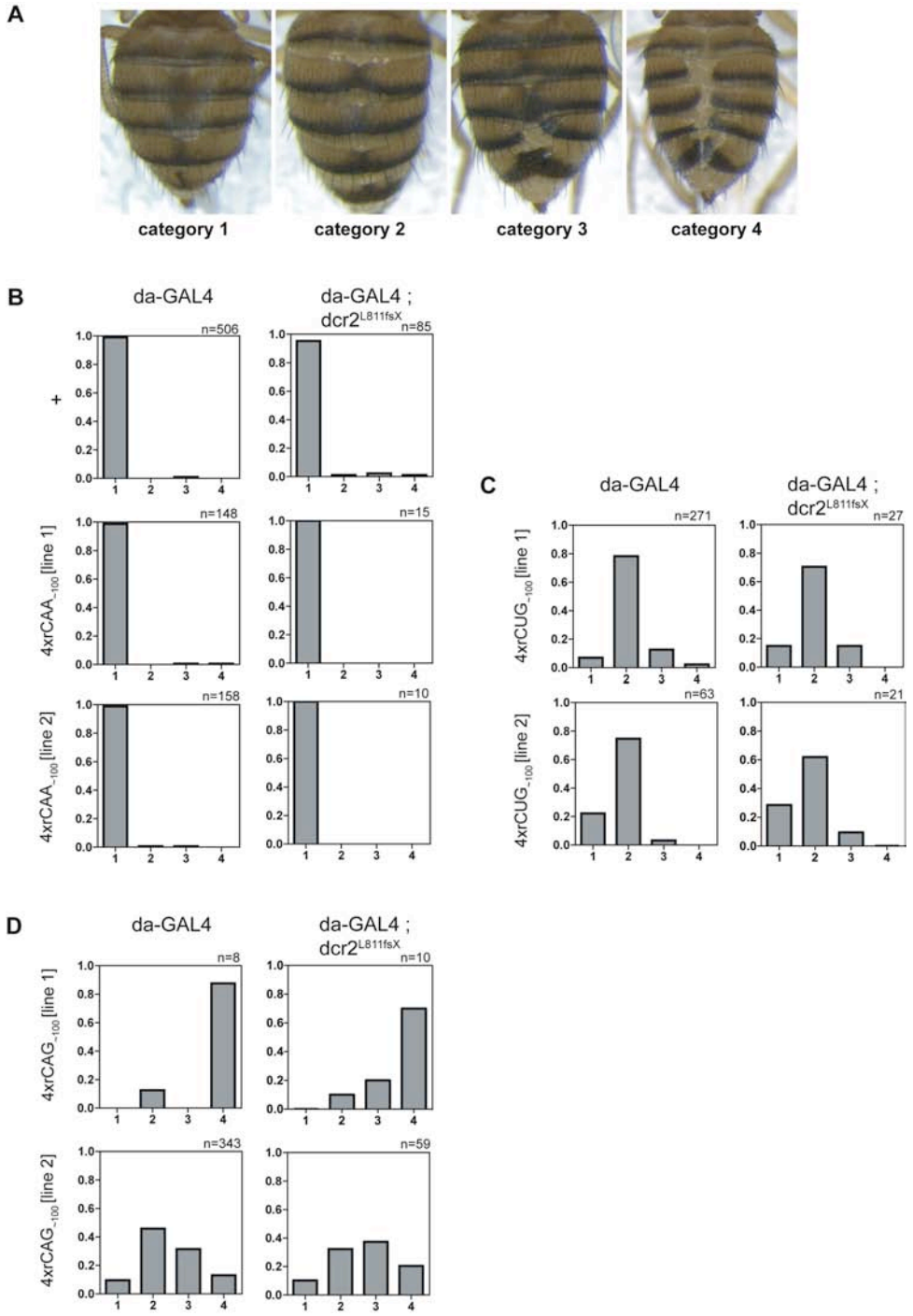
6.3.1 Dicer-2 modification of tergite phenotypes.

The tergite phenotype described in Chapter 4 provides a reproducible, quantitative read-out of hairpin RNA-mediated effects in *Drosophila*. Hence this was used to assess whether altering the level of Dicer proteins, by expressing repeats within a mutant background, is able to modify the severity of the phenotype. Modification may indicate a requirement for Dicer processing, and therefore provide evidence for a shared mechanism in hairpin and complementary repeat RNA mediated pathogenesis.

Initially altering the level of Dicer-2 was tested as this was shown to strongly modify complementary RNA mediated phenotypes [166]. Loss of one copy of Dicer-2 via introducing the $dcr2^{L811fsX}$ allele produces viable adults, however ubiquitous over-expression of *UAS-Dcr-2* via *da-GAL4* results in lethality before adulthood, thus preventing the analysis of over-expression effects on the tergite phenotype.

Each of the individual repeat constructs, $rCAG_{\sim 100}$, $rCUG_{\sim 100}$ and $rCAA_{\sim 100}$, and controls were ubiquitously expressed via *da-GAL4* in the presence of the $dcr2^{L811fsX}$ allele. If Dicer-2 function is necessary for the phenotype then this situation may result in suppression of the phenotype, as observed with complementary repeat expression [166]. Progeny were scored as previously to enable comparison of the distribution of phenotype severity within each genotype (Figure 6.3 A, Appendix 3.2). The *da-GAL4 /* + and both $4xrCAA_{\sim 100}$ control lines showed no phenotype, with almost all progeny showing wild-type tergites (Figure 6.3B). Both $4xrCUG_{\sim 100}$ lines showed a similar distribution of phenotype strength with or without $dcr2^{L811fsX}$ (Figure 6.3 C). As previously, analysis of line $4xrCAG_{\sim 100}$ *[line 1]* was complicated by reduced viability, although surviving progeny with $dcr2^{L811fsX}$ showed a strong phenotype (Figure 6.3 D). Progeny expressing $4xrCAG_{\sim 100}$ *[line 2]* showed a similar phenotype in each case, although there was a slight shift towards the stronger two phenotype categories when $dcr2^{L811fsX}$ was present (Figure 6.3 D).

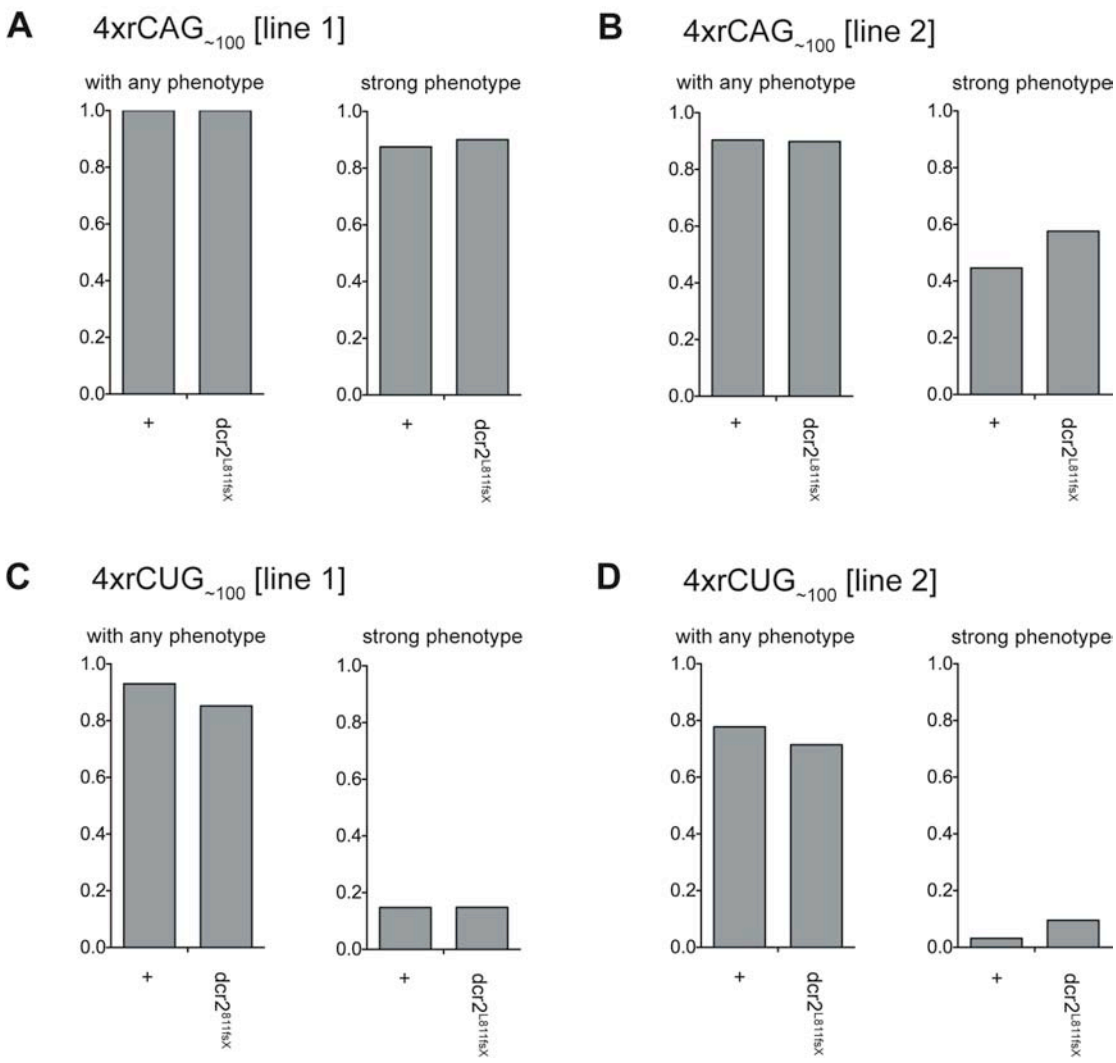
Figure 6.3 Population distribution of tergite phenotype severity with reduced Dicer-2 levels.



A, Categories used to score phenotype severity, as described previously. **B-D**, Charts show the proportion of total progeny for a particular genotype within each phenotype scoring category. In each case expression of the indicated repeat construct ubiquitously via *da-GAL4* in a wild-type background is shown on the left, while the *da-GAL4* expression with one copy of the $dcr2^{L811fsX}$ allele is shown on the right. **B**, Control lines with no repeat, or two independent $4xrCAA_{\sim 100}$ lines. **C**, Two independent $4xrCUG_{\sim 100}$ lines. **D**, Two independent $4xrCAG_{\sim 100}$ lines.

To examine whether changes were statistically significant, genotype comparisons were made for the total proportion of progeny with any phenotype, and the total proportion within the two most severe categories (a 'strong phenotype') (Figure 6.4). In each case no significant difference was observed when comparing the distribution between categories for groups expressing repeat constructs in the presence or absence of the *dcr2*^{L811fsX} allele (Fisher's exact test, Appendix 3.4).

Figure 6.4 Analysis of the effect of reducing Dicer-2 levels on the total phenotype proportion and proportion with a strong phenotype.



Charts show the proportion with any phenotype (in category 2, 3 or 4 as indicated previously) or the proportion with a 'strong phenotype' (category 3 or 4 only). Each chart compares the value for each genotype with or without one copy of the *dcr2*^{L811fsX} allele (indicated below chart). **A**, *4xrCAG~100 [line 1]*. **B**, *4xrCAG~100 [line 2]*. **C**, *4xrCUG~100 [line 1]*, **D**, *4xrCUG~100 [line 2]*.

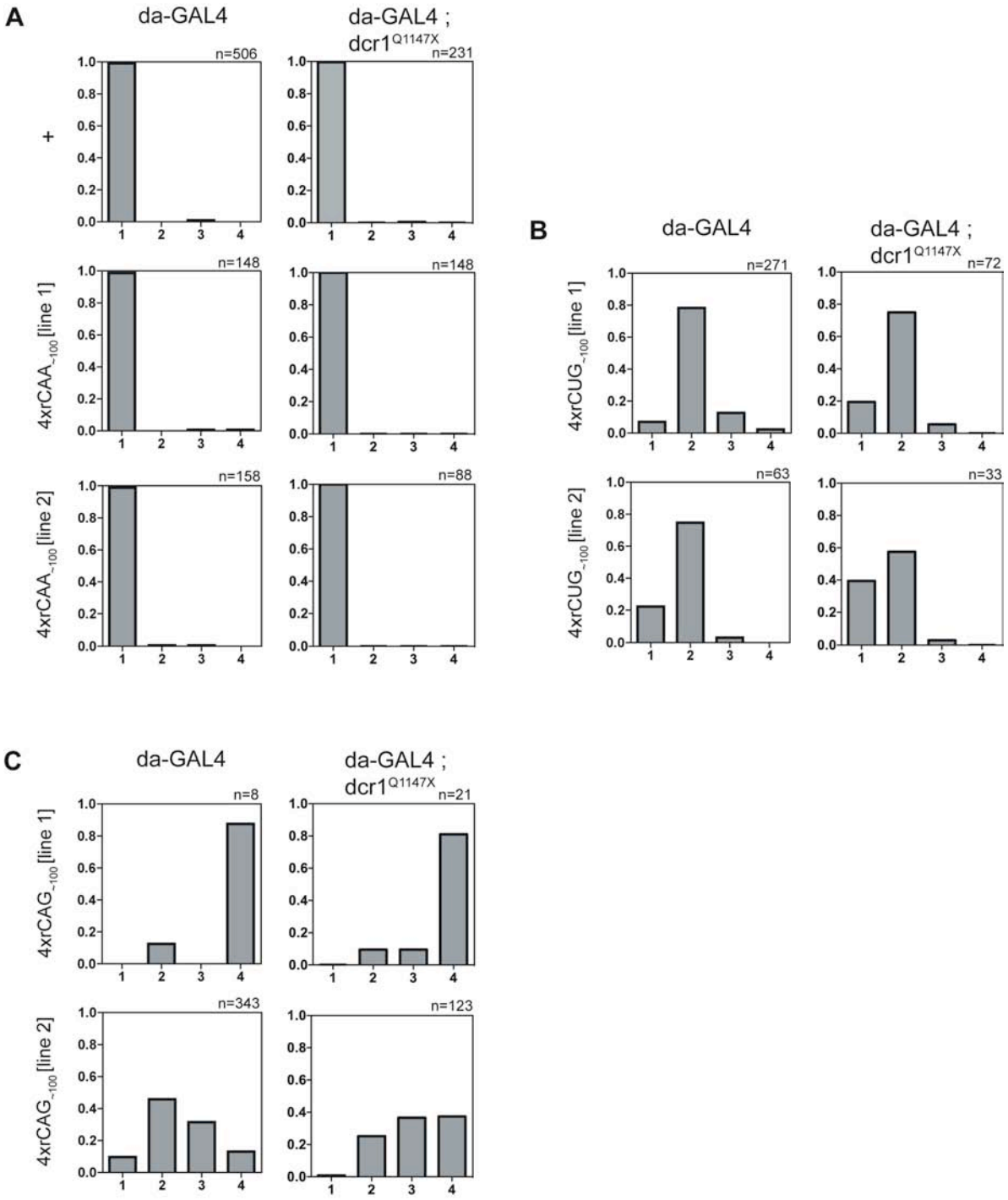
The results from these experiments support the conclusion that a reduction in Dicer-2 processing in not rate-limiting for *rCAG$_{\sim 100}$* or *rCUG$_{\sim 100}$* mediated tergite disruption in this system. This raises the possibility that tergite disruption occurs through a mechanism that is distinct from that involved in complementary repeat mediated pathogenesis. In further support of distinct pathways, *da-GAL4* driven expression of *rCAG$_{\sim 100}$ [line C],* which produces complementary repeat transcripts, gives viable adults that have a reduction in lifespan, but show no tergite phenotype (Figure 5.6).

6.3.2 Dicer-1 modification of tergite phenotypes.

Results from our group indicate that loss of Dicer-1 also leads to suppression of complementary repeat RNA-mediated eye phenotypes (S. Samaraweera, unpublished). In *Drosophila*, Dicer-1 and Dicer-2 have distinct roles in small RNA biogenesis and preferences for different RNA substrates, and therefore may have differential effects on hairpin RNA, and dsRNA [187]. Hence the ability of altered Dicer-1 levels to modify the tergite phenotypes was also tested. The effect of altering Dicer-1 levels on the tergite phenotype was examined in the same manner as for Dicer-2. Repeat constructs were ubiquitously expressed using *da-GAL4*, and Dicer-1 levels reduced by introducing one copy of the $dcr1^{Q1147X}$ allele. The distribution of phenotype severity was compared between those in a wild-type background, or with $dcr1^{Q1147X}$ (Figure 6.5, Appendix 3.3).

The *da-GAL4 /* + control and both $4xrCAA_{\sim 100}$ lines showed no phenotype with $dcr1^{Q1147X}$ (Figure 6.5 A). Both lines expressing $4xrCUG_{\sim 100}$ showed similar phenotypes in a wild-type and + / $dcr1^{Q1147X}$ mutant background, although the proportion with wild-type tergites (Category 1) did appear increased in the presence of $dcr1^{Q1147X}$ for both lines (Figure 6.5 B). As previously, numbers were limited for flies expressing $4xrCAG_{\sim 100}$ *[line 1]*, preventing meaningful statistical analysis, however surviving progeny still showed a strong phenotype (Figure 6.5 C). Progeny expressing $4xrCAG_{\sim 100}$ *[line 2]* appeared to show a stronger phenotype in the presence of $dcr1^{Q1147X}$ compared to a wild-type background, with the distribution of phenotypes shifted towards the stronger end of the scale (Figure 6.5 C).

Figure 6.5 Population distribution of tergite phenotype severity with reduced Dicer-1 levels.

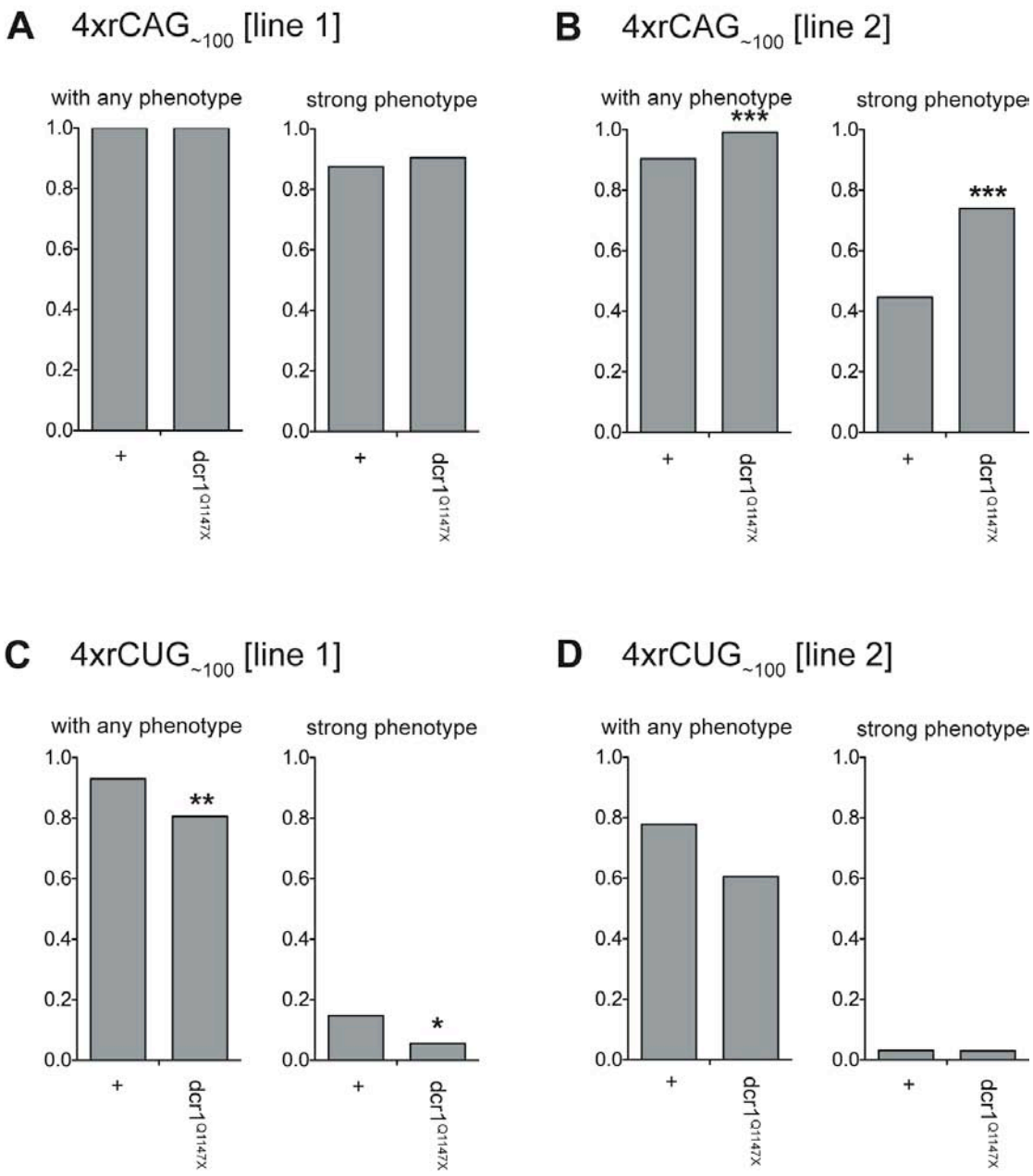


Charts show the proportion of total progeny for a particular genotype within each phenotype scoring category. In each case expression of the indicated repeat construct ubiquitously via *da-GAL4* in a wild-type background is shown on the left, while *da-GAL4* expression with one copy of the *dcr1*^{Q1147X} allele is shown on the right. **A**, Control lines with no repeat, or 2 independent 4xrCAA$_{\sim 100}$ lines. **B**, Two independent 4xrCUG$_{\sim 100}$. **C**, Two independent 4xrCAG$_{\sim 100}$ lines.

To examine whether changes were significant in the populations examined, categories were combined as before to determine the total proportion with any phenotype, and total proportion in the top two severity categories ('strong phenotype') for each genotype (Figure 6.6, Appendix 3.4). In only one of two *4xrCAG$_{\sim 100}$* lines the proportion with any phenotype, and the proportion with a strong phenotype was significantly increased with *dcr1^{Q1147X}* (Figure 6.6 B). Conversely in one of two *4xrCUG$_{\sim 100}$* lines the proportion with any phenotype was significantly reduced in the mutant background (Figure 6.6 C). Given that results were not replicated in independent lines, and that lines showing modification gave inconsistent effects these results do not provide any strong evidence that Dicer-1 processing makes a significant contribution to the common pathways leading to tergite disruption.

However this data also does not rule out the possibility of a sequence specific interaction with the Dicer-1 pathway. In this case *4xrCAG$_{\sim 100}$ [line 1]* did not provide sufficient progeny for accurate analysis while *4xrCAG$_{\sim 100}$ [line 2]* showed a significant enhancement. Since both lines gave a relatively strong phenotype, examining lines with a lower repeat RNA dosage (*2xrCAG$_{\sim 100}$*) and hence milder phenotype may be necessary to determine if this enhancement is reproducible. In contrast, both *4xrCUG$_{\sim 100}$* lines appeared to show suppression, but this was only significant for line 1. Therefore, further analysis would be required to determine whether these observations relate to an interaction with the Dicer-1 pathway, or are caused by a transgene insertion specific effect in the two lines that showed modification.

Figure 6.6 Analysis of the effect of reducing Dicer-1 levels on the total phenotype proportion and proportion with a strong phenotype.



Charts show the proportion with any phenotype (in category 2, 3 or 4 as indicated previously) or the proportion with a 'strong phenotype' (category 3 or 4 only). Each chart compares the value for each genotype with or without one copy of the *dcr1*^{Q1147X} allele (indicated below chart). **A**, *4xrCAG*$_{\sim 100}$ *[line 1]*. **B**, *4xrCAG*$_{\sim 100}$ *[line 2]*. **C**, *4xrCUG*$_{\sim 100}$ *[line 1]*, **D**, *4xrCUG*$_{\sim 100}$ *[line 2]*.

6.4 Chapter discussion

Growing evidence suggests that multiple pathways may be involved in human expanded repeat mediated disease [132, 166, 168, 219]. In our *Drosophila* model we have thus far found evidence for pathogenesis mediated by bi-directional repeat transcription from a single locus (*rCAG$_{\sim 100}$ [line C]*); complementary repeat expression from different loci ; and single-stranded hairpin repeat RNA (Chapter 3)[38]. Understanding whether these effects involve common or distinct pathogenic pathways will be important in determining their contribution to human disease. Experiments in this chapter therefore aimed to undertake preliminary comparisons between each, focussing on determining whether the Dicer processing pathways play a role in the observed effect.

In these experiments, the specific locomotion defects observed in *rCAG$_{\sim 100}$ [line C]* progeny were not replicated by neuronal expression of complementary *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$*. However, other work using this system indicates that expression of two copies of each repeat (2*xrCAG$_{\sim 100}$.rCUG$_{\sim 100}$*) in the neurons using a different driver (*Elav.2-GAL4*) results in viable progeny that show a reduced ability to climb with age [166]. In this case, flies do not appear to show the specific locomotion defect observed with *rCAG$_{\sim 100}$ [line C]* (Figure 5.1 A, B), and effects are not observed in young flies while they are in *rCAG$_{\sim 100}$ [line C]* (Figure 5.7 B*)* [166]. An explanation for these differences may be that bi-directional expression from the same genomic site leads to specific effects, that are not replicated by expression of complementary repeats from different sites. Some evidence for this exists in the form of a study showing that convergent transcription of repeats leads to apoptosis in a cell culture model [220]. Alternatively, the endogenous *cheerio* promoter may contribute to this specificity by restricting complementary CUG repeat transcription to a particular spatial or temporal pattern. This would be distinct from the *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* expression where both repeats are under UAS control. The *cheerio* gene locus at which *rCAG$_{\sim 100}$ [line C]* is inserted may also make some contribution to the specific locomotion defect. Although experiments in Chapter 5 suggest that this is not through a simple decrease or increase in cheerio levels, a more complex mechanism may impart some specificity. Further analysis using a system in which bi-directional and complementary repeat transcription can be directly compared

with similar inducible promoters would be necessary to examine these differences further.

Attempts were made to determine whether *rCAG$_{\sim 100}$ [line C]* photoreceptor degeneration is modified by alterations in Dicer-2, however no significant effect was observed. This is in contrast to previous results showing a strong enhancement is observed due to Dicer-2 overexpression with complementary GAL4/UAS driven repeat expression from different loci [166]. Preliminary attempts were also made to determine whether altering Dicer-2 levels alters locomotion defects with neuronal *rCAG$_{\sim 100}$ [line C]* expression, but identified no significant difference. This approach was complicated by the high variability involved in the phenotype. Together, no significant alterations in *rCAG$_{\sim 100}$ [line C]* phenotypes were observed with altered Dicer-2 processing.

If complementary repeat RNAs form dsRNA that is a substrate for Dicer processing, enhancement in this case may be caused by an increase in the processing of abundant dsRNA substrate. However, in the case of *rCAG$_{\sim 100}$ [line C]*, expression of the complementary rCUG transcript from the *cheerio* promoter is likely to occur at a lower level than the GAL4/UAS driven rCAG transcript, such that the amount of potential dsRNA is limiting. Therefore, increased processing when Dicer-2 is overexpressed may be limited by the amount of double-stranded repeat RNA substrate, and subsequently have no effect on the phenotype. Future experiments examining small RNA profiles in *rCAG$_{\sim 100}$ [line C]* lines may determine whether processed repeat containing small RNAs can be detected, thus indicating the involvement of Dicer processing in this case. Likewise, examining the ability of altered Dicer-1 processing to modify *rCAG$_{\sim 100}$ [line C]* phenotypes may indicate that the phenotype has a greater reliance on the Dicer-1, rather than Dicer-2, processing in this case.

Based on the evidence that single-stranded hairpin RNA can be a substrate for Dicer processing [149, 150] further experiments examined whether this pathway may be involved in the previously characterised adult tergite phenotypes. This may indicate a shared mechanism between single stranded hairpin-forming RNA mediated effects and those induced by complementary repeats that form dsRNA. No significant

changes in tergite phenotype severity were observed with reduced Dicer-2 levels, suggesting that this pathway is not limited by Dicer-2 processing. In further support of this, expression of *4xrCAG$_{\sim 100}$* or *4xrCUG$_{\sim 100}$* in neurons does not lead to production of CAG 21-mers that appear to be associated with *rCAG$_{\sim 100}$.rCUG$_{\sim 100}$* mediated effects [166]. Likewise, *4xrCAG$_{\sim 100}$* or *4xrCUG$_{\sim 100}$* expression in the eye has no external effect, and increased Dicer-2 processing does not produce a phenotype in this case [166].

Significant, but inconsistent, alterations were observed in the severity of the tergite phenotype due to a reduction in Dicer-1 levels. Interestingly, whilst one of the *rCAG$_{\sim 100}$* lines showed a mild enhancement, one of the *rCUG$_{\sim 100}$* lines showed a mild suppression, however, effects in other lines were not significant. One possibility may be that CAG and CUG repeats in this case lead to a similar tergite disruption phenotype through distinct pathways, however this does not seem likely. Alternatively reductions in Dicer-1 levels may exert a differential effect indirectly through alterations in other pathways. Dicer 1 processing is involved in the biogenesis of miRNAs that regulate many functions [187], and therefore may lead to a differential effects on tergite disruption. It is also possible that any differential effects may be due to direct processing of each repeat by Dicer-1. In support of this CUG and CAG hairpin RNA display distinct structural properties that may alter their affinity for RNA binding proteins [137], however it is currently unclear how this might lead to the observations described.

Overall it was concluded that neither Dicer-2, nor Dicer-1 play a significant role in common tergite phenotypes. This indicates that a distinct pathway may be responsible for adult tergite disruption mediated by hairpin RNA, to that identified for dsRNA. This *Drosophila* model therefore provides a system in which to examine genetic modifiers and may identify novel pathways perturbed by hairpin RNA. Further investigation of the mechanism, and specificity, involved in single-stranded hairpin-forming RNA-mediated effects will be required to determine a role in human disease.

CHAPTER 7 : Final discussion

7.1 Summary of results

Experiments described in this thesis use *Drosophila* as a model system to examine the role of expanded repeat RNA-mediated pathology as a common contributor to dominant expanded repeat disease. Previous work in our group established a system to ectopically express different repeat sequences and found that repeat RNA expression leads to specific transcriptional changes in neurons, but no phenotype when expressed in the developing eye [38, 166]. In the work for this thesis, a phenotype involving disruption to the adult tergites and reduced viability was identified due to CUG or CAG repeat RNA expression ubiquitously. Both repeats are able to form a hairpin secondary structure in this case, while a non-hairpin-forming CAA repeat RNA had no effect on adult tergites. Expression specifically within histoblast cells that give rise to the adult tergites led to a mild phenotype providing evidence that disruption is due to perturbation of this cell population.

The tergite disruption phenotype provides a biological read-out of common CUG and CAG repeat-mediated cellular perturbation and may therefore be used to examine the pathways involved in pathology. Initially the RNA binding-protein muscleblind was tested as a candidate, with results supporting that the pathway involving perturbation of muscleblind is not a major contributor to the tergite disruption phenotype. Further experiments examined repeat RNA localisation, with results showing that CUG repeat RNA localises to specific nuclear foci in muscle cells. In contrast CAG and CAA repeats both showed localisation that was almost identical, but was clearly distinct from that seen with CUG in muscle cells. Examination of different repeat contexts revealed that CUG, but not CAG and CAA was sufficient to induce specific localisation in muscle nuclei. The ability to form specific nuclear foci was limited to CUG expression in muscle cells and does not correlate with tergite disruption. It was therefore concluded that common CUG and CAG mediated tergite disruption may occur through a pathway distinct from that previously identified involving sequestration of RNA binding-proteins, including muscleblind, to nuclear foci.

Examination of hairpin-forming RNA effects in other tissues using our system has previously shown no visible phenotype in the *Drosophila* eye or neurons [38, 166]. However a single transgenic line, *rCAG$_{\sim 100}$ [line C]*, was identified that gave dominant, tissue-specific phenotypes. Expression of *rCAG$_{\sim 100}$ [line C]* in neurons leads to locomotion defects, while expression in the eye leads to mild disruption and photoreceptor degeneration, and ubiquitous expression leads to a reduction in lifespan. The presence of the insertion alone without expression, or expression of repeat RNA from a different site did not reproduce the phenotype. Analysis of the *rCAG$_{\sim 100}$ [line C]* insertion site identified its location at the *cheerio* locus, however, further experiments did not support alterations in *cheerio* levels as the cause of the phenotypes. Examination of transcription at the locus identified a complementary CUG transcript produced from the *cheerio* promoter, suggesting that bi-directional transcription of the repeat to produce complementary repeat transcripts may explain phenotypes specific to this line. Other results using our system suggest that expression of complementary repeats from different loci leads to toxicity in *Drosophila* that is dependent on Dicer processing [166]. It was therefore proposed that bi-directional transcription of complementary repeats, and formation of double-stranded repeat RNA may represent a novel pathogenic mechanism [166].

Attempts were made to compare hairpin (*4xrCAG$_{\sim 100}$ or 4xrCUG$_{\sim 100}$*), bi-directional (*rCAG$_{\sim 100}$ [line C])*, and complementary repeat (rCAG$_{\sim 100}$.rCUG$_{\sim 100}$) phenotypes in *Drosophila*, to examine the whether common pathways may be involved. Overexpression of Dicer-2, which has been shown to enhance double-stranded repeat RNA pathology, did not cause any significant enhancement of photoreceptor degeneration associated with *rCAG$_{\sim 100}$ [line C]* expression. However, this observation may be explained by the specific mechanics of the *rCAG$_{\sim 100}$ [line C]* line, in which bi-directional repeat transcription occurs from the endogenous *cheerio* locus and the transgene UAS. This may result in lower steady state levels of dsRNA, and therefore further analysis will be required to confirm the involvement of this pathway in *rCAG$_{\sim 100}$ [line C]* pathology.

To examine whether a common pathway is involved in hairpin repeat, and complementary repeat-mediated mechanisms, the effects of genetically reducing

Dicer function was examined on hairpin-forming repeat RNA-induced tergite phenotypes. Altering Dicer pathways modifies complementary repeat RNA mediated pathology [166], however, no consistent changes in the tergite phenotype were observed with reduced Dicer-1 or Dicer-2. These findings suggest that Dicer pathways are not rate limiting for hairpin repeat RNA-mediated tergite pathology, and provide evidence supporting distinct mechanisms in single-stranded repeat RNA, and complementary repeat RNA-mediated pathology.

7.2 Pathways of hairpin RNA-mediated pathology

An important finding in this study is the identification of disruption to adult *Drosophila* tergite patterning when repeats are expressed ubiquitously. Thus CUG or CAG repeat RNA is able to cause similar cellular perturbation, and may therefore involve common pathways in this *Drosophila* model. Reducing *muscleblind* levels did not lead to a consistent change in the tergite phenotype, and analysis of RNA localisation showed no correlation between the formation of specific foci, and tergite disruption. Results therefore support that tergite disruption involves a pathway independent of that involving sequestration of muscleblind.

Nonetheless, in our model it is not possible to rule out that muscleblind-dependent pathways contribute to cellular perturbation in some form. This is highlighted by results specific to CUG expression where reducing muscleblind levels mildly suppressed tergite pathology. CUG expression also led to muscle-specific foci suggesting that some interaction with the muscleblind sequestration pathway may occur. The absence of these observations with CAG expression suggest that this is not the main pathway responsible for tergite pathology, however, muscleblind-dependent pathways may contribute to some degree, especially in the case of CUG expression.

CUG RNA-mediated pathology involving the formation of RNA foci, and muscleblind sequestration, has been previously identified in *Drosophila* suggesting that the pathway is conserved [89, 142]. However, consistent with the results in this thesis, recent findings suggest that this mechanism may not be induced by CAG

repeat expression. In another *Drosophila* model, CAG repeat RNA was toxic in neurons and formed foci in muscle cells, however, alterations in splicing were not observed suggesting an alternative pathway to that involving muscleblind may be involved [132]. In this case muscleblind overexpression led to an enhanced phenotype, rather then the expected suppression based on a mechanism of muscleblind sequestration [132]. Likewise, in mouse and cell culture models, CAG repeat RNA forms foci that co-localise with muscleblind, but alterations in splicing were not observed [88, 170]. While some evidence for muscleblind mediated CAG repeat pathology is provided by a *C. elegans* model in which *C. elegans* muscleblind (CeMBL) overexpression suppressed phenotypes, it is uncertain if this occurs through splicing or some other function of CeMBL [171].

Although previous studies have focussed on muscleblind-dependent sequestration in foci, it appears that this pathway may not account for all pathology. Interactions with CUG-BP, which is not found in foci, are also important for pathology [70, 96, 145]. Similarly, a mouse model comparing transcriptional changes caused by CUG expression to those caused by loss of MBNL-1 function suggest that expanded CUG causes transcriptional changes independent of MBNL-1 dysregulation [143]. It is therefore possible that both CUG and CAG RNA contribute to pathology through a common pathway that is distinct from that involving MBNL-1 dysregulation of splicing.

The observation of specific tergite disruption induced by either CUG or CAG repeat RNA expression may provide clues to candidate pathways involved in pathology. In this case, disruption appears to involve a direct effect on specific cells that eventually migrate and differentiate to become adult epidermal cells [192, 193]. The formation of the adult abdominal epidermis involves a number of important biological processes that may be candidates for repeat mediated disruption. One example is cytoskeletal dynamics, that play an important role in the histoblast migration [193]. Actin dynamics are important for intercalation of cells during migration, and subsequent displacement of larval epithelial cells, such that altering actin dynamics leads to cleft tergites [193]. Alterations in cytoskeletal proteins were identified in previous microarray studies using our model of repeat RNA-mediated pathology, in a mouse model showing MBNL-1 independent transcriptional changes in DM1, in a

Drosophila model of CGG RNA toxicity and a *C. elegans* model of repeat RNA-mediated pathology [38, 113, 143, 171]. It would therefore be interesting to investigate whether changes in cytoskeletal dynamics contribute to repeat induced tergite disruption, by examining whether altering known components of this pathway can modify the phenotype.

7.3 Double-stranded repeat RNA-mediated pathogenesis

Work presented in this thesis, along with other results from our group, highlights a novel RNA repeat-mediated mechanism involving bi-directional transcription leading to complementary repeat transcripts that form double-stranded repeat RNA [166]. Two mechanisms have been proposed to account for this effect. The first involves producing small repeat RNAs that target endogenous repeat-containing transcripts in the genome, however we find no significant evidence for this [166]. Alternatively, abundant double-stranded repeat RNA may lead to dysregulation of normal double-stranded RNA processing pathways, in a mechanism analogous to muscleblind dysregulation of splicing. In support for this we see alterations to miRNA profiles in *Drosophila* neurons expressing complementary repeats [166].

During the writing of this thesis, an independent study was published in support of double-stranded repeat RNA-mediated pathogenesis [167]. In this case expression of complementary CUG and CAG transcripts led to increased toxicity compared to either repeat alone [132, 167]. Co-expression led to a reduction in full length repeat transcript levels and a corresponding increase of 21nt small RNAs that was Dicer-2 dependent, indicating that as in our model complementary transcripts form double-stranded RNA that is processed. In contrast to work from our lab, this model suggests that pathology is dependent on Dicer-2 processing, but not Dicer-1 with no change in processing of the endogenous miRNAs tested [167]. The authors found that small RNAs act as siRNAs to target CAG repeat containing endogenous transcripts, but interestingly not CUG endogenous transcripts, suggesting a preference for loading CUG small RNA [167]. Results from our group identified the opposite effect, with abundant small CAG RNA reads from RNA deep-sequencing, but no consistent changes in endogenous transcripts, and altered miRNA levels [166]. Different repeat

constructs and analysis methods were used in each case and this may underlie the distinctions observed. Further studies will be necessary to determine the exact basis for double-stranded RNA mediated pathology, as well as identify hallmarks such as altered miRNA profiles in human patients.

Bi-directional repeat transcription has been identified in a number of disease associated expanded repeats [119, 151, 153-155]. This process has also been identified widely throughout the genome and has functional roles [156]. Therefore, as well as the role in generating double-stranded repeat RNA, bi-directional transcription may contribute to expanded repeat pathology through other mechanisms. One area requiring further investigation is the direct role played by the process of bi-directional transcription. In this study, differences were observed when comparing bi-directional transcription from the *cheerio* insertion, to transcription of complementary repeat RNAs from different genomic loci. Some specificity in this case may be imparted by the *cheerio* promoter as well as possible effects on *cheerio* function. Further investigation will therefore be required to determine whether bi-directional transcription itself plays a direct role in pathology, or whether the generation of complementary transcripts is the critical pathogenic agent. Some evidence suggests that regulation at the level of DNA and chromatin may play a role in expanded repeat disease [151, 152, 159] and bi-directional transcription may therefore play a role in this process.

7.4 Multiple pathways contribute to expanded repeat disease

Increasing evidence suggests the potential for multiple pathways to contribute to human expanded repeat disease. Numerous studies now support the role of both expanded protein tracts, and expanded RNA transcripts as pathogenic agents. Within this thesis results are presented supporting multiple mechanisms by which repeat RNA may induce pathology (Figure 7.1)

Recently a novel mechanism was reported whereby translation of repeat RNA can be initiated independent of an ATG start codon [168]. This may result in proteins translated in any of the three possible reading frames and was shown to occur in cell

culture, SCA8 and DM1 mouse models, and human tissues [168]. This result suggests an added layer of complexity whereby repeats previously considered untranslated may contribute to pathology by producing polyglutamine, as well as a number of other homopolymeric peptides. In the case of results in this thesis, ubiquitous expression of four copies of an untranslated CUG or CAG repeat transgene led to reduced viability and tergite disruption, while expression of one copy of a translated CAG repeat transgene leads to complete lethality (C. van Eyk, in submission). Likewise expression of four copies of untranslated CAG or CUG transgenes in the eye led to no phenotype, while expression of a translated CAG transgene leads to strong disruption and lethality in some cases [37, 38, 166]. Together these results suggest that non-ATG translation is not likely to play a major role in the repeat RNA-mediated phenotypes observed in our model. Support is also provided by independent reports suggesting that untranslated CAG repeat expression does not produce polyglutamine protein [132]. Further analysis will be necessary to determine if non-ATG repeat translation occurs in *Drosophila* and if so, its contribution to pathology. Interestingly, expression of a translated GCA repeat to give polyalanine protein ubiquitously gives a tergite disruption phenotype identical to that observed with $rCAG_{\sim 100}$ (C. van Eyk, in submission). However, this analysis is complicated as the effect may equally be caused by the CAG hairpin-forming RNA in this case. Nonetheless, polyalanine provides a potential common pathogenic agent as a number of disease causing repeats could produce frameshift products encoding polyalanine. Further investigation will be required to determine the extent to which non-ATG repeat-mediated translation contributes to the tergite phenotype, as well as pathology in general.

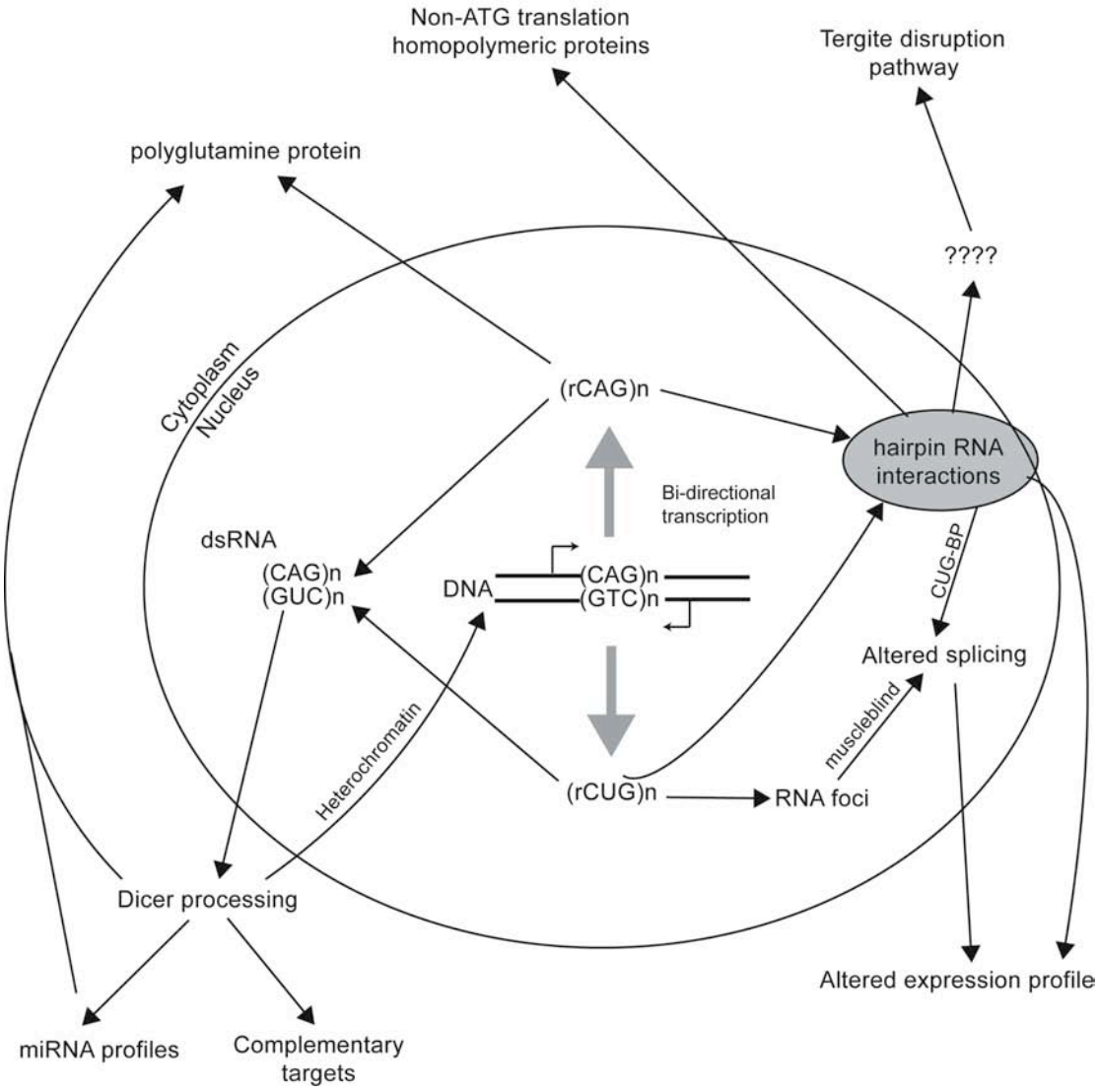
While previous efforts have focussed on a single unifying pathway in expanded repeat disease, increasing evidence is consistent with the existence of multiple non-exclusive mechanisms. Multiple pathways with different contributions may provide an explanation for some of the disease, and cell specific effects that are observed. Therefore, understanding the relative contributions from each of these mechanisms now becomes an essential part of understanding the basis for human disease. Furthermore, in the case of co-existing mechanisms, determining if and how each is influenced by the others becomes an important consideration, and may provide explanations for some of the uncertainties within the field. For example, the existence

of a single pathogenic agent has yet to account for the susceptibility of certain cell populations in each human disease. The existence of multiple pathogenic agents could provide a situation where a spatial pattern of cellular susceptibility is determined by overlapping patterns of activity from each mechanism and interactions between pathways.

A number of interesting results already exist that may indicate connections between distinct repeat mediated mechanisms. miRNAs have been shown to regulate alternative splicing networks involving CUG-BP [221], and thus double-stranded RNA mediated effects may have the ability to modulate the disregulation of splicing. Similarly, changes in Dicer processing and specific miRNAs have been shown to modify polyglutamine toxicity [160]. Likewise bi-directional transcription may regulate repeat-containing mRNA levels, while Dicer processing leads to a reduction in full length transcripts, both of which may have an effect on polyglutamine production [155, 167]. Further examination of the interactions between pathways will be important in understanding these diseases, as well as designing therapeutics where targeting of multiple pathways would be ideal.

Drosophila models have been used to examine a number of the important pathways that contribute to disease, and will be an important tool in further work. Using *Drosophila* to identify genetic pathways and specific biomarkers will help to determine the relative contribution of different pathogenic mechanisms to human disease. Similarly, the ease of genetic manipulation, and resources available for *Drosophila* may allow investigation of the interactions, and competition between pathogenic pathways that may not be possible in other systems.

Figure 7.1 Multiple pathways leading to cellular perturbation in *Drosophila* models of expanded repeat disease.



Schematic illustrating proposed pathways that contribute to repeat-mediated cellular perturbation, based on a hypothetical coding CAG / CTG expanded repeat that is bi-directionally transcribed. Potential connections between pathways are indicated by arrows.

7.5 Future directions

As the mechanisms identified in dominant expanded repeat disease become increasingly more complex, a number of important areas of future research have been highlighted. Important questions exist in terms of understanding the details of specific pathways, as well as the overall integration, and interactions between mechanisms.

Repeat RNA-mediated toxicity now appears relevant at a number of levels. Different hairpin-forming repeat RNA sequences appear to induce pathology through both common and unique pathways. Much work has focussed on the pathways involving dysregulation of splicing in CUG RNA induced pathology, but evidence exists for RNA induced pathways independent of this mechanism, and understanding these effects will be important. Understanding the common and distinct pathways perturbed by different repeat RNAs may uncover specific therapeutic targets relevant to a range of diseases. *Drosophila* provides an ideal system to undertake such studies, and using phenotypes such as tergite disruption will enable the identification of specific pathogenic pathways that inform therapeutic design.

Work in our group, and by others has now identified a novel form of RNA-mediated pathology involving the formation of double-stranded repeat RNA [166, 167]. Many questions remain in regard to the details of the mechanism involved, such as the dependence on different aspects of small RNA processing and the mechanism by which small repeat RNAs exert an effect. The specific changes identified in this work also provide biomarkers that may be used to identify the existence of this pathway in human disease, and this is now being undertaken in our lab (S. Samaraweera, unpublished). Determining the contribution made directly by bi-directional transcription from a single locus, compared to the subsequent presence of complementary repeat transcripts will also be important, and may uncover specific effects caused by bi-directional transcription. Another important question involves whether bi-directional transcription, and subsequent double-stranded RNA mediated effects are dependent on repeat length and how this might occur. This could provide an explanation for the existence of particular pathogenic repeat length thresholds, which remains an important question in the field.

Finally, determining the relative contributions, and dynamics involved in multiple repeat-mediated pathways will be an important aim for future research. For example, does Dicer processing reduce the level of repeat mRNA available for translation in the case where repeats are translated? Likewise, does the ability of repeat RNA to bind and sequester certain proteins reduce the availability of transcripts to form double-stranded RNA? Given an equilibrium is likely formed between pathways, how is this altered by repeat length or cell specific factors? Understanding this may lead to an explanation for the poorly understood basis for cell specificity, and age of onset observed in different expanded repeat diseases. Identifying both the common and distinct pathways involved in the dominant expanded repeat diseases will be essential to developing effective therapies in the future.

Appendicies

Appendix 1

Viability when each repeat is expressed via *da-GAL4*

Genotype	n	4xtransgene	balancer	proportion (4xtransgene)	95%CI	p value (vs 4xUAS)	p value (vs 4xrCAA$_{\sim 100}$ [line 1])
4xUAS	365	220	145	0.603	0.551-0.655	-	-
4xrCAA$_{\sim 100}$ [line 1]	474	299	175	0.631	0.586-0.674	0.431	-
4xrCAA$_{\sim 100}$ [line 2]	201	143	58	0.711	0.644-0.773	**0.010**	0.051
4xrCUG$_{\sim 100}$ [line 1]	152	84	68	0.553	0.470-0.633	0.327	0.104
4xrCUG$_{\sim 100}$ [line 2]	247	89	158	0.360	0.300-0.424	**<0.001**	**<0.001**
4xrCAG$_{\sim 100}$ [line 1]	373	201	172	0.539	0.487-0.590	0.087	**0.008**

Genotype	n	4xtransgene	balancer	Proportion (4xtransgene)	95%CI	p value (vs 4xUAS)	p value (vs 4xrCAA$_{\sim 100}$ [line 1])
4xUAS	85	56	29	0.659	0.548-0.758	-	-
4xrCAA$_{\sim 100}$ [line 1]	73	57	16	0.781	0.669-0.869	0.1121	-
4xrCAA$_{\sim 100}$ [line 2]	101	74	27	0.733	0.635-0.816	0.3359	0.4833
4xrCUG$_{\sim 100}$ [line 2]	17	0	17	0.000	0.000-0.195	**<0.001**	**<0.001**
4xrCAG$_{\sim 100}$ [line 1]	45	0	45	0.000	0.000-0.079	**<0.001**	**<0.001**

Raw data for viability experiments. For each genotype total population size (n) is shown along with number of progeny that express four copies of the transgene, and number that inherit the compound balancer chromosome. Proportion with four copies of the transgene, and 95% confidence interval (based on a binomial distribution) for the particular proportion are shown. P values are given for Fisher's exact test using the raw values comparing the number of 4xtransgene, and balancer progeny for each genotype to either the *4xUAS* control, or *4xrCAA$_{\sim 100}$ [line 1].*

Appendix 2.1

Distribution of progeny between phenotype categories in all repeat lines with and without *da-GAL4* driver.

da-GAL4 >	n=	Proportion [1]	Proportion [2]	Proportion [3]	Proportion [4]
da-GAL4 > +	506	0.990	0.000	0.010	0.000
+ / +	401	0.998	0.000	0.002	0.000
da-GAL4 > 4xUAS	161	0.994	0.006	0.000	0.000
+ / 4xUAS	203	1.000	0.000	0.000	0.000
da-GAL4 > 4xrCAA$_{\sim 100}$ [line 1]	158	0.986	0.000	0.007	0.007
+ / 4xrCAA$_{\sim 100}$ [line 1]	200	1.000	0.000	0.000	0.000
da-GAL4 > 4xrCAA$_{\sim 100}$ [line 2]	148	0.987	0.006	0.006	0.000
+ / 4xrCAA$_{\sim 100}$ [line 2]	241	1.000	0.000	0.000	0.000
da-GAL4 > 4xrCUG$_{\sim 100}$ [line 1]	271	0.070	0.782	0.125	0.022
+ / 4xrCUG$_{\sim 100}$ [line 1]	127	0.992	0.008	0.147	0.000
da-GAL4 > 4xrCUG$_{\sim 100}$ [line 2]	63	0.222	0.746	0.032	0.000
+ / 4xrCUG$_{\sim 100}$ [line 2]	229	1.000	0.000	0.000	0.0000
da-GAL4 > 4xrCAG$_{\sim 100}$ [line 1]	8	0.000	0.125	0.000	0.875
+ / 4xrCAG$_{\sim 100}$ [line 1]	150	1.000	0.000	0.000	0.000
da-GAL4 > 4xrCAG$_{\sim 100}$ [line 2]	343	0.096	0.458	0.315	0.131
+ / 4xrCAG$_{\sim 100}$ [line 2]	286	1.000	0.000	0.000	0.000
da-GAL4 > 4xrCAG$_{\sim 100}$ [line 3]	97	0.103	0.495	0.340	0.062
da-GAL4 > 4xrCAG$_{\sim 100}$ [line	95	0.442	0.432	0.105	0.021
da-GAL4 > 4xrCAG$_{\sim 100}$ [line	97	0.361	0.433	0.206	0.000

Proportion of total progeny (n) for each genotype that fall within each phenotype severity category.

Appendix 2.2

Preliminary analysis of tergite phenotypes with the ubiquitous *Act5c-GAL4* driver

Line	Act5C-GAL4
+	None (n = 95)
$rCUG_{\sim 100}$ [line 1]	Moderate
$rCUG_{\sim 100}$ [line 2]	n/a – low viability
$rCAG_{\sim 100}$ [line 1]	Severe - low viability
$rCAG_{\sim 100}$ [line 2]	Mild
$rCAG_{\sim 100}$ [line 3]	Moderate
$rCAA_{\sim 100}$ [line 1]	None
$rCAA_{\sim 100}$ [line 2]	None

Preliminary analysis of tergite phenotypes when repeat lines are driven with the ubiquitous *Act5c-GAL4* driver. Phenotype strength is based on a qualitative scale (mild, moderate, severe) where severe represents the worst phenotype observed of all lines, and cannot be compared directly to *da-GAL4* quantitative results. Relative severities appear to be approximately comparable between drivers where *$rCAG_{\sim 100}$ [line 1]* gave the most severe tergite phenotype with both *da-GAL4* (see Chapter 4) and *Act5c-GAL4*. Similarly, as for da-GAL4, *$rCAG_{\sim 100}$ [line 1]* and *$rCUG_{\sim 100}$ [line 2]* showed reduced viability.

Appendix 3.1

Effect on distribution of progeny between categories with and without *mbl*^{E27}

da-GAL4 >	n=	Proportion [1]	Proportion [2]	Proportion [3]	Proportion [4]
+	506	0.990	0.000	0.010	0.000
+ ; mbl^{E27}	159	1.000	0.000	0.000	0.000
4xrCAA$_{\sim 100}$ [line 1]	158	0.986	0.000	0.007	0.007
4xrCAA$_{\sim 100}$ [line 1] ; mbl^{E27}	34	1.000	0.000	0.000	0.000
4xrCAA$_{\sim 100}$ [line 2]	148	0.987	0.006	0.006	0.000
4xrCAA$_{\sim 100}$ [line 2] ; mbl^{E27}	31	1.000	0.000	0.000	0.000
4xrCUG$_{\sim 100}$ [line 1]	271	0.070	0.782	0.125	0.022
4xrCUG$_{\sim 100}$ [line 1] ; mbl^{E27}	34	0.176	0.676	0.147	0.000
4xrCUG$_{\sim 100}$ [line 2]	63	0.222	0.746	0.032	0.000
4xrCUG$_{\sim 100}$ [line 2] ; mbl^{E27}	24	0.458	0.458	0.042	0.042
4xrCAG$_{\sim 100}$ [line 1]	8	0.000	0.125	0.000	0.875
4xrCAG$_{\sim 100}$ [line 1] ; mbl^{E27}	33	0.000	0.152	0.364	0.485
4xrCAG$_{\sim 100}$ [line 2]	343	0.096	0.458	0.315	0.131
4xrCAG$_{\sim 100}$ [line 2] ; mbl^{E27}	64	0.172	0.406	0.328	0.094

Each repeat line was expressed ubiquitously via *da-GAL4* and via *da-GAL4* in the presence of one copy of the *mbl*^{E27} allele. Table shows the total number of flies scored for each genotype (n), and the proportion of the total represented by each phenotype category where 0.000 is no progeny in that category and 1.000 is all progeny in that category.

Appendix 3.2

Effect on distribution of progeny between categories with and without *dcr2*^{L811fsX}

da-GAL4 >	n=	Proportion [1]	Proportion [2]	Proportion [3]	Proportion [4]
+	506	0.990	0.000	0.010	0.000
+ ; dcr2^{L811fsX}	85	1.000	0.000	0.000	0.000
4xrCAA$_{\sim 100}$ [line 1]	158	0.986	0.000	0.007	0.007
4xrCAA$_{\sim 100}$ [line 1] ; dcr2^{L811fsX}	15	1.000	0.000	0.000	0.000
4xrCAA$_{\sim 100}$ [line 2]	148	0.987	0.006	0.006	0.000
4xrCAA$_{\sim 100}$ [line 2] ; dcr2^{L811fsX}	10	1.000	0.000	0.000	0.000
4xrCUG$_{\sim 100}$ [line 1]	271	0.070	0.782	0.125	0.022
4xrCUG$_{\sim 100}$ [line 1] ; dcr2^{L811fsX}	27	0.148	0.704	0.148	0.000
4xrCUG$_{\sim 100}$ [line 2]	63	0.222	0.746	0.032	0.000
4xrCUG$_{\sim 100}$ [line 2] ; dcr2^{L811fsX}	21	0.286	0.619	0.095	0.000
4xrCAG$_{\sim 100}$ [line 1]	8	0.000	0.125	0.000	0.875
4xrCAG$_{\sim 100}$ [line 1] ; dcr2^{L811fsX}	10	0.000	0.100	0.200	0.700
4xrCAG$_{\sim 100}$ [line 2]	343	0.096	0.458	0.315	0.131
4xrCAG$_{\sim 100}$ [line 2] ; dcr2^{L811fsX}	59	0.102	0.322	0.373	0.203

Each repeat line was expressed ubiquitously via *da-GAL4* and via *da-GAL4* in the presence of one copy of the *dcr2*^{L811fsX} allele. Table shows the total number of flies scored for each genotype (n), and the proportion of the total represented by each phenotype category where 0.000 is no progeny in that category and 1.000 is all progeny in that category.

Appendix 3.3

Effect on distribution of progeny between categories with and without *dcr1*^{Q1147X}

da-GAL4 >	n=	Proportion [1]	Proportion [2]	Proportion [3]	Proportion [4]
+	506	0.990	0.000	0.010	0.000
+ ; dcr1^{Q1147X}	231	0.996	0.000	0.004	0.000
4xrCAA$_{\sim 100}$ [line 1]	158	0.986	0.000	0.007	0.007
4xrCAA$_{\sim 100}$ [line 1] ; dcr1^{Q1147X}	148	1.000	0.000	0.000	0.000
4xrCAA$_{\sim 100}$ [line 2]	148	0.987	0.006	0.006	0.000
4xrCAA$_{\sim 100}$ [line 2] ; dcr1^{Q1147X}	88	1.000	0.000	0.000	0.000
4xrCUG$_{\sim 100}$ [line 1]	271	0.070	0.782	0.125	0.022
4xrCUG$_{\sim 100}$ [line 1] ; dcr1^{Q1147X}	72	0.194	0.750	0.056	0.000
4xrCUG$_{\sim 100}$ [line 2]	63	0.222	0.746	0.032	0.000
4xrCUG$_{\sim 100}$ [line 2] ; dcr1^{Q1147X}	33	0.394	0.576	0.030	0.000
4xrCAG$_{\sim 100}$ [line 1]	8	0.000	0.125	0.000	0.875
4xrCAG$_{\sim 100}$ [line 1] ; dcr1^{Q1147X}	21	0.000	0.095	0.095	0.810
4xrCAG$_{\sim 100}$ [line 2]	343	0.096	0.458	0.315	0.131
4xrCAG$_{\sim 100}$ [line 2] ; dcr1^{Q1147X}	123	0.008	0.252	0.366	0.374

Each repeat line was expressed ubiquitously via *da-GAL4* and via *da-GAL4* in the presence of one copy of the *dcr1*^{Q1147X} allele. Table shows the total number of flies scored for each genotype (n), and the proportion of the total represented by each phenotype category where 0.000 is no progeny in that category and 1.000 is all progeny in that category.

Appendix 3.4

Statistical comparison of tergite severity when different mutations are introduced.

Any phenotype	*da ; mbl*^{E27}	*da ; dcr2*^{L811fsX}	*da ; dcr1*^{Q1147X}
$rCAG_{\sim 100}$ [line 1]	1.000	1.000	1.000
$rCAG_{\sim 100}$ [line 2]	0.081	0.815	0.000
$rCUG_{\sim 100}$ [line 1]	0.045	0.142	0.003
$rCUG_{\sim 100}$ [line 2]	0.037	0.564	0.096
Strong phenotype	*da ; mbl*^{E27}	*da ; dcr2*^{L811fsX}	*da ; dcr1*^{Q1147X}
$rCAG_{\sim 100}$ [line 1]	1.000	1.000	1
$rCAG_{\sim 100}$ [line 2]	0.785	0.068	<0.0001
$rCUG_{\sim 100}$ [line 1]	1.000	1.000	0.0459
$rCUG_{\sim 100}$ [line 2]	0.304	0.259	1

Tables shows p values from Fisher's exact test comparing genotypes for the distribution between progeny with any phenotype (category 2, 3 and 4) and others, or between progeny with a strong phenotype (category 3 and 4) and others. In each case comparisons are made to the population expressing each repeat with *da-GAL4* alone.

Appendix 4

Kynan T. Lawlor*, Louise V. O'Keefe*, Saumya E. Samaraweera*, Clare L. van Eyk, Catherine J. McLeod, Christopher A. Maloney, Thurston H.Y. Dang, Catherine M. Suter, and Robert I. Richards. (2011) Double-stranded RNA is pathogenic in Drosophila models of expanded repeat neurodegenerative diseases. Human Molecular Genetics *20*, 3757 -3768.

*Joint first authors.

Open Access electronic copies are available from :

http://hmg.oxfordjournals.org/content/20/19/3757.abstract

The following publication contains work described in Chapter 5. The work is described in full in Chapter 5, however the publication is included for reference and to place this work in the broader context of findings from the Richards Lab.

Author contributions : K.T.L. identified and characterised phenotypes in *rCAG$_{\sim 100}$ [line C]*, and determined the requirement for bi-directional transcription in this case (as per Chapter 5). L.V.O identified and characterised eye phenotypes due to complementary repeat expression from different loci (*rCAG$_{\sim 100}$.rCUG$_{\sim 100}$*), and the contribution of Dicer pathways. S.E.S. characterised eye phenotypes, and identified neuronal phenotypes due to complementary repeat expression from different loci (*rCAG$_{\sim 100}$.rCUG$_{\sim 100}$*), the contribution of Dicer pathways, performed qRT-PCR analysis and examined expression levels of complementary repeat containing transcripts. C.L.v.E. assisted with crosses, experimental design and analysis. C.J.M. generated transgenic constructs and provided preliminary analysis of the *rCAG$_{\sim 100}$ [line C]* eye phenotype. C.A.M, T.H.Y.D. and C.M.S. performed RNA sequencing and bioinformatic analyses. R.I.R. conceived of hypotheses being tested and with K.T.L., L.V.O., S.E.S., C.L.v.E., and C.J.M. designed the experiments, analysed the data and wrote the paper. All authors discussed the results and commented on the manuscript.

K.T. Lawlor, L.V. O'Keefe, S.E. Samaraweera, C.L. van Eyk, C.J. McLeod, C.A. Maloney, T.H.Y. Dang, C.M. Suter, and R.I. Richards. (2011) Double-stranded RNA is pathogenic in Drosophila models of expanded repeat neurodegenerative diseases.
Human Molecular Genetics v. 20 (19), pp. 3757 -3768, October 2011

References

1. Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science *252*, 1711-1714.
2. La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature *352*, 77-79.
3. La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nature Reviews Genetics *11*, 247-258
4. Wells, R.D., and Ashizawa, T. (2006). Genetic instabilities and neurological diseases, Academic Press.
5. Richards, R.I. (2001). Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. Human Molecular Genetics *10*, 2187-2194.
6. Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T., and Nelson, D.L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. Cell *66*, 817-822.
7. Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M., and Wells, R.D. (2001). Sticky DNA, a self-associated complex formed at long GAA*TTC repeats in intron 1 of the frataxin gene, inhibits transcription. The Journal of Biological Chemistry *276*, 27171-27177.
8. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., and Zhang, F.P. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905-914.
9. Coffee, B., Zhang, F., Warren, S.T., and Reines, D. (1999). Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. Nature Genetics *22*, 98-101.
10. Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423-1427.
11. Babcock, M., de Silva, D., Oaks, R., Davis-Kaplan, S., Jiralerspong, S., Montermini, L., Pandolfo, M., and Kaplan, J. (1997). Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. Science *276*, 1709-1712.
12. Feng, Y., Absher, D., Eberhart, D.E., Brown, V., Malter, H.E., and Warren, S.T. (1997). FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. Molecular Cell *1*, 109-118.
13. Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide Repeat Disorders. Annual Review of Neuroscience *30*, 575-621.
14. Bakker, C.E., and Oostra, B.A. (2003). Understanding fragile X syndrome: insights from animal models. Cytogenetic and Genome Research *100*, 111-123.

15. Cossée, M., Puccio, H., Gansmuller, A., Koutnikova, H., Dierich, A., LeMeur, M., Fischbeck, K., Dollé, P., and Koenig, M. (2000). Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation. Human Molecular Genetics *9*, 1219-1226.
16. Gatchel, J.R., and Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. Nature Reviews. Genetics *6*, 743-755.
17. Subramony, S.H. (2011). Overview of autosomal dominant ataxias. Ataxic Disorders *103*, 389-398.
18. Ashizawa, T., and Sarkar, P.S. (2011). Myotonic dystrophy types 1 and 2. Handbook of Clinical Neurology *101*, 193-237.
19. Leehey, M.A. and Hagerman, P.J. (2011). Fragile X-associated tremor/ataxia syndrome. Ataxic Disorders *103*, 373-386.
20. Tsuji, S. (2011). Dentatorubral-pallidoluysian atrophy. Ataxic Disorders *103*, 587-594.
21. Gallo, J.-M., Leigh, P.N., and Andrew, A.E. (2007). Chapter 8 Spinobulbar muscular atrophy (Kennedy's disease). Motor neuron disorders and related diseases *82*, 155-169.
22. Donato, S.D., Mariotti, C. and Taroni, F. (2011). Spinocerebellar ataxia type 1. Ataxic Disorders *103*, 399-421
23. Auburger, G.W. (2011). Spinocerebellar ataxia type 2. Ataxic Disorders *103*, 423-436.
24. Paulson, H. (2011). Machado-Joseph disease/spinocerebellar ataxia type 3. Ataxic Disorders *103*, 437-449.
25. Solodkin, A., and Gomez, C.M. (2011). Spinocerebellar ataxia type 6. Ataxic Disorders *103*, 461-473.
26. Martin, J.J. (2011). Spinocerebellar ataxia type 7. Ataxic Disorders *103*, 475-491.
27. Ikeda, Y., Ranum, L.P.W. and Day, J.W. (2011). Clinical and genetic features of spinocerebellar ataxia type 8. Ataxic Disorders *103*, 493-505.
28. Ashizawa, T., and Sankara, H.S.a.A.D.r. (2011). Spinocerebellar ataxia type 10. Ataxic Disorders *103*, 507-519.
29. O'Hearn, E., Holmes, S.E., Margolis, R.L., and Sankara, H.S.a.A.D.r. (2011). Spinocerebellar ataxia type 12. Ataxic Disorders *103*, 535-547.
30. McFarland, K.N., Cha, J.-H.J., and William, J.W.a.E.T. (2011). Molecular biology of Huntington's disease. Hyperkinetic Movement Disorders *100*, 25-81.
31. Duyao, M.P., Auerbach, A.B., Ryan, A., Persichetti, F., Barnes, G.T., McNeil, S.M., Ge, P., Vonsattel, J.P., Gusella, J.F., and Joyner, A.L. (1995). Inactivation of the mouse Huntington's disease gene homolog Hdh. Science *269*, 407-410.
32. Matilla, A., Roberson, E.D., Banfi, S., Morales, J., Armstrong, D.L., Burright, E.N., Orr, H.T., Sweatt, J.D., Zoghbi, H.Y., and Matzuk, M.M. (1998). Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-pulse facilitation. The Journal of Neuroscience *18*, 5508-5516.
33. Rudnicki, D.D., Pletnikova, O., Vonsattel, J.-P.G., Ross, C.A., and Margolis, R.L. (2008). A comparison of huntington disease and huntington disease-like 2 neuropathology. Journal of Neuropathology and Experimental Neurology *67*, 366-374.

34. Cho, D.H., and Tapscott, S.J. (2007). Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. Biochimica et Biophysica Acta *1772*, 195-204.
35. Machuca-Tzili, L., Brook, D., and Hilton-Jones, D. (2005). Clinical and molecular aspects of the myotonic dystrophies: a review. Muscle & Nerve *32*, 1-18.
36. Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. Nature Reviews Genetics *6*, 729-742.
37. McLeod, C.J., O'Keefe, L.V., and Richards, R.I. (2005). The pathogenic agent in Drosophila models of 'polyglutamine' diseases. Human Molecular Genetics *14*, 1041-1048.
38. van Eyk, C.L., O'Keefe, L.V., Lawlor, K.T., Samaraweera, S.E., McLeod, C.J., Price, G.R., Venter, D.J., and Richards, R.I. (2011). Perturbation of the Akt/Gsk3-β signalling pathway is common to Drosophila expressing expanded untranslated CAG, CUG and AUUCU repeat RNAs. Human Molecular Genetics *20*, 2783-2794.
39. Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. Annual Review of Neuroscience *30*, 575-621.
40. Ordway, J.M., Tallaksen-Greene, S., Gutekunst, C.A., Bernstein, E.M., Cearley, J.A., Wiener, H.W., Dure, L.S.t., Lindsey, R., Hersch, S.M., Jope, R.S., et al. (1997). Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. Cell *91*, 753-763.
41. Watase, K., Weeber, E.J., Xu, B., Antalffy, B., Yuva-Paylor, L., Hashimoto, K., Kano, M., Atkinson, R., Sun, Y., Armstrong, D.L., et al. (2002). A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. Neuron *34*, 905-919.
42. Yoo, S.Y., Pennesi, M.E., Weeber, E.J., Xu, B., Atkinson, R., Chen, S., Armstrong, D.L., Wu, S.M., Sweatt, J.D., and Zoghbi, H.Y. (2003). SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity. Neuron *37*, 383-401.
43. Marsh, J.L., Walker, H., Theisen, H., Zhu, Y.-Z., Fielder, T., Purcell, J., and Thompson, L.M. (2000). Expanded polyglutamine peptides alone are intrinsically cytotoxic and cause neurodegeneration in Drosophila. Human Molecular Genetics *9*, 13-25.
44. DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. (1997). Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science *277*, 1990-1993.
45. Lunkes, A., and Mandel, J.L. (1997). Polyglutamines, nuclear inclusions and neurodegeneration. Nature Medicine *3*, 1201-1202.
46. Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. Nature Medicine *10 Suppl*, S10-17
47. Ciechanover, A., and Brundin, P. (2003). The Ubiquitin Proteasome System in Neurodegenerative Diseases: Sometimes the Chicken, Sometimes the Egg. Neuron *40*, 427-446.

48. Chan, H.Y., Warrick, J.M., Gray-Board, G.L., Paulson, H.L., and Bonini, N.M. (2000). Mechanisms of chaperone suppression of polyglutamine disease: selectivity, synergy and modulation of protein solubility in Drosophila. Human Molecular Genetics *9*, 2811-2820.
49. Cummings, C.J., Sun, Y., Opal, P., Antalffy, B., Mestril, R., Orr, H.T., Dillmann, W.H., and Zoghbi, H.Y. (2001). Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice. Human Molecular Genetics *10*, 1511-1518.
50. Bowman, A.B., Yoo, S.-Y., Dantuma, N.P., and Zoghbi, H.Y. (2005). Neuronal dysfunction in a polyglutamine disease model occurs in the absence of ubiquitin-proteasome system impairment and inversely correlates with the degree of nuclear inclusion formation. Human Molecular Genetics *14*, 679-691.
51. Bennett, E.J., Shaler, T.A., Woodman, B., Ryu, K.-Y., Zaitseva, T.S., Becker, C.H., Bates, G.P., Schulman, H., and Kopito, R.R. (2007). Global changes to the ubiquitin system in Huntington's disease. Nature *448*, 704-708.
52. Saudou, F., Finkbeiner, S., Devys, D., and Greenberg, M.E. (1998). Huntingtin Acts in the Nucleus to Induce Apoptosis but Death Does Not Correlate with the Formation of Intranuclear Inclusions. Cell *95*, 55-66.
53. Bence, N.F., Sampat, R.M., and Kopito, R.R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. Science *292*, 1552-1555.
54. Opal, P., and Zoghbi, H.Y. (2002). The role of chaperones in polyglutamine disease. Trends in Molecular Medicine *8*, 232-236.
55. Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature *431*, 805-810.
56. Ravikumar, B., Vacher, C., Berger, Z., Davies, J.E., Luo, S., Oroz, L.G., Scaravilli, F., Easton, D.F., Duden, R., O'Kane, C.J., et al. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. Nature Genetics *36*, 585-595.
57. Nucifora, F.C., Jr., Sasaki, M., Peters, M.F., Huang, H., Cooper, J.K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V.L., et al. (2001). Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. Science *291*, 2423-2428.
58. McCampbell, A., Taylor, J.P., Taye, A.A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G., et al. (2000). CREB-binding protein sequestration by expanded polyglutamine. Human Molecular Genetics *9*, 2197-2202.
59. Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tzvetkov, N., Strippel, N., Sakahira, H., Siegers, K., Hayer-Hartl, M., and Hartl, F.U. (2004). Cellular Toxicity of Polyglutamine Expansion Proteins: Mechanism of Transcription Factor Deactivation. Molecular Cell *15*, 95-105.
60. Bowman, A.B., Lam, Y.C., Jafar-Nejad, P., Chen, H.-K., Richman, R., Samaco, R.C., Fryer, J.D., Kahle, J.J., Orr, H.T., and Zoghbi, H.Y. (2007). Duplication of Atxn1l suppresses SCA1 neuropathology by decreasing incorporation of polyglutamine-expanded ataxin-1 into native complexes. Nature Genetics *39*, 373-379.

61. Lim, J., Crespo-Barreto, J., Jafar-Nejad, P., Bowman, A.B., Richman, R., Hill, D.E., Orr, H.T., and Zoghbi, H.Y. (2008). Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. Nature *452*, 713-718.
62. Emamian, E.S., Kaytor, M.D., Duvick, L.A., Zu, T., Tousey, S.K., Zoghbi, H.Y., Clark, H.B., and Orr, H.T. (2003). Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. Neuron *38*, 375-387.
63. Pennuto, M., Palazzolo, I., and Poletti, A. (2009). Post-translational modifications of expanded polyglutamine proteins: impact on neurotoxicity. Human Molecular Genetics *18*, R40-47-R40-47.
64. Borrell-Pages, M., Zala, D., Humbert, S., and Saudou, F. (2006). Huntington's disease: from huntingtin function and dysfunction to therapeutic strategies. Cellular and Molecular Life Sciences *63*, 2642-2660.
65. Katsuno, M., Adachi, H., Kume, A., Li, M., Nakagomi, Y., Niwa, H., Sang, C., Kobayashi, Y., Doyu, M., and Sobue, G. (2002). Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. Neuron *35*, 843-854.
66. Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y., and Orr, H.T. (1998). Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. Cell *95*, 41-53.
67. Graham, R.K., Deng, Y., Slow, E.J., Haigh, B., Bissada, N., Lu, G., Pearson, J., Shehadeh, J., Bertram, L., Murphy, Z., et al. (2006). Cleavage at the Caspase-6 Site Is Required for Neuronal Dysfunction and Degeneration Due to Mutant Huntingtin. Cell *125*, 1179-1191.
68. Ranum, L.P.W., and Day, J.W. (2004). Pathogenic RNA repeats: an expanding role in genetic disease. Trends in Genetics *20*, 506-512.
69. Cho, D.H., and Tapscott, S.J. (2007). Myotonic dystrophy: emerging mechanisms for DM1 and DM2. Biochimica Et Biophysica Acta *1772*, 195-204.
70. Lee, J.E., and Cooper, T.A. (2009). Pathogenic mechanisms of myotonic dystrophy. Biochemical Society Transactions *37*, 1281-1286.
71. Fu, Y., Pizzuti, A., Fenwick, R., Jr, King, J., Rajnarayan, S., Dunne, P., Dubel, J., Nasser, G., Ashizawa, T., de Jong, P., et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science *255*, 1256-1258.
72. Jansen, G., Groenen, P.J., Bächner, D., Jap, P.H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J.J., et al. (1996). Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. Nature Genetics *13*, 316-324.
73. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., and Thornton, C.A. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. Science *289*, 1769-1773.
74. Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*, 864-867
75. Taneja, K., McCurrach, M., Schalling, M., Housman, D., and Singer, R. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. Journal of Cell Biology *128*, 995-1002.

76. Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., and Housman, D.E. (1997). Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. PNAS *94*, 7388-7393.
77. Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. Embo Journal *19*, 4439-4448.
78. Mooers, B.H.M., Logue, J.S., and Berglund, J.A. (2005). The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. PNAS *102*, 16626-16631.
79. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M., and Baylies, M.K. (1998). The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2. Developmental Biology *195*, 131-143.
80. Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M., and Mlodzik, M. (1997). muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. Development *124*, 4321-4331.
81. Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006). The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. Differentiation *74*, 65-80.
82. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004). Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. Human Molecular Genetics *13*, 495-507.
83. Yuan, Y., Compton, S.A., Sobczak, K., Stenberg, M.G., Thornton, C.A., Griffith, J.D., and Swanson, M.S. (2007). Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. Nucleic Acids Research, *35*, 5474-5486.
84. Ranum, L.P.W., and Cooper, T.A. (2006). RNA-mediated neuromuscular disorders. Annual Review of Neuroscience *29*, 259-277.
85. Mankodi, A., Urbinati, C.R., Yuan, Q.-P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S., and Thornton, C.A. (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Human Molecular Genetics *10*, 2165-2170.
86. Chen, K.-Y., Pan, H., Lin, M.-J., Li, Y.-Y., Wang, L.-C., Wu, Y.-C., and Hsiao, K.-M. (2007). Length-dependent toxicity of untranslated CUG repeats on Caenorhabditis elegans. Biochemical and Biophysical Research Communications *352*, 774-779.
87. de Haro, M., Al-Ramahi, I., Gouyon, B.D., Ukani, L., Rosa, A., Faustino, N.A., Ashizawa, T., Cooper, T.A., and Botas, J. (2006). MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of Myotonic Dystrophy Type 1. Human Molecular Genetics *15*, 2138-2145.
88. Ho, T.H., Savkur, R.S., Poulos, M.G., Mancini, M.A., Swanson, M.S., and Cooper, T.A. (2005). Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. Journal of Cell Science *118*, 2923-2933.

89. Garcia-Lopez, A., Monferrer, L., Garcia-Alcover, I., Vicente-Crespo, M., Alvarez-Abril, M.C., and Artero, R.D. (2008). Genetic and chemical modifiers of a CUG toxicity model in Drosophila. PloS One *3*, e1595-e1595.
90. Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W., and Swanson, M.S. (2003). A Muscleblind Knockout Model for Myotonic Dystrophy. Science *302*, 1978-1980.
91. Machuca-Tzili, L., Thorpe, H., Robinson, T., Sewry, C., and Brook, J. (2006). Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts. Human Genetics *120*, 487-499.
92. Kanadia, R.N., Shin, J., Yuan, Y., Beattie, S.G., Wheeler, T.M., Thornton, C.A., and Swanson, M.S. (2006). Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. PNAS *103*, 11748-11753.
93. Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L., Roberts, R., Caskey, C.T., and Swanson, M.S. (1996). Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. Nucleic Acids Research *24*, 4407-4414.
94. Mankodi, A., Teng-Umnuay, P., Krym, M., Henderson, D., Swanson, M., and Thornton, C.A. (2003). Ribonuclear inclusions in skeletal muscle in myotonic dystrophy types 1 and 2. Annals of Neurology *54*, 760-768.
95. Kuyumcu-Martinez, N.M., Wang, G.-S., and Cooper, T.A. (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. Molecular Cell *28*, 68-78.
96. Ward, A.J., Rimer, M., Killian, J.M., Dowling, J.J., and Cooper, T.A. (2010). CUGBP1 overexpression in mouse skeletal muscle reproduces features of myotonic dystrophy type 1. Human Molecular Genetics *19*, 3614-3622.
97. Perini, G.I., Menegazzo, E., Ermani, M., Zara, M., Gemma, A., Ferruzza, E., Gennarelli, M., and Angelini, C. (1999). Cognitive impairment and (CTG)n expansion in myotonic dystrophy patients. Biological Psychiatry *46*, 425-431.
98. Sistiaga, A., Urreta, I., Jodar, M., Cobo, A.M., Emparanza, J., Otaegui, D., Poza, J.J., Merino, J.J., Imaz, H., Marti-Masso, J.F., et al. (2010). Cognitive/personality pattern and triplet expansion size in adult myotonic dystrophy type 1 (DM1): CTG repeats, cognition and personality in DM1. Psychological Medicine *40*, 487-495.
99. Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T., and Thornton, C.A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. Human Molecular Genetics *13*, 3079-3088.
100. Dhaenens, C.M., Schraen-Maschke, S., Tran, H., Vingtdeux, V., Ghanem, D., Leroy, O., Delplanque, J., Vanbrussel, E., Delacourte, A., Vermersch, P., et al. (2008). Overexpression of MBNL1 fetal isoforms and modified splicing of Tau in the DM1 brain: two individual consequences of CUG trinucleotide repeats. Experimental Neurology *210*, 467-478.
101. Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology *57*, 127-130.

102. Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., and Warren, S.T. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. Human Molecular Genetics *1*, 397-400.
103. Jin, P., Zarnescu, D.C., Zhang, F., Pearson, C.E., Lucchesi, J.C., Moses, K., and Warren, S.T. (2003). RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. Neuron *39*, 739-747.
104. Willemsen, R., Hoogeveen-Westerveld, M., Reis, S., Holstege, J., Severijnen, L.-A.W.F.M., Nieuwenhuizen, I.M., Schrier, M., van Unen, L., Tassone, F., Hoogeveen, A.T., et al. (2003). The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. Human Molecular Genetics *12*, 949-959.
105. Hashem, V., Galloway, J.N., Mori, M., Willemsen, R., Oostra, B.A., Paylor, R., and Nelson, D.L. (2009). Ectopic expression of CGG containing mRNA is neurotoxic in mammals. Human Molecular Genetics *18*, 2443-2451.
106. Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del Bigio, M.R., Jacquemont, S., Leehey, M., and Hagerman, P.J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. Brain *125*, 1760-1771.
107. Iwahashi, C.K., Yasui, D.H., An, H.-J., Greco, C.M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C.B., Hagerman, R.J., and Hagerman, P.J. (2006). Protein composition of the intranuclear inclusions of FXTAS. Brain *129*, 256-271.
108. Kiliszek, A., Kierzek, R., Krzyzosiak, W.J., and Rypniewski, W. (2011). Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. Nucleic Acids Research. First published online: May 19, 2011, doi: 10.1093/nar/gkr368.
109. Jin, P., Duan, R., Qurashi, A., Qin, Y., Tian, D., Rosser, T.C., Liu, H., Feng, Y., and Warren, S.T. (2007). Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron *55*, 556-564.
110. Sofola, O.A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., Nelson, D.L., and Botas, J. (2007). RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron *55*, 565-571.
111. Khalili, K., Del Valle, L., Muralidharan, V., Gault, W.J., Darbinian, N., Otte, J., Meier, E., Johnson, E.M., Daniel, D.C., Kinoshita, Y., et al. (2003). Puralpha is essential for postnatal brain development and developmentally coupled cellular proliferation as revealed by genetic inactivation in the mouse. Molecular and Cellular Biology *23*, 6857-6875.
112. Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nature Reviews Molecular Cell Biology *3*, 195-205.
113. Qurashi, A., Li, W., Zhou, J.-Y., Peng, J., and Jin, P. (2011). Nuclear Accumulation of Stress Response mRNAs Contributes to the Neurodegeneration Caused by Fragile X Premutation rCGG Repeats. PLoS Genetics *7*, e1002102-e1002102.
114. Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R.K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D.J., et al. (2011). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J *29*, 1248-1261

115. Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W., and Ranum, L.P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nature Genetics *21*, 379-384
116. Mutsuddi, M., Marshall, C.M., Benzow, K.A., Koob, M.D., and Rebay, I. (2004). The Spinocerebellar Ataxia 8 Noncoding RNA Causes Neurodegeneration and Associates with Staufen in Drosophila. Current Biology *14*, 302-308.
117. Villacé, P., Marión, R.M., and Ortín, J. (2004). The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs. Nucleic Acids Research *32*, 2411-2420.
118. Saunders, L.R., and Barber, G.N. (2003). The dsRNA binding protein family: critical roles, diverse cellular functions. FASEB Journal *17*, 961-983.
119. Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nature Genetics *38*, 758-769.
120. Daughters, R.S., Tuttle, D.L., Gao, W., Ikeda, Y., Moseley, M.L., Ebner, T.J., Swanson, M.S., and Ranum, L.P.W. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. PLoS Genetics *5*, e1000600-e1000600.
121. Greenstein, P.E., Jean-Paul G. Vonsattel, Russell L. Margolis, Jeffrey T. Joseph, (2007). Huntington's disease like-2 neuropathology. Movement Disorders *22*, 1416-1423.
122. Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T., et al. (2001). A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. Nature Genetics *29*, 377-378.
123. Rudnicki, D.D.S.E.H., Mark W. Lin, Charles A. Thornton, Christopher A. Ross, Russell L. Margolis, (2007). Huntington's disease-like 2 is associated with CUG repeat-containing RNA foci. Annals of Neurology *61*, 272-282.
124. Wilburn, B., Rudnicki, D.D., Zhao, J., Weitz, T.M., Cheng, Y., Gu, X., Greiner, E., Park, C.S., Wang, N., Sopher, B.L., et al. (2011). An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron *70*, 427-440.
125. Margolis, R.L., Rudnicki, D.D., and Holmes, S.E. (2005). Huntington's disease like-2: review and update. Acta Neurologica Taiwanica *14*, 1-8.
126. Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L., et al. (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nature Genetics *26*, 191-194.
127. Wakamiya, M., Matsuura, T., Liu, Y., Schuster, G.C., Gao, R., Xu, W., Sarkar, P.S., Lin, X., and Ashizawa, T. (2006). The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. Neurology *67*, 607-613.
128. Lin, X., and Ashizawa, T. (2005). Recent progress in spinocerebellar ataxia type-10 (SCA10). Cerebellum *4*, 37-42.

129. White, M.C., Gao, R., Xu, W., Mandal, S.M., Lim, J.G., Hazra, T.K., Wakamiya, M., Edwards, S.F., Raskin, S., Teive, H.l.A.G., et al. (2010). Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCdelta to mitochondria in spinocerebellar ataxia 10. PLoS Genetics *6*, e1000984-e1000984.
130. van Eyk, C.L. (2010). Investigation of RNA-mediated pathogenic pathways in a *Drosophila* model of expanded repeat disease. PhD Thesis, University of Adelaide.
131. Holmes, S.E., O'Hearn, E.E., McInnis, M.G., Gorelick-Feldman, D.A., Kleiderlein, J.J., Callahan, C., Kwak, N.G., Ingersoll-Ashworth, R.G., Sherr, M., Sumner, A.J., et al. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nature Genetics *23*, 391-392.
132. Li, L.-B., Yu, Z., Teng, X., and Bonini, N.M. (2008). RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature *453*, 1107-1111.
133. de Mezer, M., Wojciechowska, M., Napierala, M., Sobczak, K., and Krzyzosiak, W.J. (2011). Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. Nucleic Acids Research. First published online: January 18, 2011, doi: 10.1093/nar/gkq1323.
134. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J., and Krzyzosiak, W.J. (2003). RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Research *31*, 5469-5482.
135. Jasinska, A., Michlewski, G., de Mezer, M., Sobczak, K., Kozlowski, P., Napierala, M., and Krzyzosiak, W.J. (2003). Structures of trinucleotide repeats in human transcripts and their functional implications. Nucleic Acids Research *31*, 5463 -5468.
136. Michlewski, G., and Krzyzosiak, W.J. (2004). Molecular architecture of CAG repeats in human disease related transcripts. Journal of Molecular Biology *340*, 665-679.
137. Kiliszek, A., Kierzek, R., Krzyzosiak, W.J., and Rypniewski, W. (2010). Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic Acids Research *38*, 8370-8376.
138. Handa, V., Yeh, H.J.C., McPhie, P., and Usdin, K. (2005). The AUUCU repeats responsible for spinocerebellar ataxia type 10 form unusual RNA hairpins. The Journal of Biological Chemistry *280*, 29340-29345.
139. Houseley, J.M., Wang, Z., Brock, G.J.R., Soloway, J., Artero, R., Perez-Alonso, M., O'Dell, K.M.C., and Monckton, D.G. (2005). Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila. Human Molecular Genetics *14*, 873-883.
140. Mahadevan, M.S., Yadava, R.S., Yu, Q., Balijepalli, S., Frenzel-McCardell, C.D., Bourne, T.D., and Phillips, L.H. (2006). Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. Nature Genetics *38*, 1066-1070.
141. Onishi, H., Kino, Y., Morita, T., Futai, E., Sasagawa, N., and Ishiura, S. (2008). MBNL1 associates with YB-1 in cytoplasmic stress granules. Journal of Neuroscience Research *86*, 1994-2002.

142. de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N.A., Ashizawa, T., Cooper, T.A., and Botas, J. (2006). MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. Human Molecular Genetics *15*, 2138-2145.
143. Du, H., Cline, M.S., Osborne, R.J., Tuttle, D.L., Clark, T.A., Donohue, J.P., Hall, M.P., Shiue, L., Swanson, M.S., Thornton, C.A., et al. (2010). Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. Nature Structural & Molecular Biology *17*, 187-193.
144. Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K., and Junghans, R.P. (2004). RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. Science *303*, 383-387.
145. Junghans, R.P. (2009). Dystrophia myotonia: why focus on foci? European Journal of Human Genetics: EJHG *17*, 543-553
146. Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.
147. Okamura, K., Chung, W.-J., and Lai, E.C. (2008). The long and short of inverted repeat genes in animals: MicroRNAs, mirtrons and hairpin RNAs. Cell Cycle *7*, 2840-2845.
148. Mattick, J.S. (2005). Small regulatory RNAs in mammals. Human Molecular Genetics *14*, R121-R132.
149. Handa, V., Saha, T., and Usdin, K. (2003). The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. Nucleic Acids Research *31*, 6243-6248.
150. Krol, J., Fiszer, A., Mykowska, A., Sobczak, K., de Mezer, M., and Krzyzosiak, W.J. (2007). Ribonuclease Dicer Cleaves Triplet Repeat Hairpins into Shorter Repeats that Silence Specific Targets. Molecular Cell *25*, 575-586.
151. Batra, R., Charizanis, K., and Swanson, M.S. (2010). Partners in crime: bidirectional transcription in unstable microsatellite disease. Human Molecular Genetics *19*, R77-82.
152. Todd, P.K., and Paulson, H.L. (2010). RNA-mediated neurodegeneration in repeat expansion disorders. Annals of Neurology *67*, 291-300.
153. Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N., and Tapscott, S.J. (2005). Antisense Transcription and Heterochromatin at the DM1 CTG Repeats Are Constrained by CTCF. Molecular Cell *20*, 483-489.
154. Ladd, P.D., Smith, L.E., Rabaia, N.A., Moore, J.M., Georges, S.A., Hansen, R.S., Hagerman, R.J., Tassone, F., Tapscott, S.J., and Filippova, G.N. (2007). An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Human Molecular Genetics *16*, 3174-3187.
155. Sopher, B.L., Ladd, P.D., Pineda, V.V., Libby, R.T., Sunkin, S.M., Hurley, J.B., Thienes, C.P., Gaasterland, T., Filippova, G.N., and La Spada, A.R. (2011). CTCF Regulates Ataxin-7 Expression through Promotion of a Convergently Transcribed, Antisense Noncoding RNA. Neuron *70*, 1071-1084.
156. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564-1566.

157. Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nature Reviews. Molecular Cell Biology *10*, 637-643.
158. Kumari, D., and Usdin, K. (2009). Chromatin Remodeling in the Noncoding Repeat Expansion Diseases. Journal of Biological Chemistry *284*, 7413 -7417.
159. Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V., and Tapscott, S.J. (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nature Genetics *28*, 335-343.
160. Bilen, J., Liu, N., Burnett, B.G., Pittman, R.N., and Bonini, N.M. (2006). MicroRNA Pathways Modulate Polyglutamine-Induced Neurodegeneration. Molecular Cell *24*, 157-163.
161. Gambardella, S., Rinaldi, F., Lepore, S.M., Viola, A., Loro, E., Angelini, C., Vergani, L., Novelli, G., and Botta, A. (2010). Overexpression of microRNA-206 in the skeletal muscle from myotonic dystrophy type 1 patients. Journal of Translational Medicine *8*, 48.
162. Perbellini, R., Greco, S., Sarra-Ferraris, G., Cardani, R., Capogrossi, M.C., Meola, G., and Martelli, F. (2011). Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1. Neuromuscular Disorders: NMD *21*, 81-88.
163. Lee, S.-T., Chu, K., Im, W.-S., Yoon, H.-J., Im, J.-Y., Park, J.-E., Park, K.-H., Jung, K.-H., Lee, S.K., Kim, M., et al. (2011). Altered microRNA regulation in Huntington's disease models. Experimental Neurology *227*, 172-179.
164. Schaefer, A., O'Carroll, D., Tan, C.L., Hillman, D., Sugimori, M., Llinas, R., and Greengard, P. (2007). Cerebellar neurodegeneration in the absence of microRNAs. J. Exp. Med. *204*, 1553-1558.
165. Sofola, O.A., Jin, P., Botas, J., and Nelson, D.L. (2007). Argonaute-2-dependent rescue of a Drosophila model of FXTAS by FRAXE premutation repeat. Human Molecular Genetics *16*, 2326-2332.
166. Lawlor, K.T., O'Keefe, L.V., Samaraweera, S.E., van Eyk, C.L., McLeod, C.J., Maloney, C.A., Dang, T.H.Y., Suter, C.M., and Richards, R.I. (2011). Double-stranded RNA is pathogenic in Drosophila models of expanded repeat neurodegenerative diseases. Human Molecular Genetics *20*, 3757-3768.
167. Yu, Z., Teng, X., and Bonini, N.M. (2011). Triplet Repeat-Derived siRNAs Enhance RNA-Mediated Toxicity in a Drosophila Model for Myotonic Dystrophy. PLoS Genet *7*, e1001340.
168. Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M.r., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. PNAS *108*, 260-265.
169. McLaughlin, B.A., Spencer, C., and Eberwine, J. (1996). CAG trinucleotide RNA repeats interact with RNA-binding proteins. American Journal of Human Genetics *59*, 561-569.
170. Hsu, R.-J., Hsiao, K.-M., Lin, M.-J., Li, C.-Y., Wang, L.-C., Chen, L.-K., and Pan, H. (2011). Long tract of untranslated CAG repeats is deleterious in transgenic mice. PloS One *6*, e16417-e16417.
171. Wang, L.-C., Chen, K.-Y., Pan, H., Wu, C.-C., Chen, P.-H., Liao, Y.-T., Li, C., Huang, M.-L., and Hsiao, K.-M. (2011). Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in Caenorhabditis elegans. Cellular and Molecular Life Sciences 68, 1255-1267.

172. Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.
173. Bilen, J., and Bonini, N.M. (2005). Drosophila as a model for human neurodegenerative disease. Annual Review of Genetics *39*, 153-171.
174. Jackson, G.R. (2008). Guide to Understanding Drosophila Models of Neurodegenerative Diseases. PLoS Biol *6*, e53.
175. Warrick, J.M., Paulson, H.L., Gray-Board, G.L., Bui, Q.T., Fischbeck, K.H., Pittman, R.N., and Bonini, N.M. (1998). Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in Drosophila. Cell *93*, 939-949.
176. Jackson, G.R., Salecker, I., Dong, X., Yao, X., Arnheim, N., Faber, P.W., MacDonald, M.E., and Zipursky, S.L. (1998). Polyglutamine-expanded human huntingtin transgenes induce degeneration of Drosophila photoreceptor neurons. Neuron *21*, 633-642.
177. McLeod, C.J. (2006). Investigation of the pathogenic agent in a *Drosophila* model of polyglutamine disease. PhD Thesis, University of Adelaide.
178. Lawlor, K.T. (2006). Investigation of untranslated trinucleotide repeat expression in *Drosophila* neurons. Honours Thesis, University of Adelaide.
179. Ito, K., Awano, W., Suzuki, K., Hiromi, Y., and Yamamoto, D. (1997). The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. Development *124*, 761-771.
180. Wodarz, A., Hinz, U., Engelbert, M., and Knust, E. (1995). Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of drosophila. Cell *82*, 67-76.
181. Sekyrova, P., Bohmann, D., Jindra, M., and Uhlirova, M. (2010). Interaction between Drosophila bZIP proteins Atf3 and Jun prevents replacement of epithelial cells during metamorphosis. Development *137*, 141-150.
182. Cherbas, L., Hu, X., Zhimulev, I., Belyaeva, E., and Cherbas, P. (2003). EcR isoforms in Drosophila: testing tissue-specific requirements by targeted blockade and rescue. Development *130*, 271-284.
183. Lin, D.M., and Goodman, C.S. (1994). Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. Neuron *13*, 507-523.
184. Freeman, M. (1996). Reiterative Use of the EGF Receptor Triggers Differentiation of All Cell Types in the Drosophila Eye. Cell *87*, 651-660.
185. Harrison, D.A., Binari, R., Nahreini, T.S., Gilman, M., and Perrimon, N. (1995). Activation of a Drosophila Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. The EMBO Journal *14*, 2857-2865.
186. Kopp, A., and Duncan, I. (2002). Anteroposterior patterning in adult abdominal segments of Drosophila. Developmental Biology *242*, 15-30.
187. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. Cell *117*, 69-81.
188. Bateman, J.R., Lee, A.M., and Wu, C.t. (2006). Site-Specific Transformation of Drosophila via øC31 Integrase-Mediated Cassette Exchange. Genetics *173*, 769-777.
189. Sambrook, J. (2001). Molecular Cloning: A Laboratory Manual, Third Edition, (Cold Spring Harbor Laboratory Press).

190. Sullivan, W., Ashburner, M., and Hawley, R.S. (2000). Drosophila Protocols. CSHL Press.
191. Rudolph, T., Lu, B., Westphal, T., Szidonya, J., Eissenberg, J., and Reuter, G. (1999). New type of CyO and TM3 green balancers. Drosophila Information Service *82*, 99-100.
192. Madhavan, K., and Madhavan, M.M. (1995). Defects in the adult abdominal integument ofDrosophila caused by mutations intorpedo, a DER homolog. Roux's Archives of Developmental Biology *204*, 330-335.
193. Ninov, N., Chiarelli, D.A., and Martin-Blanco, E. (2007). Extrinsic and intrinsic mechanisms directing epithelial cell sheet replacement during Drosophila metamorphosis. Development *134*, 367-379.
194. Minakuchi, C., Zhou, X., and Riddiford, L.M. (2008). Kruppel homolog 1 (Kr-h1) mediates juvenile hormone action during metamorphosis of Drosophila melanogaster. Mechanisms of Development *125*, 91-105.
195. Bischoff, M., and Cseresnyes, Z. (2009). Cell rearrangements, cell divisions and cell death in a migrating epithelial sheet in the abdomen of Drosophila. Development *136*, 2403-2411.
196. Katti, M.V., Ranjekar, P.K., and Gupta, V.S. (2001). Differential Distribution of Simple Sequence Repeats in Eukaryotic Genome Sequences. Molecular Biology and Evolution *18*, 1161-1167.
197. Sobczak, K., Michlewski, G., de Mezer, M., Kierzek, E., Krol, J., Olejniczak, M., Kierzek, R., and Krzyzosiak, W.J. (2010). Structural diversity of triplet repeat RNAs. The Journal of Biological Chemistry *285*, 12755-12764.
198. Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000). A-to-I Pre-mRNA Editing in Drosophila Is Primarily Involved in Adult Nervous System Function and Integrity. Cell *102*, 437-449.
199. Robinson, D.N., Smith-Leiker, T.A., Sokol, N.S., Hudson, A.M., and Cooley, L. (1997). Formation of the Drosophila ovarian ring canal inner rim depends on cheerio. Genetics *145*, 1063-1072.
200. Sokol, N.S., and Cooley, L. (1999). Drosophila filamin encoded by the cheerio locus is a component of ovarian ring canals. Current Biology *9*, 1221-1230.
201. Li, M.-g., Serr, M., Edwards, K., Ludmann, S., Yamamoto, D., Tilney, L.G., Field, C.M., and Hays, T.S. (1999). Filamin Is Required for Ring Canal Assembly and Actin Organization during Drosophila Oogenesis. The Journal of Cell Biology *146*, 1061-1074.
202. Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference - a crash course. Trends in Genetics *21*, 339-345.
203. Guo, Y., Zhang, S.X., Sokol, N., Cooley, L., and Boulianne, G.L. (2000). Physical and genetic interaction of filamin with presenilin in Drosophila. Journal of Cell Science *113 Pt 19*, 3499-3508.
204. Shulman, J.M., and Feany, M.B. (2003). Genetic modifiers of tauopathy in Drosophila. Genetics *165*, 1233-1242.
205. Blard, O., Feuillette, S., Bou, J., Chaumette, B., Frebourg, T., Campion, D., and Lecourtois, M. (2007). Cytoskeleton proteins are modulators of mutant tau-induced neurodegeneration in Drosophila. Human Molecular Genetics *16*, 555-566.
206. Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., et al. (2009). FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Research *37*, D555-D559.

207. Dubnau, J., Chiang, A.-S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., et al. (2003). The staufen/pumilio pathway is involved in Drosophila long-term memory. Current Biology *13*, 286-296.
208. Bolduc, F.o.V., Bell, K., Rosenfelt, C., Cox, H., and Tully, T. (2010). Fragile x mental retardation 1 and filamin a interact genetically in Drosophila long-term memory. Frontiers in Neural Circuits *3*, 22-22.
209. Le Mee, G., Ezzeddine, N., Capri, M.l., and Ait-Ahmed, O. (2008). Repeat length and RNA expression level are not primary determinants in CUG expansion toxicity in Drosophila models. PloS One *3*, e1466.
210. Kramer, J.M., and Staveley, B.E. (2003). GAL4 causes developmental defects and apoptosis when expressed in the developing eye of Drosophila melanogaster. Genetics and Molecular Research *2*, 43-47.
211. Franceschini, N., and Kirschfeld, K. (1971). Les phenomenes de pseudopupille dans l'oeil compose deDrosophila. Kybernetik *9*, 159-182.
212. Landis, G.N., Bhole, D., and Tower, J. (2003). A search for doxycycline-dependent mutations that increase Drosophila melanogaster life span identifies the VhaSFD, Sugar baby, filamin, fwd and Cctl genes. Genome Biology *4*, R8-R8.
213. Gargano, J.W., Martin, I., Bhandari, P., and Grotewiel, M.S. (2005). Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in Drosophila. Experimental Gerontology *40*, 386-395.
214. Borst, A. (2009). Drosophila's View on Insect Vision. Current Biology *19*, R36-R47.
215. Zuker, C.S. (1996). The biology of vision of Drosophila. PNAS *93*, 571 -576.
216. Sheen, V.L., Jansen, A., Chen, M.H., Parrini, E., Morgan, T., Ravenscroft, R., Ganesh, V., Underwood, T., Wiley, J., Leventer, R., et al. (2005). Filamin A mutations cause periventricular heterotopia with Ehlers-Danlos syndrome. Neurology *64*, 254-262.
217. Masruha, M.R., Caboclo, L.O.S.F., Carrete, H., Jr., Cendes, I.L., Rodrigues, M.G., Garzon, E., Yacubian, E.M.T., Sakamoto, A.r.C., Sheen, V., Harney, M., et al. (2006). Mutation in filamin A causes periventricular heterotopia, developmental regression, and West syndrome in males. Epilepsia *47*, 211-214.
218. Sokol, N.S., and Cooley, L. (2003). Drosophila filamin is required for follicle cell motility during oogenesis. Developmental Biology *260*, 260-272.
219. Pearson, C.E. (2011). Repeat associated non-ATG translation initiation: one DNA, two transcripts, seven reading frames, potentially nine toxic entities! PLoS Genetics *7*, e1002018-e1002018.
220. Lin, Y., Leng, M., Wan, M., and Wilson, J.H. (2010). Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. Molecular and Cellular Biology *30*, 4435-4451.
221. Kalsotra, A., Wang, K., Li, P.-F., and Cooper, T.A. (2010). MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. Genes & Development *24*, 653-658.

Corrections

Chapter 1
Page 10, paragraph 2 should read “rather *than* enhancement”

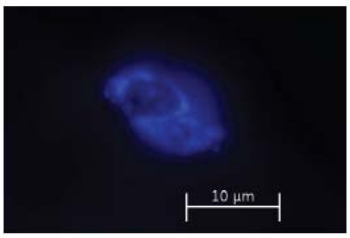
Chapter 2
Page 35, **Quantification of tergite disruption**, should include the paragraph:
The scoring scheme was based on the number and severity of disrupted tergites, using particular morphological attributes to define each category, thus minimising any experimenter bias. Preliminary data showed no significant difference (data not shown) between populations when scoring 'experimenter blind'. As such, remaining experiments were not scored blind. The order in which genotypes were scored each day was randomised and data from multiple sets of progeny obtained from multiple sets of parents on different days was used in each case.

Page 36, **Quantification of locomotion phenotype**, should include the paragraph:
Scoring involved reviewing the video to tally the time in seconds that each fly spent either upright (walking or standing) or on its back As the possibility for experimenter bias in this case appeared negligible scoring was not done 'blind'.

Page 40, **Climbing assays**, should include the clarification :
n = 3 biological replicates (sets), with 20-25 animals per genotype, per biological replicate (set), for a total of 60-75 animals examined for each genotype. A climbing score representing each biological replicate (set) was obtained by calculating the mean from 5 consecutive trials for each genotype. A final genotype score was obtained by calculating the mean of all 3 biological replicates.

Chapter 3
Page 46, **Figure 3.1** legend should include the paragraph:
Fisher's exact test does not include a calculation of standard deviation, or standard error, however 95% confidence intervals were calculated for each particular proportion. As this involved a separate calculation these values are included in Appendix 1, rather than as error bars.

Chapter 4
Page 69, In **Figure 4.1 C**, DAPI staining was poorly reproduced in the printed version. Images were chosen based on being representative of each genotype in regard to repeat RNA staining (Cy3 signal), with DAPI included as a guide to the location of the nucleus only. As such the relative levels of DAPI signal do not change the interpretation of the data. A modified version (to improve visibility in printed form) of the DAPI staining shown in 4.1 C is included below.

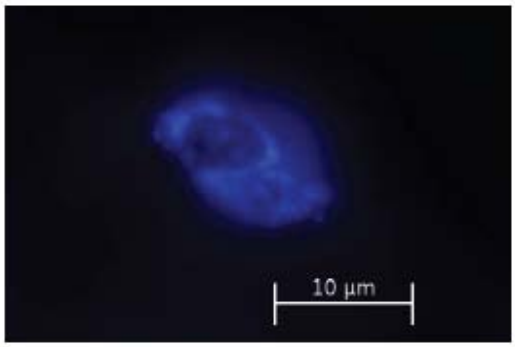


Page 77, paragraph 1, should include the paragraph:
In this study CUG-specific RNA localisation patterns were observed in independent samples from independent transgenic lines and thus the result appears robust. However, as quantification of foci was not performed, further analysis would be necessary to confirm the more subtle differences in CUG-specific localisation patterns observed in different repeat expression contexts.

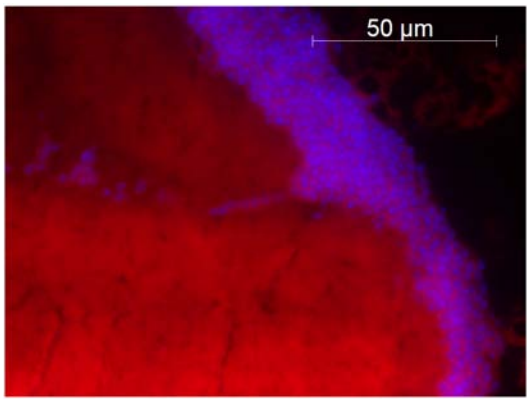
Page 80, paragraph 2, should include the sentences:
Confocal microscopy was not performed in this case. Techniques allowing higher imaging resolution may confirm the absence of neuronal foci in *Drosophila* with more certainty.

Scale bars were initially not included in fluorescent micrographs. Examples for each type of image taken are included below to aid in interpretation of these results.

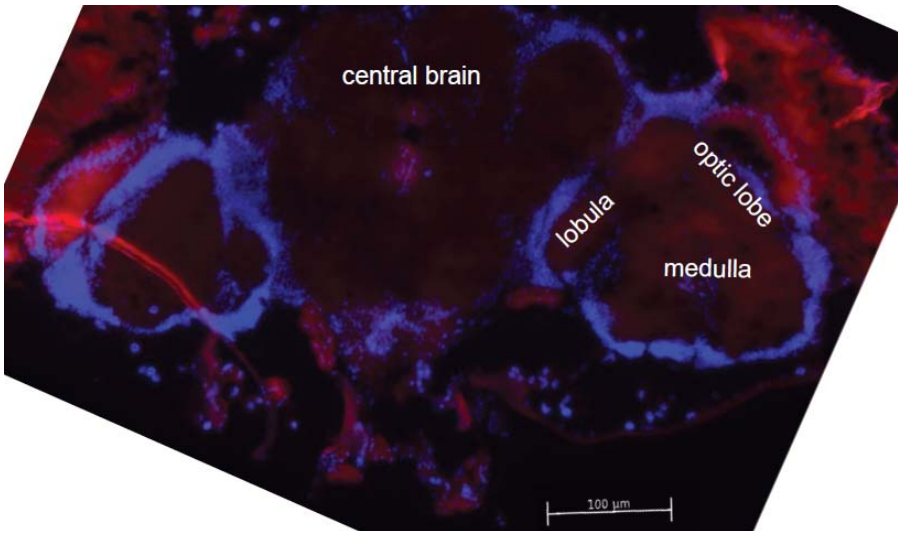
An example of muscle nuclei (As in 4.1, 4.3, 4.4, 4.5). All images were captured and cropped in the same way such that scale is identical :



An example of an adult brain at higher magnification (as in 4.6 B-D) :



An example of an adult brain at lower magnification (as in 4.6 A). In this case landmarks within the brain are annotated to further aid in interpretation.



Page 82, paragraph 1, should read :
"... support *the conclusion* that pathways ….."

Chapter 5
Page 89, paragraph 3, should read:
" … indicate that rather *than* the insertion directly disrupting …."

Chapter 6
Page 118, Figure 6.2 figure legend, should state:
All flies were aged for 35 days before sectioning (Materials and Methods).

Chapter 7
Page 135, paragraph 2, should read:
"In support *of* this we see alterations to miRNA profiles...."

Page 135, paragraph 3, should read:
".... indicating that, as in our model, complementary transcripts form double-stranded RNA that is processed."