Whole-genome analysis of the transcriptional network underlying male sexual behaviour in *Drosophila melanogaster*.

A thesis submitted for the degree of Doctor of Philosophy at the University of Oxford

By

Elizabeth L. Ashley

Linacre College
Department of Physiology, Anatomy and Genetics
University of Oxford, South Parks Road,
Oxford, IX1 3PT

Trinity Term 2012

Supervisor: Dr. Stephen F. Goodwin

Abstract

The robust behavioural courtship ritual displayed by *Drosophila melanogaster* males is governed by their underlying nervous system (NS). Two key genes of the Sex Determination Hierarchy, *fruitless (fru)* and *doublesex (dsx),* determine most neuronal substrates for sexual behaviour. In this study we aim to better understand the role *fru* plays in determining these neural substrates, as a means of better understanding the relationships between brain, behaviour and genes, and thus how the development of neuronal networks shape innate and species-specific behaviours.

fru has two major functions: control of male sexual behaviour, and viability in both sexes. Alternative splicing of *fru* produces transcription factors driven by four promoters: P1 transcripts are sex-specifically spliced (only viable in the male), and P2-4 transcripts are crucial to both sexes survival. The resulting proteins contain a BTB protein-protein interaction domain at the N-terminus, and one of four C-terminal zinc-finger (ZnF) DNA binding domains. Male-specific proteins (Fru^{M}) contain an additional 101 amino acid N-terminal domain, and one of three alternative C_2H_2 ZnF domains (Fru^{MA}, Fru^{MB}, Fru^{MC}). These male-specific isoforms are expressed in the central nervous system (CNS) beginning in the late stage larvae (L3), peaking during pupation and on into adulthood. Little is known, however, about the roles of the individual isoforms, and no clear transcriptional targets have been identified. The central aims of this thesis are to document the wild-type expression patterns of Fru isoforms throughout development in the CNS, create and characterise isoform-specific mutants, and to identify and evaluate putative transcriptional targets of Fru. Combining these findings will lead to a better understanding of the underlying molecular functions of individual Fru^{M} isoforms, as a means to understanding their roles in sexual behaviour.

Expression analysis of Fru^{M} isoforms throughout development in the NS is described. To further characterise the role of individual Fru^{M} isoforms, isoform-specific mutants in fru^{A} and fru^{B} exons were generated using site-specific homologous recombination (HR). These novel mutants were validated by PCR and Fru isoform-specific antibody stainings. Mutants were analysed in fru^{M}- and *fru*-null genetic backgrounds, to distinguish the roles of sex-specific vs. common isoforms. These analyses included: fertility, viability and morphology. Fru^{A} was found to have a role in wing extension with possible repercussions for song production. Fru^{B} was found to be developmentally lethal, in addition to having defects in male courtship behaviour.

To understand the role of *fru* in the NS, downstream transcriptional targets of Fru^{M} isoforms were identified. DNA adenine methyltransferase identification (DamID) was used to identify putative transcription targets of Fru^{M} isoforms. The Dam protein methylates DNA in *Drosophila* in a sequence-specific manner allowing targets of Fru to be isolated. Candidate genes were identified using computational analysis (including gene ontology, peak analysis and motif analysis) along with a biologically significant connection with *fru*. The relationship between *fru* and six candidate genes were characterised using RNAi. The results of these studies advance our knowledge of how Fru^{M} isoforms influence the development and physiology of the NS underlying male sexual behaviour in *Drosophila*.

Table of Contents

1 Introduction 15
1.1 The importance of studying behaviour 16
1.2 *Drosophila melanogaster*: A model for identifying and studying genes involved in behaviour. 16
1.2.1 Tools for studying genes that influence behaviour 17
1.2.2 Genes can influence the CNS to generate behaviour 19
1.3 Courtship behaviour. 21
1.3.1 Visual cues 24
1.3.2 Acoustic cues 25
1.3.3 Chemosensory cues 25
1.4 The sex determination Hierarchy 26
1.5 Fruitless is a complex gene that controls male sexual behaviour 29
1.5.1 The sex-specific transcripts of P1 30
1.5.2 The common transcripts: P2 -4 32
1.5.3 The Fru protein 33
1.5.4 Mutating fru 36
1.5.5 Conservation of Fru 40
1.6 Gene targeting by homolgous recombination 41
1.6.1 Ends-out technique 42
1.6.2 Ends-in technique 44
1.7 Putative downstream targets of Fru 46
1.8 DamID as a method to identify putative downstream targets of Fru 48
1.9 Aims and objectives 52

2 Materials and Methods 53
2.1 Drosophila 54
2.1.1 Drosophila stocks: 54
2.1.2 Wild-type strain 56
2.1.3 Rearing conditions 57
2.1.4 Drosophila microinjection 57
2.1.5 Collection of tissues: Dissecting techniques 60
2.2 Visualisation techniques 63
2.2.1 Immunoflourescence 63
2.2.2 Confocal microscopy 64
2.2.3 Image processing 65
2.2.4 Processing of images for figures 65
2.3 Bacterial protocols 66
2.3.1 Strains and plasmids 66
2.3.2 Culture media 67
2.3.3 Antibiotics and indicators 67
2.3.4 Transformation techniques: 67
2.4 General molecular biology protocols 68
2.4.1 Restriction digest of DNA 68
2.4.2 Miniprep 69
2.4.3 DNA gel electrophoresis 69
2.4.4 Visualisation and photography of gels 69
2.4.5 Gel extraction of DNA fragments 70
2.4.6 Subcloning DNA into plasmid vectors 70
2.4.7 Ligation 70
2.4.8 Ligation of plasmid containing large inserts 70
2.4.9 Southern blot 70
2.4.10 Quantification 74

2.4.11 Polymerase chain reaction (PCR) 75
2.4.12 DNA sequencing 75
2.4.13 Multi-site directed Mutagenesis (mSDM) 77
2.4.14 Buffers: 77
2.5 Bioinformatics 78
2.5.1 Corrections 78
2.5.2 R 79
2.6 Behavioural analysis 81
2.6.1 Complementation 81
2.6.2 Viability 81
2.6.3 Fertility 82
2.6.4 Courtship behaviour 82

3 Homologous Recombination 84
3.1 Introduction 85
3.2 Aims and objectives 86
3.3 Understanding *wild-type* Fru expression 86
3.3.1 Antibody stainings identify expression of Fru^M in wild-type CNS 87
3.4 Mutating the *fru* locus 95
3.4.1 Previous attempts to create null mutations of fru isoforms fru^A and fru^B 95
3.4.2 Ends-in homologous recombination 98
3.5 Targeting the *fru* locus 99
3.5.1 Integration at the target locus 100
3.5.2 Analysis of initial integration events 102
3.5.3 Resolution of the duplication 104
3.5.4 Characterising novel fru isoform mutants 111
3.5.5 Characterising the cantonised fru mutants 120
3.6 Discussion 126

4 Identifying Putative Downstream Targets of Fru using DamID 131
4.1 Introduction 132
4.1.1 DamID v ChIP-chip 134
4.2 Aims and objectives 136
4.3 DNA binding mutants 137
4.3.1 Site-directed mutagenesis primer designs 137
4.3.2 Sub-cloning the fru fragment 138
4.3.3 Multi site-directed mutagenesis 140
4.3.4 Mutagenesis 141
4.3.5 Transformation and sequencing to confirm mutagenesis 142
4.3.6 3-way ligation, microinjection and balancing 143
4.4 DamID 145
4.4.1 Creating fusion lines 146
4.5 Normalising the data 154
4.5.1 Moving average 156
4.5.2 Peak finding 156
4.6 Discarding a replicate 156
4.7 Discussion 157

5 DamID Analysis 160
5.1 Introduction 161
5.1.1 Aims and objectives 162
5.2 Comparison of the datasets 163
5.3 Flymine 164
5.3.1 Defining gene lists 165
5.3.2 Chromosome distribution is even throughout development 166
5.3.3 Tissue specific expression of Fru targets 167
5.3.4 Protein domain analysis shows high homeodomain and immunoglobin enrichment 170

5.3.5 Pathway enrichment 172
5.4 Comparing significant targets between the datasets 174
5.5 Gene ontology and gene list analysis further distinguishes isoform Fru^{MA} from the Fru^{MB} and Fru^{MC} isoforms 175
5.5.1 The Fru^{MA} isoform: A distinct isoform with a physiological role 178
5.5.2 The Fru^{MB} isoform: A functionally broad isoform with a clear role in the nervous system 180
5.5.3 Isoform Fru^{MC}: An isoform with narrow specificity and a clear role in the nervous system 182
5.5.4 GO by time point: Common elements link the isoforms 185
5.5.5 Venny by time point: Gene targets change throughout development....186
5.5.6 DNA binding mutants: identifying the effects of co-factors 189
5.6 The representation of the nervous system gene ontology group 191
5.7 Motif analysis: Identifying Fru DNA-binding sequences 194
5.8 Visualisation of the peaks 196
5.8.1 Signalmap™: Fruitless self-regulates promoters three and four 196
5.8.2 UCSC genome browser 196
5.9 Genes of interest....200
5.9.1 Merging, intersecting and subtracting datasets can identify potential priority genes of interest 200
5.10 Genes of interest: Further analysis 201
5.10.1 Chronologically inappropriate morphogenesis (Chinmo) 201
5.10.2 POU protein domain 2 (pdm2)....203
5.10.3 Kruppel (Kr)....204
5.10.4 flapwing (flw)....205
5.10.5 Longitudinals lacking (Lola) 206
5.10.6 Tramtrack (Ttk) 207
5.10.7 RNAi analysis of fertility 207
5.10.8 Analysis of the MOL 209
5.11 Discussion 210

6 Discussion 216
6.1 Introduction 217
6.2 Homologous recombination mutants define isoform-specific functions of Fru 218
6.2.1 fru^{A} function is necessary for normal courtship behaviour and wing extension 218
6.2.2 fru^{B} mutants are homozygous lethal and show dramatic courtship defects 219
6.2.3 fru^{MC} is Solely Responsible for the Formation of the MOL 221
6.2.4 Summary of HR mutations 221
6.3 DamID data identifies large datasets of putative targets of Fru, which change through development....222
6.3.1 Downstream targets of Fru share common functions 223
6.3.2 The Isoforms of Fru have Individual Functions....224
6.3.3 Fru binds cooperatively with unspecified proteins to further expand its scope of regulation 226
6.3.4 Summary of DamID data 228
6.4 Moving Forward: Using this study as a basis for future investigations....229
6.4.1 Unanswered Questions: Immediate steps 229
6.4.2 Direction: Understanding the underlying network governing Fru function...231
6.5 Concluding Remarks 233

References 234

List of Tables

Table 1-1 Common mutations of the *fru* locus....39
Table 2-1 Drosophila stocks represents an unknown exact allele for *w* or *y*....56
Table 2-2 Primary antibodies used in this study....64
Table 2-3 Secondary antibodies used in this study....64
Table 2-4 Bacterial strains and plasmids....66
Table 2-5 Plasmids used in this study....66
Table 2-6 Primers used in this study.....76
Table 2-7 PCR cycling programme for SDM....77
Table 3-1 Viability of Fru^{219} derived mutations....108
Table 3-2 Viability crosses for the second resolution mutants....112
Table 3-3 Fertility crosses depicting the effect of isoform mutations on sucessful reproduction....115
Table 3-4 Viability of cantonised mutants....121
Table 3-5 Fertility of cantonised HR mutants....123
Table 4-1 Primers required for site directed mutagenesis....138
Table 4-2 Restriction enzymes used to excise *fru* fragments from pUAST-NDam-$fru^{MA/\ MB/\ MC}$....140
Table 4-3 Fragments required for 3-way ligation of the NDam plasmid....144
Table 4-4 Mutations identified using site directed mutagenesis....145
Table 4-5 PCR cycling programme for DamID....152
Table 5-1 Number of genes identified per dataset....165
Table 5-2 Candidate genes of immediate interest....201
Table 5-3 Fertility data for candidate genes using RNAi....208

List of Figures

Figure 1-1 The courtship behaviour of *Drosophila melanogaster* 22
Figure 1-2 The sex determination hierarchy in *Drosophila melanogaster* 28
Figure 1-3 The gene structure and transcripts of *fruitless* 30
Figure 1-4 Muscle of Lawrence..... 34
Figure 1-5 Mutations of the *fru* gene 37
Figure 1-6 Schematic representation of 'ends-out' targeting by homologous recombination 44
Figure 1-7 Schematic representation of 'ends-in' targeting by homologous recombination 46
Figure 2-1 Crossing scheme for balancing chromosomes 59
Figure 3-1 Fru^{M} staining in wild-type larval CNS 89
Figure 3-2 Fruitless staining of the pupal wild-type 91
Figure 3-3 Fruitless staining of the adult wild-type 94
Figure 3-4 Schematic of donor construct for gene targeting of *fru* locus 97
Figure 3-5 I-CreI site 98
Figure 3-6 Molecular representation of gene targeting process at the *fru* locus 100
Figure 3-7 Verification of integration of the donor DNA 103
Figure 3-8 Resolution of the *fru* locus..... 104
Figure 3-9 Verification of targeted events and PCR verification of HR mutants derived from Fru^{219} 108
Figure 3-10 PCR verification of HR mutants derived from $Fru^{80.1}$ 110
Figure 3-11 HR mutant CNS stained with anti-Fru^{A} 116
Figure 3-12 HR mutant CNS stained with anti-Fru^{B} 117
Figure 3-13 Evidence of a MOL in $fru^{\Delta A}$ and $fru^{\Delta B}$ HR mutants 119
Figure 3-14 Malformations associated with $fru^{\Delta B}$ 122
Figure 3-15 Courtship Behaviour Analysis of HR mutants 124
Figure 4-1 ChIP-chip v DamID 136
Figure 4-2 Sub-cloning the *fru* fragment 140
Figure 4-3 Quikchange multi site-directed mutagenesis schematic 141
Figure 4-4 Mutations in the DNA binding domain of *fruitless* 142
Figure 4-5 3-way ligation of mutated *fru* fragment into the *p*UAST-NDam plasmid 143
Figure 4-6 Confirmation of 3-way ligation 145
Figure 4-7 Schematic representation of the NDam-Fru^{M} proteins 148
Figure 4-8 Examples of DamID samples prior to analysis by Nimblegen 153
Figure 4-9 Creating a moving average file 155
Figure 4-10 Discarding a replicate 157
Figure 5-1 Flowchart depicting bioinformatic analysis of Fru DamID datasets .. 163
Figure 5-2 Boxplots representing the DamID datasets, and their spread 164
Figure 5-3 Chromosome distribution of DamID datasets 167
Figure 5-4 Tissue specific expression of Fru target gene. 169
Figure 5-5 Protein domains enriched across the DamID datasets 171
Figure 5-6 Significant genes in each dataset show a commonality between isoforms Fru^{MB} and Fru^{MC} 175
Figure 5-7 Overview of significant GO representation in the DamID datasets.... 177
Figure 5-8 GO by isoform: Isoform Fru^{MA} 179
Figure 5-9 Venn diagrams depicting overlap of Fru^{A} 180
Figure 5-10 GO by isoform: Isoform Fru^{MB} 182
Figure 5-11 Venn diagrams depicting overlap of Fru^{B} 182
Figure 5-12 GO by isoform: Isoform Fru^{MC} 184

Figure 5-13 Venn diagrams depicting overlap of Fru^{C} 185
Figure 5-14 Venn diagrams show the overlap between individual time points... 188
Figure 5-15 Fru^{MB} adult and Fru^{mutMB} 190
Figure 5-16 Comparison of the DNA binding mutant with other isoforms in adults..... 191
Figure 5-17 Representation of downstream Fru targets in the nervous system development ontology..... 193
Figure 5-18 Identifying putative MOTIFs in DamID targets..... 195
Figure 5-19 Ttk and Kr as potential MOTIF sequences for Fru..... 196
Figure 5-20. Fruitless self-regulates at promoters P3 and P4..... 198
Figure 5-21 DamID datasets shown as a .wig file on UCSC genome browser... 202
Figure 5-23 DamID binding of *chinmo* 203
Figure 5-24 DamID binding of *pdm2*..... 203
Figure 5-25 DamID binding of *Kr*..... 204
Figure 5-26 DamID binding of *flw*..... 205
Figure 5-27 DamID Binding of *lola* 206
Figure 5-28 DamID binding to *ttk*..... 207
Figure 5-29 Formation of a MOL in candidate gene down regulated flies..... 210
Figure 6-1 Potential Cooperative Binding of Fru^{MB}..... 227

Acknowledgements

Firstly I would like to thank Stephen Goodwin, my supervisor, for the opportunity to work on this project and for his help and advice, particularly during the transition period between Glasgow and Oxford. I also owe much gratitude to Megan Goodwin, for all the training in the lab, and for help and guidance throughout the project, without which I'm not sure I'd have made it to the end!

Thank you to the past and present members of the Goodwin lab: Liz, John, Tony, Suzanne, Hania, Carolina, Caroline and occasional visitors Darren and Aaron. Thank you for the conversations when working out the kinks in my project.

Thank you also to my friends in Glasgow, Oxford and from pre-thesis time! You have put up with a variety of rubbish excuses as to why I can't meet up – I promise there will be fewer of them from now on in.

And finally, thank you to my family, Ali, Suze, Adam and mum and dad. You are the ones that were on the receiving end of the rants on subject matters you knew nothing about – thank you for listening, it was all worth it I promise! Looking forward to many more family get-togethers.

This work has been supported by a four-year DTG BBSRC studentship.

Author's declaration

I hereby declare that all the work reported in this thesis is my own unless stated otherwise in the text. None of the work has been previously submitted for any other degree at any other institution. All sources of information used in the preparation of this thesis are indicated by reference.

Glossary

Anatomical Terms

Central nervous system	CNS
Muscle of Lawrence	MOL
Mushroom bodies	MB
Nervous system	NS
Oenocyte	oe
Peripheral nervous system	pNS
Ventral nerve cord	VNC

Genetic Terms

Amino acid	aa
apontic	*apt*
Canton S	CS
Castor	*Cas*
Chromatin immunoprecipitation on a tiling array	ChIP-chip
Complementary DNA	DNA
Curly of Oster	*CyO*
Cuticular Hydrocarbon	CH
Cyclin E	*CycE*
Dam protein (*E. coli)*	Dam
DNA adenine methyltranscferase identification	DamID
doublesex	*dsx*
Double strand break	DSB
Double stranded DNA	dsDNA
ebony	*e*
Ecdosyne receptor	EcR
Ethyl methylsulphonate	EMS
flapwing	*flw*
Flippase site-specific recombinase	FLP
Flippase recognition site	FRT
fruitless	*fru*
Gene of interest	GOI
Green Fluorescent Protein	GFP
head involution defective	*hid*
Heat shock	HS
Homologous recombination	HR
Humeral	*Hu*
Hunchback	*Hb*
Krupple	*Kr*
longitudinals lacking	*lola*
Membrane-bound mouse lymphocyte marker CD8	mCD8
Messenger RNA	mRNA
mini-white	*mW+*
Mosaic analysis with a repressible cell marker	MARCM
Multiple cloning site	MCS
Pou domain (2)	*Pdm(2)*
prospero	*pros*

Protein of interest	POI
RNA interference	RNAi
roundabout homolog 3	*robo3*
rosy	*ry*
Serrate	*Ser*
seven up	*svp*
Sex-lethal	*Sxl*
Sex determination hierarchy	*SDH*
shibire	*shi*
Small interfering RNA	siRNA
Stubble	*Sb*
Temperature sensitive	TS
Tramtrack	*Ttk*
Transcription start site	TSS
transformer	*tra*
transformer-2	*tra-2*
Tubby	*Tb*
Upstream Activating Sequence	UAS
white	*w*
Wild-type	WT
Yeast Transcriptional Activator GAL4	GAL4
yellow	*y*

Measurements

Acidity $-\log_{10}$(Molar concentration of H^+ ions)	pH
Base pair	bp
Centimetre	cm
Degrees Centigrade	°C
Gram	g
Hertz	Hz
Hour(s)	hr(s)
Kilobases	kb
Kilovolt	kV
Litre	L
Melting temperature	T_m
Metre	m
Microgram	mg
Micrometre	mm
Micromolar	mM
Millilitre	mL
Millimolar	mM
Million years ago	Mya
Minute	min
Molar	M
Nanogram	ng
Nanometre	nm
Percent	%
Second	sec or s
Revolutions per minute	RPM
Units	U

Volt	V
Weight per volume	w/v

Behavioural Parameters

Courtship index	CI
Interpulse interval	IPI
Wing Extension Index	WEI

Chemicals

2' Deoxyribonucleic acid	DNA
Ethanol	EtOH
Ethidium Bromide	EtBr
Ethylene diamine tetraacetic acid	EDTA
Hydrochloric acid	HCl
Isopropyl-ß-D-thiogalactopyranoside	IPTG
Messenger RNA	mRNA
Natural Goat Serum	NGS
Paraformaldehyde	PFA
Phosphate buffered saline	PBS
Potassium Chloride	KCl
PBS with Triton-X	PBT
PBS with Triton-X and 5% NGS	PTN
Ribonucleic acid	RNA
Sodium chloride	NaCl
Sodium dodecyl sulphate	SDS
Saline sodium citrate	SSC
Salt, sodium phosphate EDTA	SSPE
Sex peptide	SP
Sodium hydroxide	NaOH
Transfer RNA	tRNA
Tris, acetic acid, EDTA	TAE
Tris EDTA	TE

Computing and Statistical Terms

Adobe illustrator file	ai
False discovery rate	FDR
Familywise error	FEW
Gene ontology	GO
General feature format file	.gff
Logarithmic scale	log
PAIR file	.PAIR
Portable document format	pdf
Significance test: probability value	p-value
Standard deviation	s.d.
Standard error of the mean	s.e.m.
Tagged image file format	tiff
University of California, Santa Cruz	UCSC
Wiggle file	.wig

Miscellaneous

Broad-complex/Tramtrack/Bric-à-brac	BTB
Calmodulin binding protein	CBP
Inverted repeats	IR
Methylation	Me
Neuroblast	NB
Overnight	O/N
P transposable element	*P*-element
Polymerase chain reaction	PCR
Ribonuclease A	RNaseA
Tandem affinity purification	TAP
Tobacco etch virus	TEV
Ultraviolet	UV
Zinc finger	ZnF

1 Introduction

1.1 The importance of studying behaviour

Behaviour is the reaction or response of an animal to a stimulus. Stimuli can be internal, such as hunger, or external, such as smell. The ability to respond appropriately to such stimuli contributes to the survival of the animal. For example, hunger cues should trigger feeding behaviours/ food intake to promote survival. More complex animals show a wider range of behaviours, which are broadly classed as innate or learned [3, 4].

Innate behaviours are a class of behaviours which are performed with no prior experience, suggesting that there is a genetic component [5]. Comparatively, learned behaviours arise from observation, experience and imitation [6, 7]. By studying innate behaviours we can gain insight into how individual genes contribute to the development of an animal such that it is capable of performing a given behaviour. In particular, the development and function of the central nervous system (CNS) plays an important role in directing innate behaviours.

To investigate how individual genes influence the development and function of the CNS, a suitable model must be identified. Such a model should be neurobiologically, genetically, and behaviourally tractable. For these reasons, *Drosophila melanogaster*, the fruit fly, has been used for decades in the study of the genetics and neurobiology of behaviour.

1.2 *Drosophila melanogaster*: A model for identifying and studying genes involved in behaviour

Drosophila melanogaster is a well-studied organism which displays a wide range of simple and complex behaviours. Over time, researchers described many

behaviours, and developed simple assays for quantifying them [8, 9]. Moreover, a variety of forward and reverse genetic tools were developed to investigate the contribution of individual genes to a behaviour [3].

1.2.1 Tools for studying genes that influence behaviour

Since Konopka and Benzer first studied clock genes in *Drosophila* in 1971 a wide variety of tools have been developed to aid understanding of behaviour: genetic markers, generation of mutations via EMS (**e**thyl **m**ethyl**s**ulphonate), random *P*-element insertions, mutant lines, various expression lines, RNAi (RNA interference), the GAL4 system, and more recently targeted homologous recombination (HR), and whole-genome screening methods [10-16].

Initially, mutants were induced via non-targeted methods: X-ray and EMS, which were screened to identify specific mutations which could be linked to specific phenotypes. Over time these techniques were refined, identifying safer mutagenesis methods, allowing further understanding of the inheritance of mutations. Sequencing of the genome in 2000 further added to the robustness of the model by providing the sequences required to design primers, identify function in genes and aid analysis of whole genome expeirments taking *Drosophila* genetics to the forefront of understanding genes, brains and behaviour [8, 17].

1.2.1.1 P-transposable elements (P-elements)

P-element insertions were perhaps one of the biggest breakthroughs in fly genetics: A *P*-element is a transposon specific to *Drosophila,* and identified by its flanking 31 basepair (bp) inverted repeats, which allow insertion of DNA into the genome via germline transformation during the embryo stage [18-20]. The system

is controlled by immobilising the transposase element on one plasmid, and a *P*-element flanked by inverted repeats on another. These two plasmids are co-injected into recently fertilised embryos [19, 20]. Only when the two elements are together can the DNA be 'jumped' into the genome [21]. The jump occurs once (as after the jump the two elements are separated), thus not disrupting the genome continually [19]. Progeny from the injection process are screened (via an incorporated marker gene) through a series of crosses, and balanced to identify the chromosomal insertion of the *P*-element. These methods along with standard mutation tools such as EMS can be used to identify and characterise new mutations.

1.2.1.2 GAL4/UAS system

GAL4/UAS is a binary expression system identified by Brand and Perrimon in 1993 which revolutionised the way in which we are able to characterise the expression of genes [22]. Exploiting the yeast transcription factor GAL4, which is not endogenously expressed in *Drosophila*, the system comprises of the promoter of the gene of interest (GOI) cloned and fused upstream of the yeast transcription factor GAL4 coding sequence [23]. This GAL4 line is crossed to a fly line containing the upstream activating sequence recognized by GAL4 fused to a reporter gene. When combined in the progeny of these flies the constructs force the reporter gene to be expressed in cells which would usually express the GOI, allowing the expression pattern to be documented [24, 25]. More recently, this system has been further exploited. A GAL4 collection has been designed with the intention of being able to drive expression in a small clusters of cells, giving this potential powerful tool the ability to sub divide the nervous system (NS) to further

our understanding of expression patterns, and the effect of genes depending on the location and timing of their expression [26].

1.2.1.3 RNA interference (RNAi)

Alternatively, RNAi is a useful tool for loss-of-function studies. Where overexpression lines increase the function of a gene by fusing the cDNA of a GOI to an **u**pstream **a**ctivation **s**equence (UAS) to increase gene expression in a subset of cells or tissue, RNAi knocks down the level of gene expression. This occurs by small interfering ribonucleic acid (siRNA) molecules blocking messenger RNA (mRNA) action, potentially revealing the role of the gene by depicting its absense. RNAi can be unreliable as different lines can have greatly differing levels of knockdown [13]. Further the potential for off-target effects is high, as genes with a similar sequence to that used in the RNAi construct could potentially be affected [27]. Where possible, multiple lines should be considered in optimising an RNAi experiment to identify a line which produces a strong effect for the gene in question, and limits off target effects [13].

1.2.2 Genes can influence the CNS to generate behaviour

The size and complexity of their brains (a small enough number of neurons that individual neurons and circuits can be identified and studied, yet complex) make *Drosophila melanogaster* an excellent tool for understanding the interactions between genes and their behaviours, and the genetic tools which were honed over the years have identified the roles that these genes take on [8, 11, 12]. Even complex behaviours can be built upon basic innate behaviours, meaning that often the underlying paradigms that are used by the fruitfly can help to elude to complexities seen in higher organisms [8].

There are a number of tools availabile that have helped to identify the way in which genes influence the CNS. Developmental genes are notoriously difficult to study as, often, their deletion leads to early-stage lethality in the fly. To overcome this *shibire^{ts} (shi^{ts}),* a temperature sensitive locus, can be used in conjunction with the GAL4/UAS system. When combined with a mutant of interest, *shi^{ts}* acts to inhibit the cell it is expressed in at temperatures below 29 °C, the action of which can subsequently be turned off by raising the flies incubation conditions to above 29 °C enabling effects to be characterised in a viable fly [28, 29].

Further specifying the effect of genes and their influence on the CNS was achieved using GFP to trace neurons, and more recently **m**osaic **a**nalysis with a **r**epressible **c**ell **m**arker (MARCM) both of which further refine our understanding: tracing neurons from their origin to identify their end using a flourescent marker (or similar) while MARCM uses the FLP/ FRT system to create individually labelled homozygous mutant cells in an otherwise wild type (WT) fly [30, 31] . Both these methods allow for understanding of how small clusters of neurons interact with each other. These tools are advancing, and the field of optogenetics appears to be the new standard for observing the CNS. By combining optical methods with genetic understanding, this field is starting to tease apart the finer aspects of CNS control, by exciting a specific neuron and observing and characterising the effects on the behaviour of the fly [32].

Courtship behaviour has been identified as an excellent paradigm for studying the genetic basis of behaviour in the CNS: complex and innately inherited [5, 33]; sexually dimorphic (with genes identified to be involved in these dismorphisms) [9, 34, 35]; and a series of individual behaviours from which individual genes, and

those genes involved in interlinking the sequence can be identified, make it ideal for studying the link between genes and behaviour [5, 33].

1.3 Courtship behaviour

Courtship behaviour in *Drosophila* is a sequence of precise behaviours displayed by the male fruit fly when attempting to attract a mate. It begins with the male fly following and orientating to the female [35]. The male approaches from behind the female and taps her abdomen with his foreleg. This behaviour is thought to allow the transfer of chemosensory cues (pheromones) to the male's gustatory sensillae on the front legs [36]. The male then extends one wing, singing a species-specific courtship song consisting of sine song and pulse song [37-39]. It is the interpulse interval (IPI) which identifies/ differentiates a species and allows the female to identify a conspecific male [40]. At this point, if the female is stimulated by the song and other early courtship behaviours she will slow down, allowing the male to lick the genital plate with his proboscis. Finally, the male attempts copulation before mounting the female, grasping her with his forelegs and, if copulation is successful, transferring a mixture of sperm and seminal fluids. If at this point, copulation is unsuccessful, the male would return to the early stages of courtship (Figure 1-1). The only behaviours displayed by the female are acceptance (displayed as a slowing down and moving of the genital plate) or rejection (kicking, running away, extruding the ovipositor) [35, 41]. The process of courtship lasts approximately 20 minutes in wild type *Drosophila* [42].

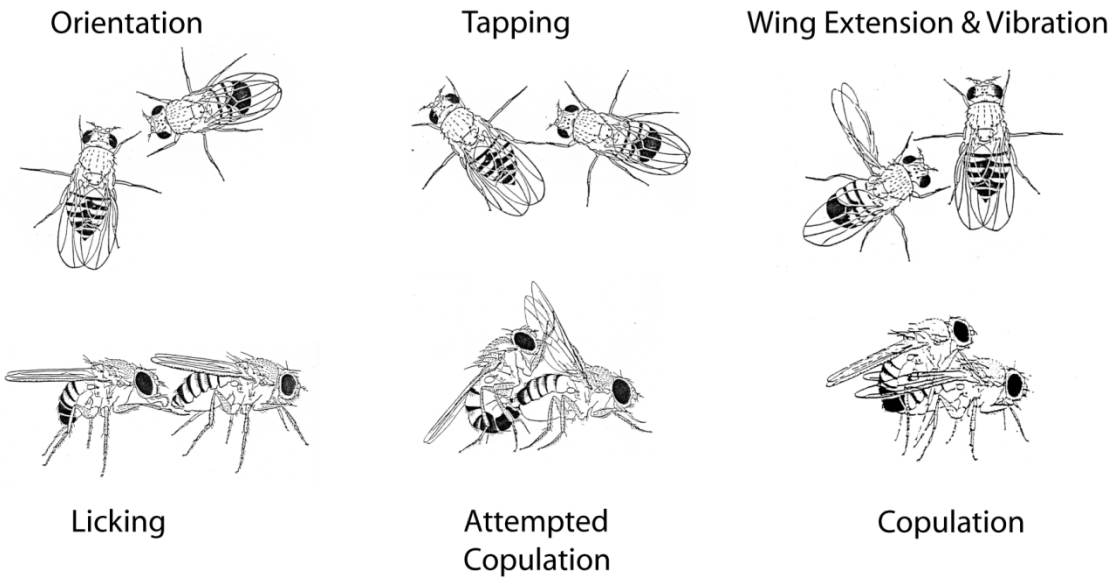


Figure 1-1 The courtship behaviour of *Drosophila melanogaster*. The male fly performs a series of behaviours to attract a female and convince her to allow him to copulate. During this process a number of visual and sensory cues are used, including the transfer of pheromones. The courtship steps include orientation, tapping, wing extension and vibration to produce a species-specific song, licking, attempted copulation and finally, copulation. Figure originally published in Burnet and Connolly, 1974 [43].

Orientation

The male orientates to the target female at approximately 20 °. Here visual cues are exchanged: the female is still running around at this point, requiring the male to be able to track her [44, 45]. The male receives olfactory cues from the female, in the form of cuticular hydrocarbons (CH) which identify her as being receptive to mating [5, 36, 46, 47].

Tapping and following

The male fly taps the female with his front leg receiving chemosensory cues through gustatory receptors in the legs [2, 36, 46, 47]. If the female is still moving he maintains an orientation of about 20 ° to her.

Species-specific song

The song produced by male flies both promotes mating and delivers information which identifies his species [48-52]. There are two components to song: sine and pulse. The sine is a low frequency humming at 160 Hz [38]. Pulse song is a rhythmic song with an intrapulse frequency of 240 Hz [53] consisting of trains of pulses separated by a species-specific gap, called the IPI. The female identifies the species of the male by interpreting the IPI - the time between consecutive pulses, and the sinusoidal cycling of the IPI [49, 50, 52-54]. At this point, if the female is interested, she will slow down allowing the male to proceed with the next steps of courtship [37].

Licking and attempted copulation

If the female is receptive, the male licks her genitalia with his proboscis receiving auditory and gustatory cues. These are stronger than those detected through olfaction [36, 46]. He then performs attempted copulation, simulating the final act of courtship behaviour.

Copulation

If the female is still receptive, she moves her vaginal plates allowing the male access for copulation. During copulation the male transfers seminal fluids along with the sperm, which include a cocktail of chemicals that alter the physical state of the female making her unreceptive to other males, and converting fat production to egg laying, which increases [55]. These proteins have been shown to be produced in the accessory glands, seminal vesicles, ejaculatory duct, ejaculatory bulb and testes and include proteases and protease inhibitors, lectins, prohormones, peptides and antioxidants [56]. This was initially shown through dissecting out accessory glands from the male and implanting them in females,

which lead to such responses being simulated [57, 58]. Since then, the components of seminal fluid have been investigated, in an attempt to understand the method by which the changes observed in the female are instigated. The most understood of which is sex peptide (SP), a peptide involved in egg production, receptivity, feeding and sleep behaviour [55, 59, 60]. It has been shown to affect expression levels of 52 genes in the head and abdomen of a mated female [61]. Up to 146 other proteins in the seminal fluid have been identified through a variety of methods, however few have been directly associated with a specific role [56]. Those which have include the sperm storage proteins AcP62F and AcP29AB which have been shown to affect female transcriptome [56]. Such post-mating responses last around one week [62].

The various cues of courtship behaviour play a strong role in successful copulation. Sex of the fly is detemined through smell, taste, look and behaviour [5]. Without the pheromones (chemosensory cues), auditory and gustatory cues, courtship behaviour would be considerably less efficient, as seen in mutants deficient for these sensory abilities: *smellblind* and *olfactory-D* to name a couple [35]. Acuity of these senses has therefore been selected for as an advantage, over time.

1.3.1 Visual cues

The visual stimuli of courtship behaviour are both static (physical characteristics) and dynamic (positioning, wing display) which aid in initiation of courtship, allowing both sexes to correctly identify a mate [44]. The female is physically larger, with a lighter thorax whereas the male is smaller with a dark thorax [44, 45]. The first stage of courtship behaviour involves the male identifying a suitable female through her shape and movements [44, 45, 63]. Further to this, in the early

stages of courtship, the female will regularly run away requiring the male to track her to keep proximity [64, 65]. This has been shown with flies deficient for the *white (w^-)* gene which are compromised in courtship due to decreased visual acuity which increases the amount of courtship behaviour required, and reduces their courtship success (due to the behaviours performed being less stimulatory) [44, 45, 66-68].

1.3.2 Acoustic cues

Courtship song is achieved by the male extending and vibrating his wing in a species-specific manner [5, 54]. The song can be stimulatory to both sexes and identifies the mate as being of the correct species [50, 51]. This important step has been characterised by observing differences between wingless males courting, and wingless males courting in the presence of artificial song being played. In these experiments, the wingless males courting in the presence of artifical song are more successful [48]. This aspect of behaviour requires both Fru and Dsx for complete execution [39].

1.3.3 Chemosensory cues

Most chemicals exchanged during courtship behaviour are detected by taste (gustatory receptors) and smell (olfactory receptors) [5, 36, 46, 47]. The main gustatory receptors for courtship behaviour are found in the proximal legs and are stimulated by tapping, further receptors are found in the proboscis. Olfactory receptors are located on the antennae and maxillary palps and are engaged during running, orientating, and courtship song [36, 46, 69].

The equivalent of sex pheromones in *Drosophila melanogaster* are CH's which are synthesised in a sexually dimorphic manner [36]. Female CH's have a stimulatory

effect on the male, which alone can induce aspects of courtship behaviour, and when absent decrease their attractiveness to the male [40, 70]. Males also produce CH which act as an anti-aphrodisiac. When absent (or undetectable) a higher incidence of intermale courtship is observed showing an important role in suppressing inappropriate advances [71, 72]. Overall, it is thought that these chemicals act to identify the sex of the fly during the initial steps of courtship [73].

Underlying all of these behaviours are a cascade of genes pivitol to the display of these behaviours in the correct order. These genes instigate the sexual dimorphism both physically and behaviourally and form the sex determination hierarchy (SDH).

1.4 The sex determination Hierarchy

In *Drosophila melanogaster*, sex is cell autonomous, in that each cell determines its sex independently of those that surround it (except in the germline/ gonads) [35, 74]. Specifically, this means that sex is controlled by factors acting within each individual cell or tissue, rather than adjacent cells. Sex in flies is determined by the ratio of sex chromosomes to autosomes, Figure 1-2.

In females, the ratio is 2: 2 sex chromosomes to autosomes which leads to the active form of the gene *Sex lethal (Sxl)*, a master regulator protein. *Sxl* is an RNA-binding splice regulator, which through alternative splice sites affects splicing both in *transformer* (*tra)* downstream of it, and itself. In the early embryo destined to be female, *Sxl* is transcribed from the early promoter E (P_E) which produces a functional protein Sxl. Sxl protein acts to transcribe *transformer (tra)*, while concurrently binding exon three of the *Sxl* protein preventing late embryo translation of promoter L (P_L) in the *Sxl* gene allowing continued translation of *Sxl*.

This process is auto-regulatory. The main downstream target of Sxl regulation in the sex-determining pathway is *tra,* which encodes an RNA binding protein. Upon splicing by Sxl*,* Tra binds its co-factor Tra-2 to regulate both *doublesex (dsx)* and *fruitless (fru)* through a common 13 nucleotide sequence (at the 3′ end of *dsx*, and the 5′ end of *fru*). In females, Tra/ Tra-2 introduce an alternative splice in *dsx* pre-mRNA to generate a female-specific transcript which produces Dsx^F [75]. Dsx^F is a transcription factor largely responsible for the somatic differences in females (i.e. their anatomy and physical attributes). Further, the Tra/ Tra-2 complex binds *fru* P1 pre-mRNA causing it to use an alternative splice which introduces a premature stop codon into the transcript. As a result, no Fru protein is produced in females (Fru^F) [76, 77].

In males, *Sxl* is not transcribed as the ratio of sex chromsomes to autosomes is 0.5. No *Sxl* is transcribed from P_E which subsequently allows P_L to be transcribed in the late embryo, creating a protein via an alternative splice site in the third exon. As a result of not having functional *Sxl* protein, Tra is not produced and the Tra/ Tra-2 complex does not form allowing *dsx* to use a default splice-site, producing male specific isoforms. Doublesex's male isoform (Dsx^M) accounts mainly for the male somatic structures as its counterpart does in females such as in pigmentation, the sex combs and internal reproductive organs. However, *dsx* does have a role in the NS and behaviour in both sexes.

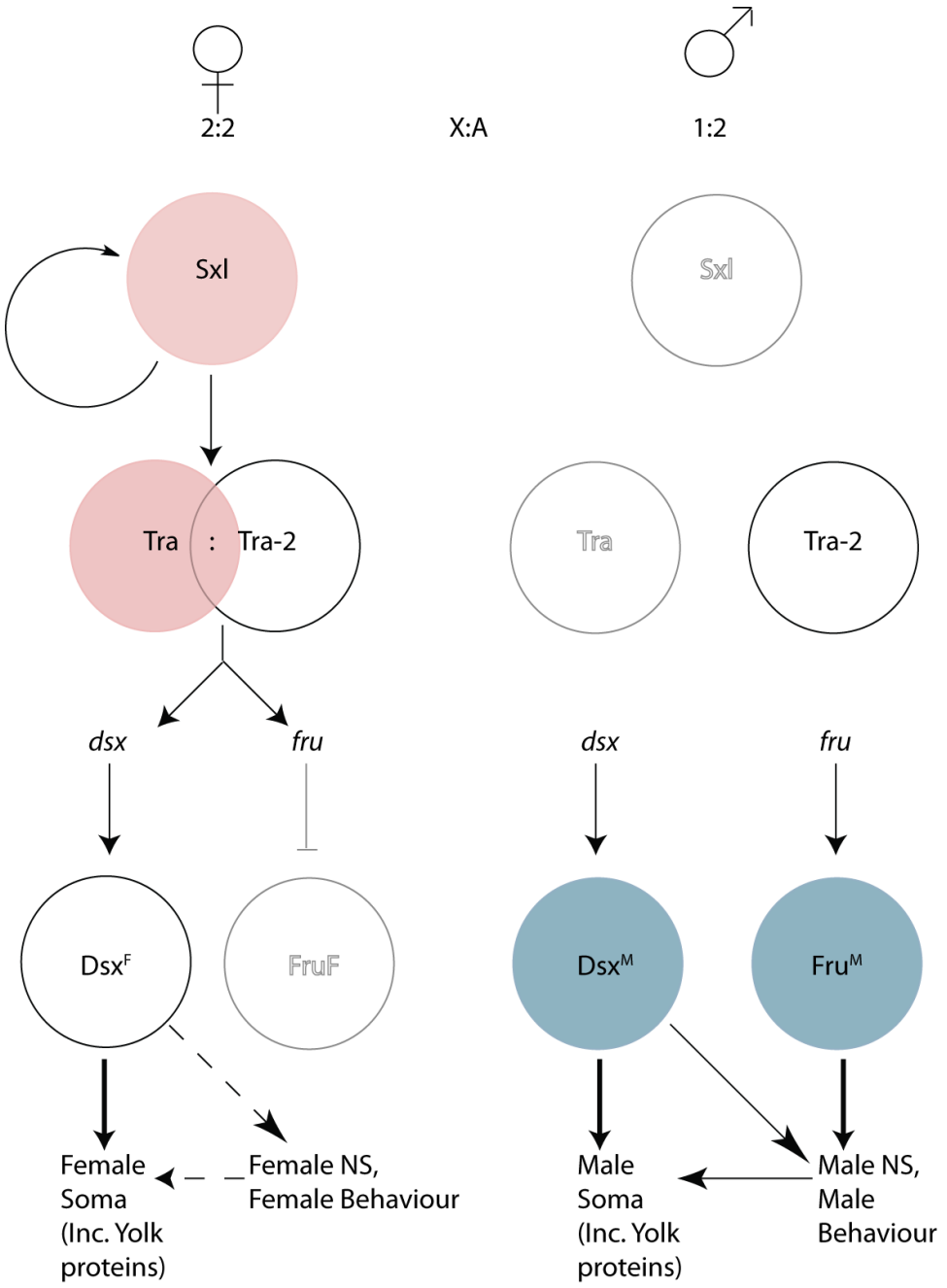


Figure 1-2 The sex determination hierarchy in *Drosophila melanogaster*. In females the ratio of sex chromosomes to autosomes (X: A) is 2: 2 resulting in active Sxl which in turn activates the Tra: Tra-2 complex causing a sex-specific splice to female versions of Dsx and Fru being produced. The female Fru contains a premature stop codon preventing it from being translated into a functioning protein. In the situation where a 1: 2 ratio arises *Sxl* is not activated which means the Tra: Tra2 complex does not form (due to a lack of Tra transcription). This causes Dsx and Fru to be transcribed in their default, male state. Image modified from Billeter et al., 2006 [2].

In males, the inactivity of Tra allows *fru* transcripts from P1 to follow a default splice, producing a functional protein with a 101 amino acid N-terminal male-specific addition. The primary gene regulating courtship behaviour in flies is *fruitless*.

1.5 Fruitless is a complex gene that controls male sexual behaviour

fruitless is a complex gene spanning 130 kb, which encodes a BTB (**b**ric a brac, **t**ramtrak, **b**road) C_2H_2 zinc finger (ZnF) transcription factor (Figure 1-3) with two distinct functions: control of male sexual behaviour and viability in both sexes [78]. *fru* has four promoters (P1-4): P1 transcripts are sex-specifically spliced via the SDH, whereas those derived from P2-4 are common to both genders (common promoters) [2, 34, 78-80].

The BTB domain is common to all transcripts [78], augmented by one of four alternatively spliced 3' ends (A-D). However, Northern blot analysis has shown that sex-specific P1 transcripts contain only the A, B and C 3' ends [2].

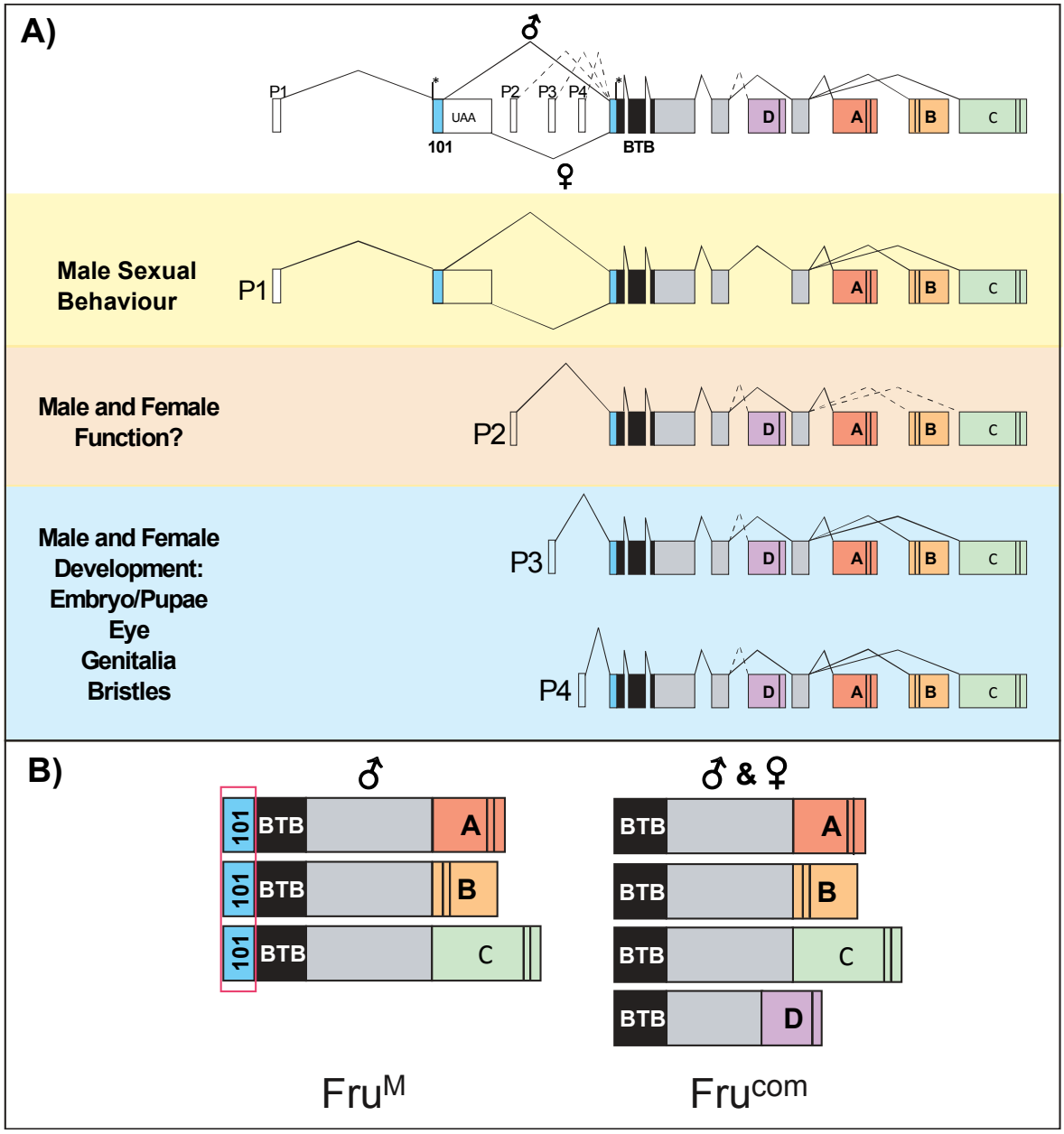


Figure 1-3 The gene structure and transcripts of *fruitless*. The complex *fru* gene has four separate promoters allowing it to produce a variety of transcripts. (A) The functions associated with individual promoters are described on the left, while the specific splicing associated with these promoters are shown on the right. (B) Composition of proteins expressed from the *fru* locus. All Fru proteins share a BTB-domain and one of four DNA-binding domains. Male-specific isoforms (Fru^{M}) differ from common isoforms (Fru^{com}) at the N-terminus due to an additional 101 amino acids (red box). M. Neville, unpublished.

1.5.1 The sex-specific transcripts of P1

There are three sex-specific isoforms of *fru* transcribed from P1: Fru^{MA}, Fru^{MB}, Fru^{MC} [2, 78, 79]. These transcripts differ from the non sex-specific transcripts by a 101 amino acid sequence at the 5′ end [78]. Splicing at the 3′ end of the gene allocates one of the three ZnF exons (A, B, C) [79]. This creates six potential

transcripts (three female, three male) [79]. These transcripts are transcribed in 20 distinct cluster within the CNS comprising about 2000 neurons (2 % of total) [81], collectively controlling courtship behaviour in *Drosophila* alongside *dsx*, however the individual functions of each isoform remain unclear [82, 83].

The P1 transcripts are initially transcribed at the L3 larvae timepoint, prior to which no sex-specific activity can be detected [79]. By the mid-pupae stage expression peaks, before dipping slightly through to the adult where it maintains a regular expression level through seven days [81]. To date, mutations have been isolated in the fru^{C} isoform, which has allowed us to elucidate some of its functions (its role in viability, fertility and formation of the Muscle of Lawrence (MOL)) [2]. Other than Fru^{MC}, very little is known about the other sex-specific isoforms. Fru^{MA} is only expressed in a subset of Fru expressing cells [1, 84]. This suggests that Fru^{MA} could have a role in restricted subsets of Fru-expressing cells/neurons.

The mutant allele of isoform C is referred to as $fru^{\Delta C}$. It was first isolated in an EMS screen in 2006 [2]. It contains a 7 bp deletion of nucleotides 77- 83 in the Fru^{C} exon of *fru* which is predicted to cause a frameshift/ stop codon, disrupting the protein, rendering it non-functional. The mutant is semi-lethal (52 % in females, and 44 % in males), with lethality occurring during late pupal development. $fru^{\Delta C}$ mutants have decreased male fertility (31 %), have several defects in courtship behaviour, and lack the MOL. The MOL is an abdominal muscle specific to male flies which is thought to be involved in aspects of courtship behaviour, primarily the un-curling of the abdomen during copulation [2, 85]. It is the innervating neuron of the MOL which is required to be male in order for the MOL to form. When the $fru^{\Delta C}$ isofrom is not present, these innervations no

longer direct the formation of the MOL, and instead, four smaller muscles are seen in place of the MOL, similar to what is observed in the female abdomen. In a small subset of *fru*$^{\Delta C}$ mutant males (~ 10 %), however, a vestigial MOL is formed, suggesting that partial MOL formation can occur in the absence of Fru^{MC} protein [38].

1.5.2 The common transcripts: P2 -4

In addition to its role in regulating sexual behaviour, *fru* also affects viability. This vital function is mediated by transcripts derived from the P2-P4 promoters, which are non-sex-specifically expressed [34, 78, 79]. The *fru*$^{\Delta C}$ mutation affects both the sex-specific and common transcripts of Fru, and so viability effects resulting from this line represent the common Fru^{C} isoform function [2].

Both Lee *et al*., and Ryner *et al*., have attempted to elucidate the role of Fru^{com} through antibody staining and *in situ* hybridisation respectively [78, 81]. Antibody staining showed that there is strong expression in the larvae in both sexes, although the isoform responsible for this staining is currently unknown. In the pupal and adult stages, staining is seen only in the male CNS and mimics the staining of Fru^{M} suggesting no common transcripts are expressed here. The exception being the four-day pupae where some broad low level expression was seen in both sexes, however the biological significance of this has not been confirmed [81]. Outside the CNS Fru^{com} expression was documented in the imaginal discs, muscles and gonads during L3 larvae; in the appendages (which are imaginal disc derived), epidermal tissue, and a subset of the developing flight muscles in the pupae; and similar muscles, follicle cells and ovaries in the adult. In all instances this staining was seen at similar intensities in both sexes in all stages [34, 78, 81, 82, 86].

fru's role in behaviour was uncovered as a result of identifying and characterising flies with aberrant courtship behaviour [2, 79]. Although its sex-specific transcripts are starting to be understood, very little is known about the common transcripts. These transcripts are derived from P2-4 and are transcribed in both males and females. They are denoted by the superscript 'com' followed by the isoform letter, e.g. Fru^{comA}. Dornan *et al*. created a *P*-element mutation derived from the fru^4 mutant, which produced transcripts that recapitulated the antibody staining of the com isoforms. [86]. Northern analysis showed P1 transcripts were expressed in the head and body of adult males, and male pupal head, confirming these transcripts as male specific. In comparison, P2 transcripts were found in pupal and adult heads of both sexes; P3 transcripts were seen at low levels prior to L3 larvae (onset at around 10 hr.), the female pupae and adult and in the internal genitalia of both sexes; P4 was observed through all stages of development in both sexes, except in the internal genitalia of adult males [86]. Song *et al*. added to this by examining the embryonic CNS where they identified transcripts derived from the P3 and P4 promoters, and suggested that P3 transcripts were sufficient for forming WT *FasII* and *BP102* axon tracts [80]. Further, P3- and P4-derived transcripts are both thought to be involved in development, during the embryo and pupal stages, of the eye, genitalia and bristles [81].

1.5.3 The Fru protein

All Fru transcripts comprise two domains: an N-terminal BTB domain and the C-terminal ZnF domain, which binds DNA. These two domains comprise the methods by which Fru can bind other proteins and target genes for transcriptional regulation, leading to contol of behaviour.

1.5.3.1 The BTB domain

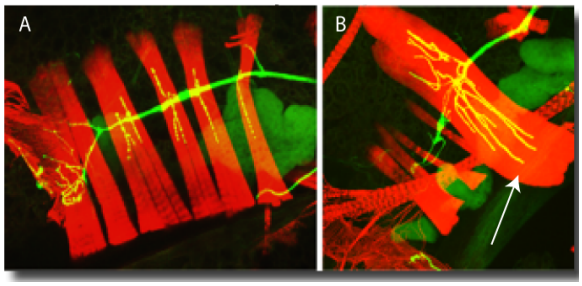


Figure 1-4 Muscle of Lawrence. The muscle of Lawrence (phalloidin - Red) and the Fru neuron (UAS-mcD8::GFP) that innervates it. (A) Female showing the WT condition, with no MOL present. (B) Male showing the WT condition, with a WT MOL present, identified by a white arrow [1, 2].

The BTB domain is an evolutionary conserved, complex protein-protein interaction domain which has a large interaction surface due to its unique folding pattern allowing it to expose residues for both dimerisation and interactions with other proteins [87, 88]. The domain is widely conserved through evolution, and is present in many other model organisms such as *Caenorrhabitis elegans, Arabidopsis thaliana*, and in humans in a variety of roles [27, 37].

The BTB domain has been implicated in diverse cellular processes such as chromatin remodelling, insulating activity, carcinogenesis, repression and activation of transcription [27, 37, 88-90]. For example, *lola*, a BTB ZnF with 20 predicted splice isoforms has been shown to regulate olfactory projection neurons during development [91]. Further, *tramtrack* (after which the domain is named; *Ttk*) has been implicated in, amongst other things, specification of daughter cell fates during asymmetric cell divisions of the **p**eripheral **n**ervous **s**ystem (pNS), showing the strong links such domains have with the development and formation of the CNS [34, 92-94].

Although it is conserved evolutionarily, Fru is notably different from other BTB ZnF proteins due to its differential expression in the sexes [37]. Some BTB proteins have been shown to facilitate homomeric interactions, such as Bcl-6. Furthermore, when in the presence of N-CoR or SMRT, Bcl-6 has been shown to interact in a heteromeric fashion. This identifies the potential for Fru to act in a

similar fashion, which could provide evidence for a manner by which Fru regulates a number of downstream targets [95-98]. Although it is unknown whether or not the latter two occur, homodimers of Fru have been shown to occur in the yeast-2-hybrid system (M. Neville, unpublished).

1.5.3.2 The C_2H_2 Zinc Finger

Fru ZnF domains are encoded within one of four alternatively spliced 3' ends, although the sex-specific promotor has been shown to only produce three transcripts containing one of three alternative C_2H_2 type DNA binding domains (A-C) [78]. C_2H_2 ZnFs create a stable structure capable of binding DNA sequence: four fingers of DNA are held together by zinc molecules which control the secondary protein structure allowing successful temporary binding to DNA of a specific sequence through this DNA binding domain [99]. The sequence of the C_2H_2 ZnF is crucial for this to occur. Cysteine residues placed periodically through the sequence allow the successful binding of the zinc molecules and therefore retention of the secondary structure [100].

Over 20 different types of ZnF have been identified, with the C_2H_2 structure being most abundant [101-103]. These structures have been found to bind in a variety of patterns, depending on the number of structures present in the protein, with some evidence for the ‘one finger, many partners’ binding pattern [99, 104-106], which should be considered as a potential method by which Fru binds its downstream targets.

Fru isoforms have been shown to have different functions [2], working individually or in combination. For example, Fru^{MC} has been shown to control the formation of the somatic MOL (Figure 1-4) [76], whereas serotonin expression in the abdominal

ganglion, whose neurons innervate the accessory glands, vas deferens and over half the neuronal terminals of the anterior ejaculatory duct, has been shown to require all the isoforms, and further, in some instances require the presence of *dsx* [2, 85]. This specificity is likely to be a result of the structure of the ZnFs (and therefore the sequence(s) to which they bind) [99].

The alternative splice sites present within *fru* provide potential to create many different transcripts by combining different ZnF containing 3´ exons. This ability to bind different DNA sequences combined with the potential for the BTB domain to add complexity through dimerization provide scope for diverse interactions, therefore Fru potentially regulates numerous genes to fullfill its role in the control of courtship behaviour.

1.5.4 Mutating fru

Past studies have generated several deletion mutants, placing *fru* within the SDH, and providing an understanding of how behaviour is controlled. The most severe of these mutations is fru^F, a point mutation at the male splice site derived from an homologous recombination (HR) screen, which shows almost a complete lack of courtship due to the forced use of the female splice site (Table 1-1) [82]. Less severe mutants are just as informative, for example fru^3, which shows certain steps of courtship, but not all, suggesting the involvement of *fru* in subsets of behaviours, as well as courtship behaviour as a whole [107]. The fru^1 mutation, an inversion breakpoint found 3.3 kb upstream of the transcription start site (TSS), causes ectopic Fru expression, causing male flies to court other males and females indiscriminately [82]. fru^1 mutants also display chaining behaviour, where males follow, orientate and sing to one another, making distinct chains of male flies courting each other. Significantly, all mutants show a level of sterility which is

behavioural rather than a loss of function within the sperm or lack of constituents in the seminal fluid [108]. Mutants for the common isoforms affect viability, such as *Df(3R)fru*^{sat15} (*fru*^{sat15}). Furthermore, mutants generated that affect the common transcripts fail to eclose, have malformed leg joints or uninflated wings [2, 82, 107]. More recently, characterisation of newly created mutants involves crossing them to well defined deficiencies: *Df(3R)fru*$^{4\text{-}40}$ (*fru*$^{4\text{-}40}$), *Df(3R)fru*^{w24}*/TM3, Sb, e* (*fru*^{w24}), and *Df(3R)fru*^{sat15}*/TM3, Sb, e* to identify their effects. *fru*$^{4\text{-}40}$ removes promoters one and two, *fru*^{sat15} deletes all common coding exons, and *fru*^{w24} removes the entire locus.

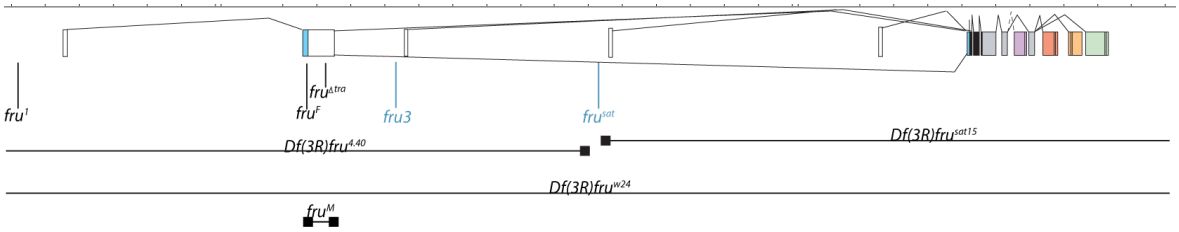


Figure 1-5 Mutations of the *fru* gene. The *fru* gene is depicted showing the exons (boxes) and alternative splicing options (solid lines attached to boxes). The common mutations of the *fru* gene are depicted as solid horizontal lines in the case of large deletions, with a square defining the end point of the deletion (no end point depicts the deletion ending outside the frame of the image). P-elements are depicted as blue arrows and blue text. *fru*^{1} is an inversion breakpoint; *fru*^{F} is a point mutation and *fru*$^{\Delta tra}$ a 261 bp small deletion in the S exon. Modified from Lee and Hall, 2004 [85].

Mutant Name	Type of Lesion	Molecular phenotype	Behavioural phenotype	Reference
fru^{satori}	*P*-element insertion slightly upstream of P3	Low-level expression of Fru^{M}. Disrupts P1 transcripts.	Shows very little evidence of courtship.	Ito *et al*. [108]
fru^{1}	Inversion breakpoint found 3.3 kb upstream of the TSS	Absence of Fru^{M} expressing neurons, ectopic expression in novel cells of P1 transcripts.	Strong chaining, court indiscriminately,	Goodwin *et al.,* Villella *et al*. [79, 107]
fru^{3}	*P*-element insertion at position 91B marginally upstream of P3.	Near absence of expression of *fru*^{M} P1 transcripts.	Poor receptiveness to females and no pulse song.	Goodwin *et al*, Villella *et al*. [79, 107]
fru^{F}	HR induced point mutations at the male splice site	No *fru*^{M}, *fru* P1 null.	No mating behaviour.	Demir and Dickson [82]
fru^{M}	HR induced female specific splice site deleted forcing male splicing.	Fru^{M} expressed in females.	In females it causes them to attempt to mate with other females. Show male behaviour.	Demir and Dickson [82]
fru$^{\Delta TRA}$	HR induced deletion eliminating the Tra binding sites and forces male specific splicing.	Fru^{M} expressed in females.	In females it causes them to attempt to mate with other females. Show male behaviour.	Demir and Dickson [82]
Df(3R) fru$^{4-40}$	Deletion of at least 70 kb covering	Expression of sex-specific (P1) and P2 transcripts greatly diminished.	Little or no evidence of courtship behaviour.	Anand *et al.* [34]
Df(3R) fru^{sat15}	Extends centromere-proximal from within *P*-element insertion of *satori* mutant.	No common coding region expression. Also deletes some local genes downstream of *fru*.	Homozygous lethal, no courtship behaviour associated with Fru is exhibited.	Ito *et al*.[108]
Df(3R) fru^{w24}	Third chromosome deficiency at 91A and 91D2-3.	Removes all of *fru* gene.	Little or no behaviour associated with Fru is exhibited. Homozygous lethal.	Ryner *et al*. [78]

Table 1-1 Common mutations of the *fru* locus. Existing *fruitless* mutants and their behaviours, showing the range of phenotypes possible due to the type of mutation. See Also, Figure 1-5.

Originally, it was thought that *fru* was sufficient for the control of male sexual behaviour entirely, and that the somatic differences outside the CNS were controlled by *dsx* [47, 109]. This has since been shown to not be the case, in fact, both genes have been shown to be necessary for complete sexual behaviour [5, 39], with *dsx* being shown to be required not only for somatic, physical attributes as was initially described, but also for the correct performance of song (specifically sine song) [39]. Further, Rideout *et al*. showed that *dsx* was required for development of male specific neurons that co-express *fru*, and proposed that *dsx* and *fru* act together to form the neuronal basis of male sexual behaviour [110].

Courtship behaviour is controlled by transcripts from the P1 promoter which are known as the Fru^{M} transcripts and are only successfully translated in the male fly. The multiple 3′ ends of *fru* allow a single gene to encode multiple proteins which have the potential to control different functions. This may be one way in which a single gene can control many different processes [5]. Six transcripts have been identified as being transcribed from the distal P1, sex-specific promoter [79], three in the male, and three in the female, however only the three male transcripts are expressed as fully functional proteins.

By observing the behaviours of mutants, it has been possible to identify the behaviours in which *fru* plays a part. Most *fru* mutants are large mutations, which lead to the line being sterile. However, these have been used to attribute most courtship behaviour to *fru* such as in the case of fru^{1} where an inversion breakpoint 3.3 kb upstream of the TSS just distal of the *fru* P1 promoter leads to

indiscriminate courting and subsequently male- male chaining [79]. *fru*^{F} and *fru*^{M} were crucial in confirming the specifics of *fru*: that it is necessary and sufficicent in creating a fly capable of identifying a fly of the opposite sex and behaving correctly toward it to be able to successfully copulate [82]. By exploiting *fru*^{F} (a fly line incapable of producing male specific transcripts) and *fru*^{M} (a fly line incapable of producing female specific transcripts) Fru was shown to control male courtship behaviours and MOL formation, by forcing male flies to translate female *fru* transcripts causing them to fail to court more than 95 % of the time. Further, when females were forced to produce male Fru^{M}, they behaved like males (while looking female), completing aspects of courtship and chaining when placed with males genetically altered to smell like females (using *oe*-GAL4/ UAS-*tra*, which modifies production of pheromones from the oenocyte to those produced by the female) [111].

These mutations have led to *fru* being linked to most of the previously described courtship behaviours; with the exception that the species-specific song has been shown to be a combination of *fru* and *dsx's* control. The method by which *fru* controls all these functions has long been pondered, and identification of strong conservation of the BTB and ZnF domains within the insect order lead to the investigation of these domains in the role of *fru's* function [37].

1.5.5 Conservation of Fru

fru is only found within the insect order, however within this order elements of its structure are highly conserved, showing it to be essential within this class. Specifically, the BTB domain and C_2H_2 ZnF domains are strongly conserved around the motifs that sequence each domain [37]. No conservation is seen in the inter-connecting regions [37].

These specific regions encompassing the binding domain motifs have been strictly conserved between *Drosophila melanogaster* and *Anopheles gambiae* over 250 million years (Mya) of evolution [37, 112]. *Anopheles gambiae* exhibit similar alternative splicing to generate a number of BTB-ZnF transcripts. It also invokes sex-specific splicing which leads to only male transcripts being fully transcribed, and whereby female transcripts are inhibited by the incorporation of a STOP codon [37]. Further to this, evidence suggests that *fru* may represent an ancestral gene of sexual determination in Dipteran insects [37]. In fact, Gailey et al., propose that the *fru*-dependent muscle, the MOL, is a vestigial muscle exluded several times in fruitfly evolution *(Drosophila erecta, Drosophila yakuba* amongst others) [37, 113].

Further investigations into the overlap between these two species found that isoform Fru^C sequence from each species can partially rescue function in the other species, and in females (who would not normally produce a MOL) was sufficient to produce a MOL in the majority of cases [2, 37].

1.6 Gene targeting by homologous recombination

Gene targeting by HR is a popular method for introducing specific mutations in *Drosophila*, first described by Rong and Golic [84]. Based on the well established technique of *P*-element transformation, it has been sucessfully used to modify several loci [20, 82, 83, 114, 115]. First pioneered in yeast, theoretically any endogenous sequence can be modified, however the structure of the genome means that some areas are more accessible than others [84, 116]. In 2000, a technique was described for use in *Drosophila* [84].

Rong and Golic refined the technique for *Drosophila* by generating linear donor DNA, avoiding the need to find a compatible cell line [2]. Although other forms of mutagenesis were available (*P*-element transformation or EMS), HR proved to be advantageous in specifically manipulating long genes and those with complex *cis*-acting regulation as the specificity of the targeting process left other elements intact [114].

For *Drosophila*, there are two methods available for successful recombination. In both instances a 'donor' plasmid carries the sequence of donor DNA – which incorporates homologous sequence for the target region, the mutation to be introduced, a marker gene for following the mutation through genetic crosses, and construct sequences used in the targeting process. This donor *P*-element is incorporated into the genome randomly [20], and subsequent steps/ crosses are followed to complete the process of gene targeting. From this point, there are two strategies that can be followed, depending on the desired outcome. The two methods of HR are "ends-out" and "ends-in".

1.6.1 Ends-out technique

This technique is most commonly used when a null mutation is required, to completely disrupt a genomic locus. Successfully used in yeast and mice, initially this process was unsuccessful in *Drosophila* [117]. By modifying the crossing scheme to occur in females rather than males, ends-out was shown to be capable of repairing or creating mutations at a target locus [2].

In ends-out targeting the marker gene is located between two regions of homology, which is inserted via a double strand break (DSB) at the target locus (Figure 1-6). Successful ends-out in females is thought to occur due to the heat

shock (HS) element of the process. Heat shock activity acting on the promoters for *I-SceI* and FLP are limited to certain periods of spermatogenesis in males, which is not a limitation in females [84], further, males do not undergo meiotic recombination in *Drosophila*.

After introduction of the target DNA and its linearization, there are two methods of integration: A target gene can be used as the homologous region to replace or modify the gene, this method would allow deletions within the gene to be incorporated. Alternatively, if flanking regions are used the gene can be completely deleted, using sequence either side of the gene to target the HR insertion, thus looping out the gene which is subsequently excised. This process leaves the possibility of interference from the marker gene. To ensure against such an event, flanking loxP sites can be used to remove the marker gene by CreI site-specific recombinase, post-targeting. Ends-in, a similar method, can be used to achieve the same end point by slightly differing means - which in some cases can be advantageous.

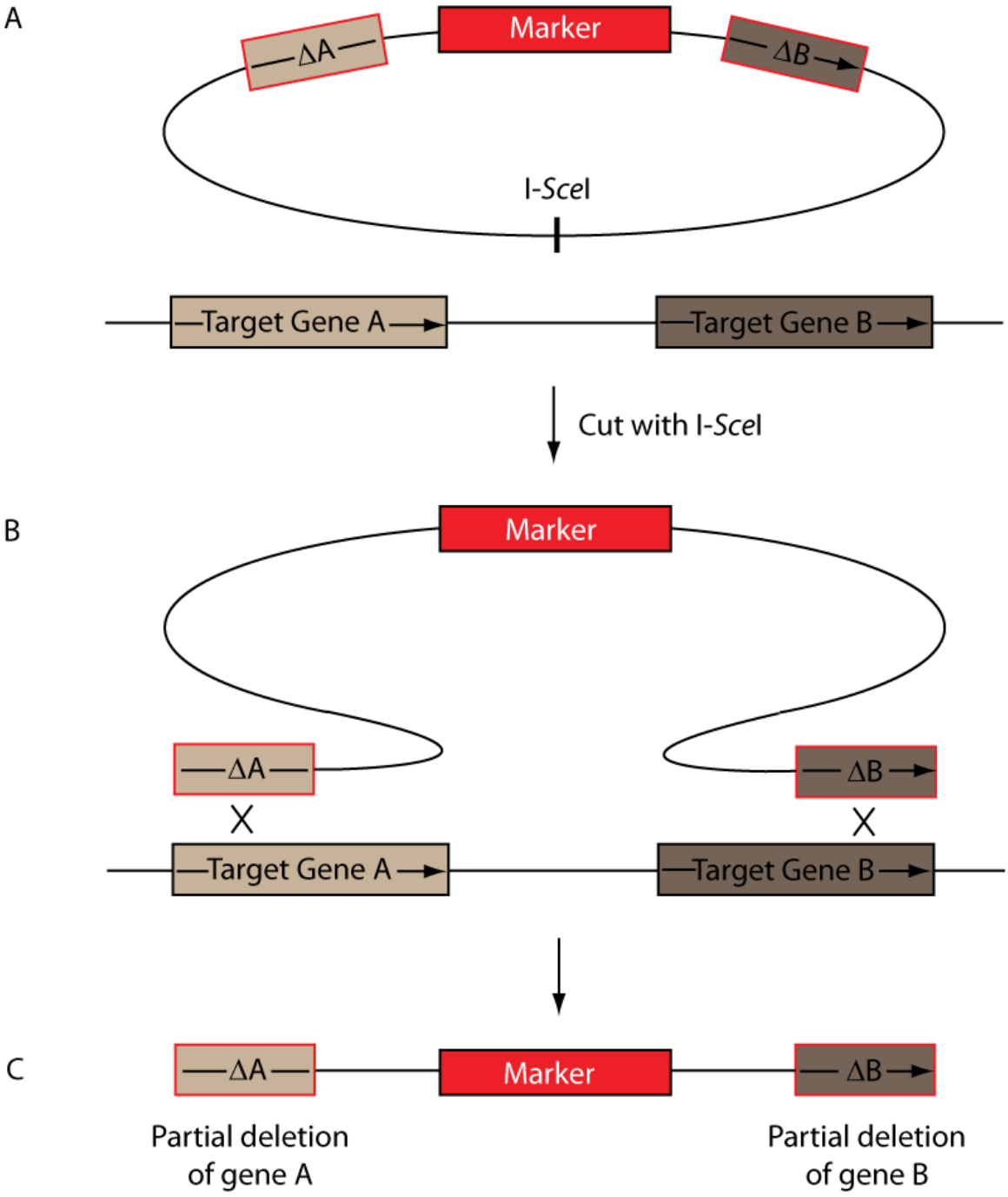


Figure 1-6 Schematic representation of 'ends-out' targeting by homologous recombination. (A) Ends-out targeting of a small locus containing two small genes, the circular targeting vector, and target locus, corresponding to the donor DNA stretch carried by the targeting vector. The red box signifies the marker gene, placed between two regions of homology to target locus (ΔA and ΔB). (B) The targeting vector is cut, and linearised at the *I-SceI* recognition site using *I-SceI* endonuclease. This produces a DSB allowing the targeting construct to replace the target locus by homologous recombination (in a number of occurrences). (C) As the regions of homology are positioned at the extreme ends of the two genes, the genes are replaced by truncated versions and the intervening sequence is deleted. J. Walker, unpublished.

1.6.2 Ends-in technique

Originally used by Rong and Golic in 2000 to resolve a mutation in the yellow gene, this technique uses a straight-forward replacement of the target DNA,

Figure 1-7. However, it requires a second step for removing the resultant tandem duplication which results from the integration of the donor DNA within the homology region. To resolve this unwanted duplication, an I-*CreI* endonuclease isolated from *C. reinhardtii* is incorporated [118]. *I-CreI* recognises a unique cut site engineered into the donor construct. This cut site induces a DSB between the duplication, allowing the removal of one copy – as the endogenous repair machinery will often remove the duplication whilst repairing the break in the DNA. Often, this also removes the marker gene making the occurrence visible. A number of fly lines retain the mutated form of the duplication.

Ends-in is more subtle than Ends-out in that small specific deletions can be introduced without altering the genomic locus including sequences involved in *cis*-regulation and alternative splicing. This means that *fru*, a complex gene, can be modified with limited concern for damaging the overall function of the resultant locus.

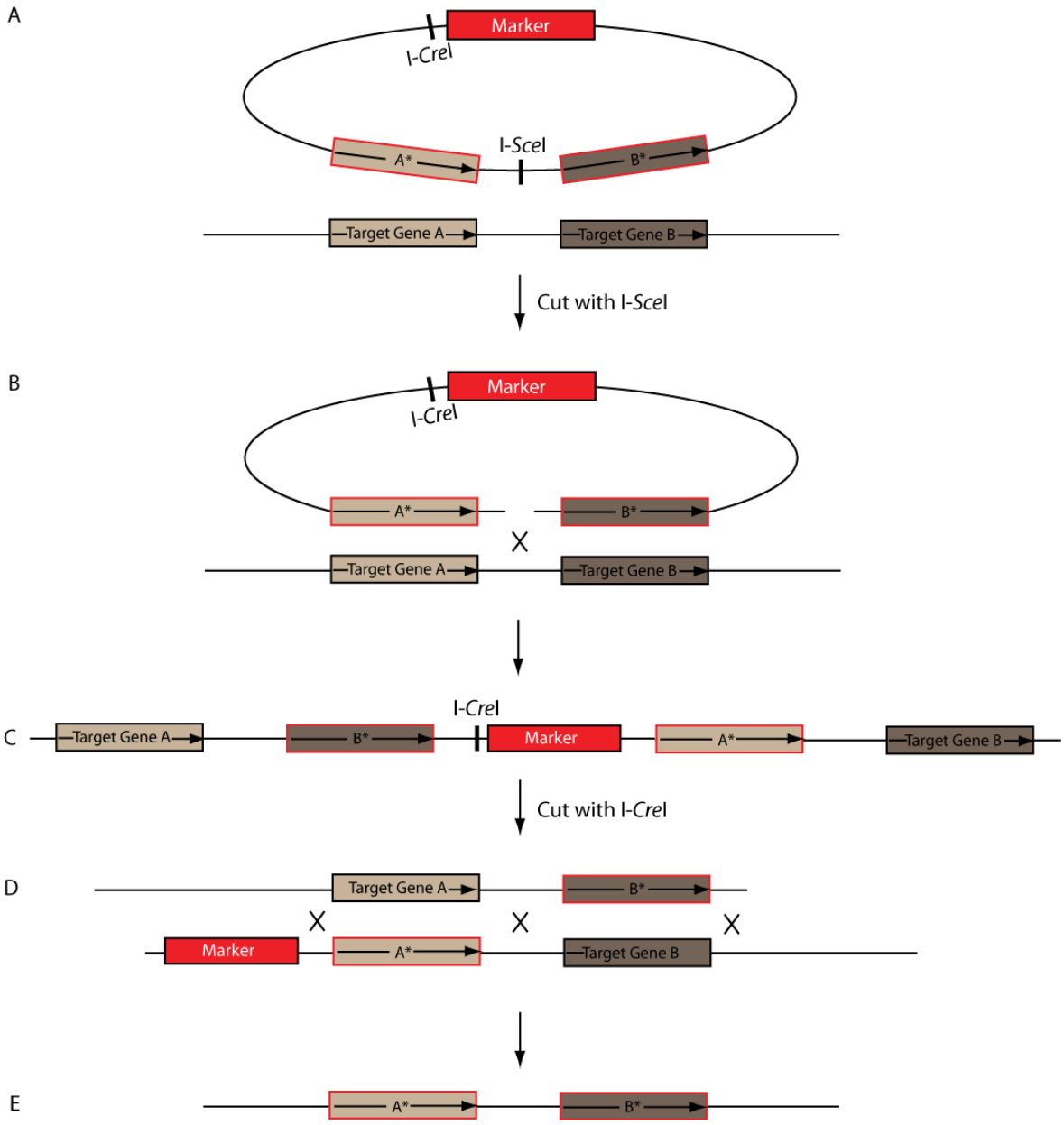


Figure 1-7 Schematic representation of 'ends-in' targeting by homologous recombination. (A) Ends-in targeting of a locus containing two small genes containing a circular targeting vector and target locus, corresponding to a donor DNA stretch carried by a targeting vector. An *I-SceI* site is located between the two points of homology for targeting which contain small deletions desired to by recombined into the genome (A* and B*). The vector also contains a marker gene (red box) and an *I-CreI* recognition site for the resolution of the duplication incurred. (B) The targeting vector is cut at the *I-SceI* site using the site-specific *I-SceI* endonuclease. This produces a DSB and allows the targeting vector to line up against the native genomic DNA using the regions of modified homology, and to insert via homologous recombination. (C) This process incurs a tandem duplication of the target region with both a WT and modified copy present. (D) A DSB is created using the I-*CreI* endonuclease on the engineered I-CreI site between the duplication. This process is repaired by homologous recombination presenting a number of potential resolutions (D). The required resolution contains only a single copy of the target genes, each containing the engineered mutation (E).

1.7 Putative downstream targets of Fru

Identifying specific functions for Fru isoforms gives insight to the downstream actions of Fru. However, to further investigate this, identification of the targets

these actions are instigated by would be required. One way of identifying such targets is by DNA adenine methyltransferase identification, more commonly known as DamID.

To date, our knowledge of *fru* is good in terms of its position within the SDH. Further, we are aware of the various behaviours in which Fru has a role in and are starting to understand its interactions with *dsx* and how these two genes control courtship behaviour in flies. Yet it remains unknown what genes lie downstream of *fru* in order for *fru* to exert its effects on the development of the CNS and behaviour. What does Fru transcriptionally target enabling it to have these wide ranging controls over a variety of behaviours? To be able to pursue this question, a new way of investigating *fru* must be considered: a whole genome approach.

Some methods have already been used to suggest some targets of Fru, such as the Ecdosyne receptor (EcR) [119]. This target was identified by microarray analysis (custom design [120]) of *fru* P1 transcripts deficient flies ($fru^{4\text{-}40/\ P14}$ and $fru^{w12/\ ChaM5}$) compared to WT flies (where WT were two strains: CS and *w*Berlin), EcR is suggested to be targeted by Fru during metamorphosis where the nervous system is re-modeled for adult function. Dalton *et al*. showed through courtship behavior assays, co-expression staining, and RNAi, that EcR-A (which is expressed in the CNS) is co-expressed in a subset of Fru expressing cells, and when reduced causes a significant increase in male- male courtship, a known phenotype of *fru* mutants, however EcR-A did not effect male-female courtship, suggesting a role specific to the suppression of male- male courtship [119].

Studies have shown that proteins containing a BTB protein binding domain often bind to other transcription factors of the same family [88]. Since *fru* is a member of this family, it is possible that it interacts with these other family members.

Potential interactors include *lola* and *ttk* [88, 121]. These examples have been shown to be involved in development of the brain and/ or courtship behavior [81, 122, 123]. Dimerising with other BTB- containing transcription factors would allow Fru to target many different types of genes by partnering with a variety of transcription factors.

1.8 DamID as a method to identify putative downstream targets of Fru

Recently genetics has benefited from the advances in whole genome analysis. For Fru, this allows the realistic possibility to identify its targets. By analysing Fru targets on a whole genome scale we might be able to create a better understanding of the distinct functions of *fruitless*.

DamID is a common method for identifying targets of a protein of interest (Figure 1-8) [1, 124]. The Dam protein was first identified in *Escherichia coli* (*E. coli*) as a protein that methylates the sequence GATC. DamID can be exploited in *Drosophila* as they have no native methylation of GATC sequences [14].

Fusion Protein

To identify binding sites of a protein of interest (POI), a fusion of Dam with the POI is designed, which methylates GATC sequences surrounding the genome at which the POI binds. This methylation spreads up to several kb from the point of binding identifying a large section of DNA which is within close proximity or encorporates the binding point of the target [125]. This methylating is controlled by a leaky UAS system [36].

Leaky UAS system

The leaky UAS system designed by Brand et al., produces low levels of transcript by exploiting a local, non-specific, strong promoter [22]. The effects of such a promoter, in this instance the uninducible minimal promoter Hsp70, is low level transcription (or 'leaky' transcription) of the Dam fusion gene, by proximity [1, 14, 22, 126]. This is necessary as Dam has been shown to be lethal at higher levels due to excessive non-specific methylation [79, 125]. To separate specific targets from those Dam is able to bind on its own, the Dam 'leaky' construct is used as a control for non-specific binding. The data can then be filtered to remove this non-specific binding to identify real targets of the gene of interest [1, 126].

Creating a dataset

Triplicate experimental and control samples are digested with *DpnI* which specifically cuts methylated GATC sequences, those being the points of methylation created by the Dam fusion [14]. Adaptors are then ligated to the 'sticky ends' of the GATC sequences protecting them from *DpnII* digestion, which digests away the remaining DNA (not targeted by the GOI), leaving only the target regions of DNA [14]. These sequences are then amplified using a PCR process and are fluorescently labelled with either Cy3 or Cy5 (incorporating a dye-swap replicate to account for differences between the two fluorophores) and hybridised to a tiling array to assess the results [14].

DamID is an effective technique as it is sensitive to transient DNA protein interactions, requires relatively little material and gives high reproducibility between samples. However some disadvantages include the spread of methylation, reported as far as 2- 5 kb from the binding site resulting in potentially

several genes in the target area in which a single gene is the true target. It is currently not possible to define flanking DNA. DamID represents the average protein binding in cells over time and so may under represent some targets. Further, it cannot prove transcription occurs, only state that transcription machinery sat on the DNA at this point, and cannot define whether transcription has been repressed or activated. To compensate for these disadvantages, bioinformatics analysis is used to identify true targets.

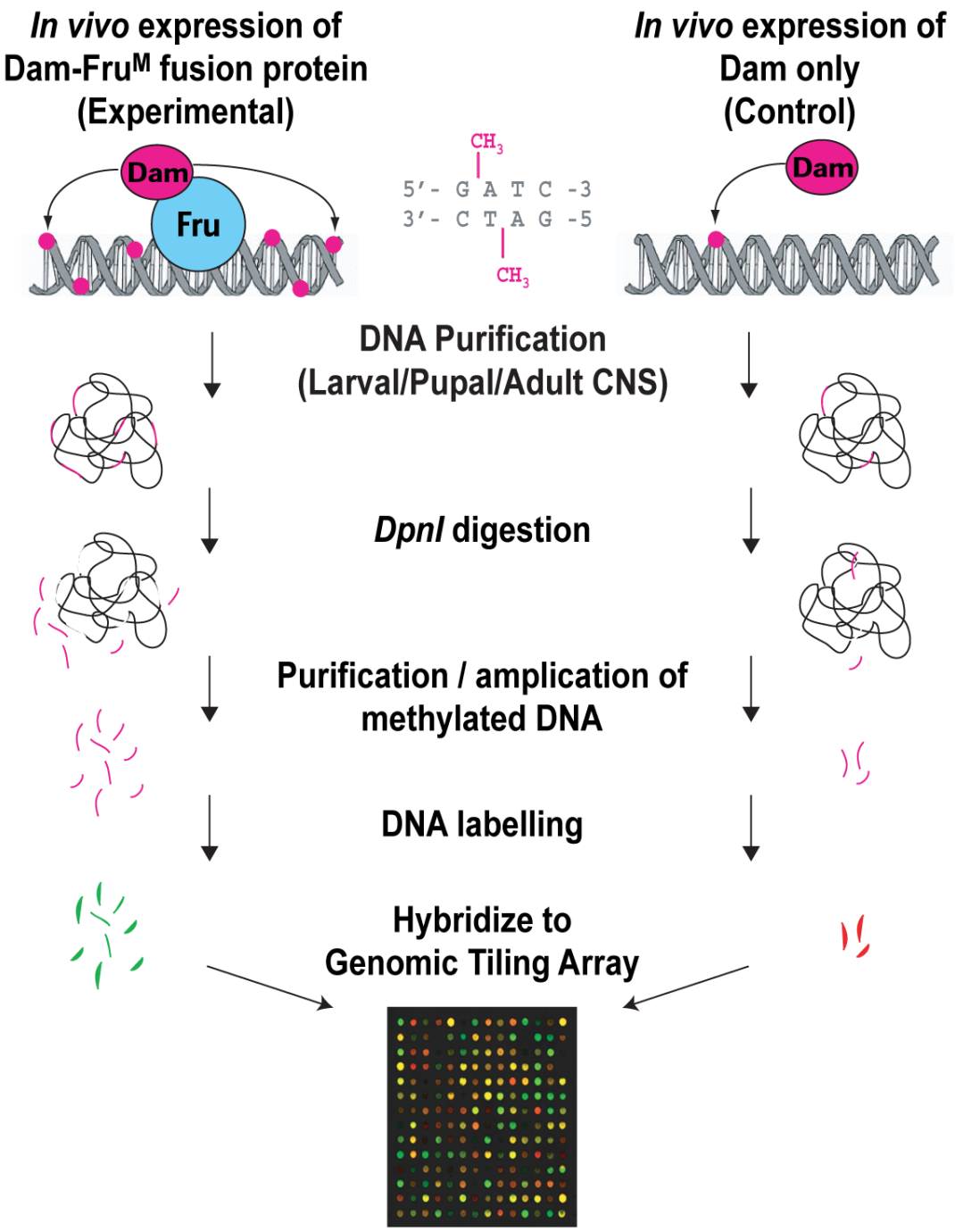


Figure 1-8 Schematic of genome-wide mapping of Fru^{M} binding sites using DamID.
A schematic diagram of the DamID technique. DamID is based on the creation of a fusion protein consisting of *E. coli* adenine methyltransferase (Dam) and the protein of interest (Fru). Upon trace expression, Dam methylates the GATC sequences surrounding the points on the genome, which are targeted by the fused protein of interest. *DpnI* is used to digest out the targeted DNA, which is then protected from further digestion by the ligation of an adaptor. Remaining non-targeted DNA is degraded away and the targeted DNA is amplified before being labelled with fluorescent markers, and hybridised on a genome tiling array (Roche Nimblegen). Each array contains 2.1 M probes of approximately 45 bases in length at about 300 bp intervals covering the *Drosophila* euchromatic DNA on a single array. M. Neville, unpublished.

1.9 Aims and objectives

Given the diversity of processes that *fru* regulates, we want to understand how this one locus with its multiple promoters and isoforms can generate a range of functions which combine to produce phenotypes associated with courtship behaviour. This body of work concerns understanding how *fru* expressed during development through to the adult fly has such a wide range of functions. To do this I will be studying the differences between sex-specific Fru isoforms: Fru^{MA}, Fru^{MB} and Fru^{MC}. These isoforms are expressed in specific cells in the nervous system, acting to masculinise the neurons. I will be using a mutational approach to investigate the function of specific Fru isoforms. In addition I will use a genome-wide approach to identify Fru transcriptional targets, helping us to understand how Fru can control the shape, physiology and survival of neurons during development that control male sexual behaviour in the mature fly.

2 Materials and Methods

2.1 Drosophila

2.1.1 Drosophila stocks:

All *Drosophila* stocks used in this work are described in Table 2-1. A full description of all mutations and balancer chromosomes used can be found in FlyBase (http://flybase.org).

Genotype/Strain	Phenotype	Reference
Canton S (*CS*)	Isogenic wild-type strain	Flybase, 2007
w^{1118}	*white*$^{-}$ mutant strain used for microinjections	Flybase, 2007
w^{h}; ap^{Xa}/ *CyO*; +	Strain isogenised to WT CS strain. 2nd chromosome balancer (*CyO*) and dominant marker (ap^{Xa})	Flybase, 2007
w^{+}; +; *TM3 Sb e*/ *TM6b Hu e*	Strain isogenised to wild-type CS strain. 3rd chromosome balancers.	Flybase, 2007
w^{1118}; +; *PUAST{NDam}*/ *TM3 Sb e*	*P*-element insertion on the 3rd chromosome carrying the DamID NDam construct with no fusion gene for control purposes.	M. Neville, unpublished.
w^{1118}; +; *PUAST{NDamfru*^{MA}*}*/ *TM3 Sb e*	*P*-element insertion of the Dam coding sequence fused to fru^{MA} on the third chromosome, balanced.	M. Neville, unpublished.
w^{1118}; +; *PUAST{NDamfru*^{MB}*}*/ *TM3 Sb e*	*P*-element insertion of the Dam coding sequence fused to fru^{MB} on the 3rd chromosome, balanced.	M. Neville, unpublished.

w^{1118}; +; PUAST{NDamfru^{MC}}/ TM3 Sb e	*P*-element insertion of the Dam coding sequence fused to *fru^{MC}* on the 3rd chromosome, balanced.	M. Neville, unpublished.
w^{1118}; +; PUAST{NDamfru^{mutMA}}/ TM3 Sb e	*P*-element insertion of the Dam coding sequence fused to *fru^{mutMA}*, with mutated DNA binding domain.	This study.
w^{1118}; +; PUAST{NDamfru^{mutMB}}/ TM3 Sb e	*P*-element insertion of the Dam coding sequence fused to *fru^{mutMB}*, with mutated DNA binding domain.	This study.
w^{1118}; +; PUAST{NDam}/ TM3 Sb e	*P*-element insertion on the 3rd chromosome carrying the DamID NDam construct with no fusion gene for control purposes.	M. Neville, unpublished.
M690 *y* w*/ Y hs-hid; Sco hs-I-SceI hs-FLP/ CyO*	*P*-element insertions on the Y and 2nd chromosomes; used during first stage of gene targeting by HR to excise the donor *P*-element from genome and create a DSB in the extrachromosomal DNA prior to integration at the desired locus.	Dickson, B.J. unpublished[1]
M699 *y* w* ; ey-FLP*	*P*-element insertion on 2nd chromosome, used during the first stage of gene targeting by HR.	Dickson, B.J. unpublished[1]
M700 *y* w* ey-FLP; Pin/ CyO*	*P*-element insertion on X chromosome in a 2nd chromosome balancer background, with *Pin* as a dominant marker for the other 2nd chromosome; used during the first stage of gene targeting by HR.	Dickson, B.J. unpublished[1]

M701 *y* w* ey-FLP; +; Ly / TM3 Sb e*	*P*-element insertion on X chromosome, in a 3rd chromosome balancer background, with *Ly* a dominant marker for the other 3rd chromosome; used during the first stage of gene targeting by HR.	Dickson, B. J. unpublished[1]
w; +; hs-I-CreI Sb*	*P*-element insertion on the 3rd chromosome; used to create a DSB in the targeted locus to resolve the duplication created in the first stage of HR.	Rong *et al.* 2002 [127].
Df(3R) fru^{sat15}/ TM3 Sb e	3rd chromosome deficiency. Extends centromere-proximal from within fru^{SAT} *P*-element insert. Removes entire *fru* common coding region, as well as several other genes.	Ito *et al.* 1996 [108].
Df(3R)fru^{w24}/ TM3 Sb e	3rd chromosome deficiency breaking at *fru* locus. Breakpoints at 91A and 91D2-3.	Ryner *et al*. 1996 [78].
w^{+}; Df(3R) fru$^{4-40}$/ Tm6, Hu, e	3rd chromosome deletion of at least 70 kb covering the *fru* locus and removing transcripts from P1 and P2.	Anand *et al.*, 2001 [34].
w^{-}; +; Df(3R)ED5812/ TM6C	3rd chromosome deletion mutation deleting 91B2-91B5.	DGRC

Table 2-1 Drosophila stocks ([1] - http://www.imp.ac.at/research/barry-dickson/) * represents an unknown exact allele for *w* or *y*.

2.1.2 Wild-type strain

Unless otherwise stated the WT line used in this thesis is *Canton S (CS)*.

2.1.3 Rearing conditions

2.1.3.1 Stocks

Stocks were kept at 18 ˚C when not in use and changed into new vials once every 20 days. When in use, lines were kept at 25 ˚C, 40 % humidity and changed into new vials once every ten days. One exception is the stock *w^{1118}; +; PUAST{NDamJ}/ TM3 Sb e*, a slow-growing stock, which was kept at 25 ˚C and turned once every 14 days. Both temperature controlled rooms are maintained in a 12: 12 hr. light-dark cycle. During this study, the lab moved from the University of Glasgow to the University of Oxford, therefore work completed in the first two years used flies brought up on Glasgow medium, and subsequently Oxford medium was used after the move.

2.1.3.2 Glasgow medium

10 g agar, 15 g sucrose, 30 g glucose, 35 g dried yeast, 15 g maize meal, 10 g wheat germ, 30 g treacle, 10 g soy flour per litre of distilled water.

2.1.3.3 Oxford medium

7.69 g agar, 20 g maize, 2.3 g soya, 4 g yeast, 70 g malt, 20 ml molasses, 5 ml acid mix (500 ml propionic acid, 32 ml orthophosphoric acid), 30 ml of 10 % nipagin per litre.

2.1.4 Drosophila microinjection

2.1.4.1 DNA solution

Plasmids for microinjection were purified from *E. Coli* DH5α (Table 2-4), using the Plasmid Maxi Kit (QIAGEN) following the manufacturer's instructions.

2.1.4.2 Microinjection

Microinjection was performed by Genetic Services Inc. (Cambridge, MA) 30 - 50 µg of DNA solution was microinjected into w^{1118} flies, which were returned at the larval stage, approximately 48 hr. after injection.

2.1.4.3 Selecting P-element transformants

Larvae developed at 25 °C in the food vial in which they were transported. Single adults were collected as they eclosed and were mated to w^{1118} flies from the Goodwin lab of the opposite sex. The resulting progeny were screened for eye colour. The *P*-element used in this study contains the *mini-white* (mW^{+}) gene, which confers red eye colour to flies in a w^{1118} background. Depending on the insertion, the strength of this eye colour varies, thus progeny of these crosses were screened for the presence of non-white eyes, indicating *P*-element integration.

2.1.4.4 Determining the location of the insert

To determine the chromosomal location of the insert, transformant flies were crossed to strains containing balancer chromosomes with different dominant genetic markers on either the second or the third chromosomes, as follows: CyO/ Xa (II), TM3/ TM6 (III). The *P*-elements used in this work contain the mW^{+} gene which is used as a marker when mapping the chromosomal location of the insert. The genetic background of all balancer stocks was w^{-}. To identify an insertion on the X chromosome, a male peach eyed fly is crossed to a female balancer. A result showing only white eyed males identifies an X chromosome insertion, Figure 2-1.

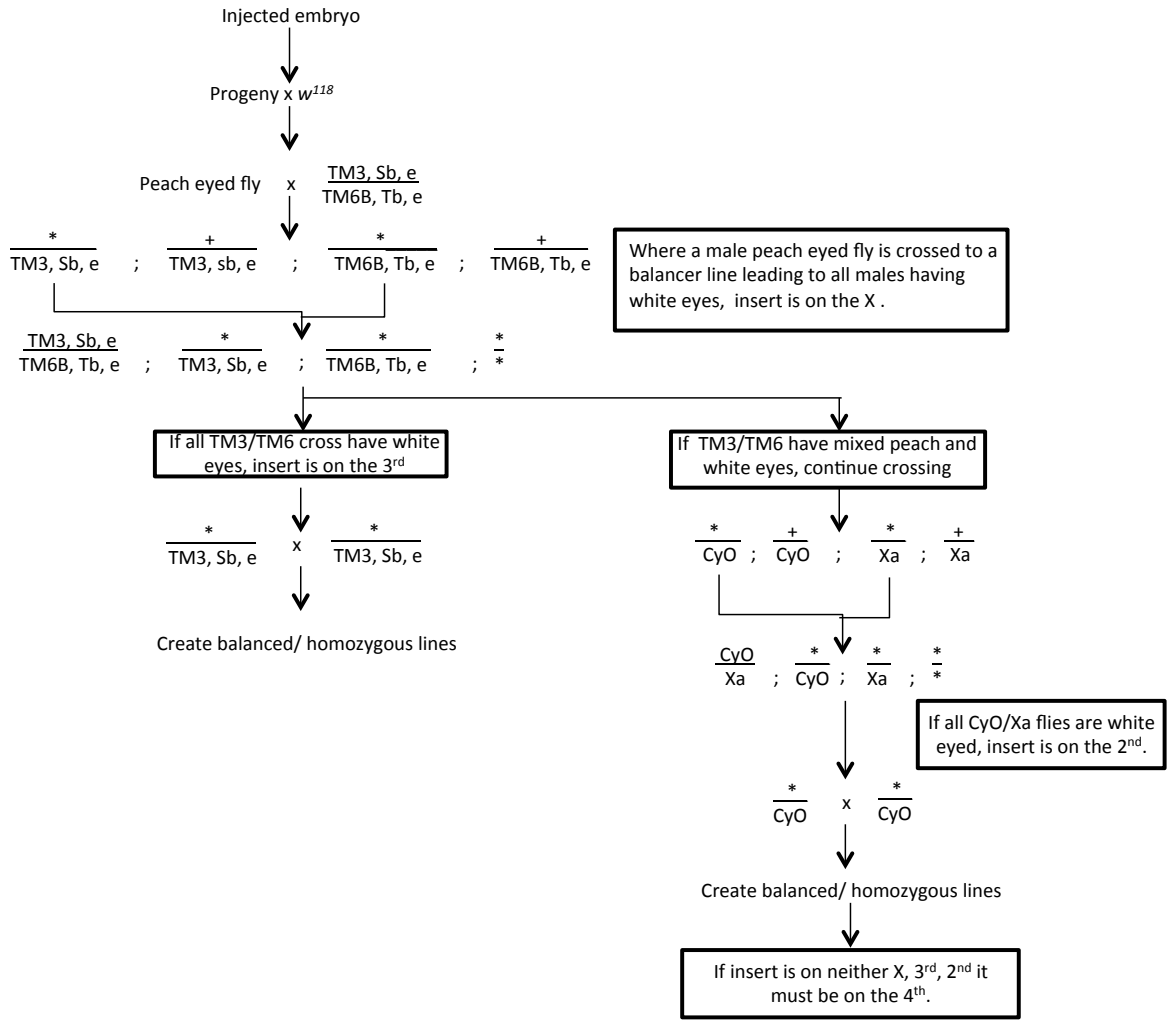


Figure 2-1 Crossing scheme for balancing chromosomes. The above series of crosses identifies the location of an insert in the genome. Once the location is identified, a line is established: homozygous viable, or a heterozygous stock over a balancer chromosome.

As a third chromosome insertion is preferred in this study, the mutants were first crossed to a third chromosome balancer. To determine whether an insertion occurred on the third chromosome, flies heterozygous for the presence of the *P*-element were crossed to *w^-; +; TM3 Sb e/ TM6b Hu e*. The orange eyed progenies *w^-; +; {w^+}/ TM3 Sb e* and *w^-; +; {w^+}/ TM6b Hu e* were then crossed together. Insertion on the third chromosome was determined by the absence of to *w^-; +; TM3 Sb e/ TM6b Hu e* with orange eyes.

Similarly, to determine insertions on the second chromosome, orange eyed flies with the *P*-element insertion were crossed to *w^h; ap^{Xa}/ CyO;* +. The orange eyed progenies *w^h; {w^+}/ CyO;* + and w^h; *{w^+}/ ap^{Xa};* + were then crossed together.

Insertion on the second chromosome was determined by the absence of w^{h}; ap^{Xa}/ *CyO;* + with orange eyes. After excluding chromosomes one, two and three the insertion was assumed to be on the chromosome four.

After determining the chromosome of insertion, each line was balanced with the appropriate balancer: CyO for a second chromosome insertion and TM3 or TM6B for a third chromosome balancer. Homozygous lines were kept as both a homozygous stock or as a homozygous stock with floating balancers. Lethal lines were kept as heterozygotes with the appropriate balancer chromosome.

To account for other *de novo* mutations incurred during the injection, and subsequent determination of the chromosomal location of insertion, *P*-element containing stock lines were outcrossed to a WT strain for six generations.

2.1.5 Collection of tissues: Dissecting techniques

2.1.5.1 Larvae

L3 'wandering' larvae (L3 larvae leave the food and wander up the vial, before eventually everting their spiracles and stopping) were collected and sexed by the size of the developing gonads. Male L3 larvae have larger gonads than females [128]. Males were collected in PBS or DamID buffer (2.4.14.2). To dissect the brain and VNC, the larvae were dissected by tearing the larvae in half with the forceps. The posterior end was discarded. A midline tear was created from the open posterior section along the dorsal midline towards the anterior end of the larvae. The brain was isolated from the fat and imaginal discs, and transferred to a pre- chilled 24-well culture dish containing 2 % paraformaldehyde (PFA) (or a 1.5 ml Eppendorf pre-chilled on dry ice if following the DamID protocol) using a

siliconised pipettor and incubated for 30 mins on ice (or flash frozen on dry ice for DamID protocol and stored at -80 ˚C), before continuing with the required staining protocol. For DamID, 45 larval CNSs were used per experiment.

2.1.5.2 Pupae

L3 wandering larvae were picked and sexed by observing the developing gonads, as described above (2.1.5.1). Males were collected and transferred to a Petri dish lined with Whatmann paper soaked in PBS (2.4.14.3). After 46 hr. the pupae were dissected. The pupal casing was removed, starting with the operculum, using forceps. The rest of the pupal casing was peeled back to reveal the pupa. As in the larval stage, the pupa was ripped at the mid-point and the posterior section discarded. Using a P20 Gilson pipette, buffer was pipetted onto the open end of the pupa to dislodge the fat. As the VNC became visible, the pupa was ripped further up towards the brain until the whole CNS was visible. The CNS was lifted from the pupal shell from the VNC upward until it was completely free. The brain and VNC were transferred to a pre- chilled 24 - well culture dish containing 2 % PFA (or a 1.5 ml Eppendorf pre-chilled on dry ice if following the DamID protocol) using a siliconised pipettor and incubated for 30 mins on ice (or flash frozen on dry ice for DamID protocol and stored at -80 ˚C), before continuing with the required staining protocol. 45 pupal CNSs were required per DamID replicate.

2.1.5.3 Adult

Virgin males were isolated from females and left to mature in small groups for 48 hrs. At this point, the fly was anesthetised with carbon dioxide gas, and dipped briefly in ethanol before being transferred to a dissecting dish containing PBS (or DamID buffer for use with DamID protocol). Using forceps, the wings and legs

were removed. The proboscis was removed, and the opening created was used to tear the head capsule to expose the brain. Trachea and fat body were removed being careful not to break the nerve trunk connecting the brain to the VNC. Dissection scissors were used to cut off the abdomen at the bottom of the thorax from which a cut was made transversely from anterior to posterior along the length of the thorax just above the wing nodes on each side. The top of the thorax was lifted away from the remaining fly, separating the majority of the wing muscles with it. The remaining wing muscles were removed, care was taken not to dislodge the VNC, and any remaining trachea surrounding the VNC were cleared to detach the bottom of the VNC. The VNC was lifted out towards the brain, taking the brain with it. Any remaining trachea/ debris was removed, and the brain and VNC transferred to a pre- chilled 24 - well culture dish containing 2 % PFA (or a 1.5 ml Eppendorf pre-chilled on dry ice if following the DamID protocol) using a siliconised pipettor and incubated for 30 mins on ice (or flash frozen on dry ice for DamID protocol and stored at -80 ˚C), before continuing with the required staining protocol. For DamID 30 adult brains and VNC were used per replicate.

2.1.5.4 Muscle of Lawrence

Males were isolated from females for five to seven days at 25 °C. The flies were anesthetised with carbon dioxide gas, and dipped in ethanol before being transferred to a dissecting dish containing PBS. The abdomen and head were removed by cutting the abdominal-thoracic join with dissection scissors. A ventral cut posterior to anterior, down the central ventral abdomen, was made to the point of the last cuticular ridge. Similar cuts were made posterior to anterior down the outer edges of the cuticular dorsal side of the carcass and the internal organs

of the fly were removed retaining only the hard cuticular dorsal side of the thorax. The remaining cuticle was trimmed as necessary to allow it to sit flat on the microscope slide. The cuticle was transferred to a pre- chilled 24- well culture dish containing 2 % PFA and incubated for 30 mins before continuing with the required staining protocol.

2.2 Visualisation techniques

2.2.1 Immunoflourescence

Samples were dissected in PBS and fixed in 2 % paraformaldehyde (in PBS) for 30 mins on ice. Dissected brains were washed for five minutes, three times in PBS, and for 10 mins, three times in PBT (1 x PBS, 0.4 % Triton X-100). The samples were blocked in PTN (PBT, 5 % NGS) for one hour. The CNSs were incubated overnight in primary antibody diluted in PTN at 4 ˚C (Table 2-2), and then washed five times for 20 mins in PBT. The samples were incubated in secondary antibody diluted in PTN overnight at 4 ˚C (Table 2-3). Washes were repeated (5 x 20 mins in PBT) followed by 5 washes for 10 mins in PBS. The samples were left in the final PBS wash overnight at 4 ˚C. The samples were mounted the following morning in VectaShield (Vector Lab) on polylysine treated microscope slides (VWR International) and stored at 4 ˚C.

For conjugated antibodies, the same protocol was followed, bar the second staining step and washes. Therefore, after washing in PBT, post primary antibody incubation, the samples were washed in PBS before mounting in Vectashield.

Antibody	Description	Dilution	Source/ Reference
Anti-Fru^M	Rabbit, polyclonal against male-specific 101 aa domain Fru^M	1: 500	Billeter *et al*., 2006 [2].
Anti-Fru^A	Rat, polyclonal against zinc finger domain of Fru^A	1: 100	Song et al. 2002 [80].
Anti-Fru^B	Rat, polyclonal against zinc finger domain of Fru^B	1: 200	J Walker, unpublished.
Anti-Fru^C	Rat, polyclonal against zinc finger domain of Fru^C	1: 200	J Walker, unpublished.
Anti-Nc82	Stains against neuropil	1: 10	DSHB

Table 2-2 Primary antibodies used in this study

Antibody	Description	Dilution	Source/Reference
Anti-rat IgG-Alexa Fluor 488	Goat, polyclonal against rabbit IgG, conjugated to Alexafluor 488.	1: 300	Molecular Probes, Inc.
Anti-rabbit IgG-Alexa Fluor 546	Goat, polyclonal against rabbit IgG, conjugated to Alexafluor 546	1: 300	Molecular Probes, Inc.
Anti-mouse IgG-Alexa Fluor 546	Goat, polyclonal against mouse IgG, conjugated to Alexafluor 546	1: 300	Molecular Probes, Inc.

Table 2-3 Secondary antibodies used in this study

2.2.2 Confocal microscopy

Stained whole mounts were examined with a Leica SP5 equipped with x 10, x 20, x 63 (oil immersion) objectives. The Argon (Ar) and Helium Neon (HeNe) lasers can excite fluorochromes within the 458 nm, 476 nm, 488 nm, 514 nm, 543 nm and 633 nm visible spectrum. Alexa fluor-488-conjugated secondary antibodies were excited with the 488 nm line and emit in the green spectrum. TRITC-conjugated phalloidin and Alexa fluor 546-conjugated secondary antibodies were excited with the 543 nm line, and emit in the red spectrum.

For double-labelling, the 488 nm and 543 nm lines were sequentially used for each optical section through the sample to excite both fluors individually and avoid bleed-through. Stacks of optical sections were generated at 1 μm spacing.

The MOL, stained with TRITC-conjugated Phallodin, was visualized using a Zeiss Lumar V12 equipped with 80 x magnification.

2.2.3 Image processing

After acquisition, confocal images were opened in Leica software LAS AF and converted to 'Maximal-Z-projections' which superimpose serial optical sections into a single image, allowing consideration of the overall expression patterns. Brightness and contrast of images was altered in Photoshop (Adobe).

2.2.4 Processing of images for figures

Pseudocolours were added to images generated by the confocal microscope using Leica LAS AF. Alexa fluor 488 signal was coloured green, and Alexa fluor 546 magenta. Unless otherwise stated, all images in this study are maximal Z stacks showing multiple slice images of the tissue stacked together to show overall staining patterns.

Images obtained using the Leica microscope were processed using Adobe Photoshop 7.0. Adobe Systems Inc., (San Jose, CA). White balance was adjusted using the 'Level' Function. Brightness and contrast were also adjusted as necessary. Images were cropped and saved as .TIFF files.

Figures were assembled using Adobe Illustrator 10 (Adobe Systems Inc., San Jose, CA). Images were saved as Illustrator images: .ai.

2.3 Bacterial protocols

2.3.1 Strains and plasmids

Strain	Genotype	Reference
DH5α	F^-, ϕ80d*lacZΔ*M15*Δ (lacZYA-arg*F) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_k^- ,m_k^+) *pho*A *sup*E44 λ^- thi-1 gyrA96 relA1	Hanahan *et al*., 1983 [129].

Table 2-4 Bacterial strains and plasmids

The plasmids used in this study, whose construction is not otherwise described elsewhere, are listed in Table 2-5.

Plasmid	Size (bp)	Description	Antibiotic Resistance	Reference
*p*ED7	8,090	Vector for gene targeting by HR. Contains multiple cloning sites (MCS), *P*-element transposition sequences, a *mW*+ marker gene and I-*Cre*I recognition site.	Ampicillin	Gift from B. J. Dickson, unpublished.
*p*Sc-B	3,500	Small vector used in this study to sub-clone fragments for mutagenesis. Contains MCS, *Lac-Z*, pUC origin, f1 origin, loxP sites.	Ampicillin	Strataclone
p-ZErO™-2	3,297	Small vector used in this study to sub-clone fragments for mutagenesis. Contains MCS, *Lac-Z*, f1 origin, pMB1 origin.	Kanamycin	Invitrogen

Table 2-5 Plasmids used in this study

2.3.2 Culture media

SOC: (**S**uper **O**ptimal **B**roth with catabolite repression) 2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose, pH 7.0.

L-Broth: 10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl per litre of water; pH adjusted to 7.0 with NaOH.

L-Agar: As above with the addition of Bacto-agar (Difco) to 1.5 % (w/v).

NZY+ Broth: Combine 10 g NZ-amine, 5 g yeast extract, 5 g sodium chloride and dissolve in de-ionised water to a final volume of 1000 ml using sodium hydroxide, pH to 7.5 and autoclave. After autoclave add additional filter sterilised supplements: 12.5 ml 1 M magnesium chloride, 12.5 ml 1 M magnesium sulphate, and 20 ml 20 % w/v glucose.

All culture media were sterilised by autoclaving at 120 ºC for 15 mins.

2.3.3 Antibiotics and indicators

Antibiotics were added to liquid broth at room temperature, and to molten agar at 50 ºC. Final concentrations of antibiotics used were: 100 µg/ml for ampicillin (Sigma), and 25 µg/ml for kanamycin (Sigma).

2.3.4 Transformation techniques:

2.3.4.1 Heat shock using DH5α competent cells:

DH5α cells were thawed on ice for 20 mins (from -80 ˚C). 1 µl DNA was added and incubated for 20 mins on ice. The cells were heat shocked for 1 min at 42 ˚C.

The solution was made to 1 ml with LB and left to recover at 37 ˚C for 1 hr., shaking at 250 revolutions per minute (rpm). 100 μl was plated on an agar plate containing suitable antibiotic resistance and incubated overnight at 37 ˚C.

2.3.4.2 Electroporation using ElectroTen-Blue cells:

A cuvette (0.2 cm gap) and shocking chamber were pre-chilled for 10- 15 mins at -20 ˚C. SOC broth was pre-warmed to 37 ˚C in a waterbath. 1 μl DNA was added to 40 μl cells. The Biorad MicroPulser™ electroporator was turned on and the cells/ DNA solution transferred to the pre-chilled cuvette and inserted in to the shocking chamber. 'Pulse' (2.5 kV) was pressed, and immediately 960 μl SOC was added, pipetting up and down to mix the solution. The solution was transferred to a fresh tube, and recovered at 37 ˚C for 1 hr., shaking vigorously. The solution was spun down, and the full quantity plated onto an agar plate with suitable antibiotic resistance. The plate was incubated overnight at 37 ˚C.

2.4 General molecular biology protocols

2.4.1 Restriction digest of DNA

DNA was digested in the appropriate buffer according to manufacturer's recommendations (New England Biolabs; NEB). 2 units of enzyme were added per μg of DNA to ensure complete digestion. Digests were carried out at 37 °C (unless specified otherwise) for 1 hr. for plasmid DNA, or up to 4 hrs for genomic DNA.

2.4.2 Miniprep

Minipreps were performed using Qiagen Miniprep kit (Qiagen) following manufacturer's instructions.

2.4.3 DNA gel electrophoresis

DNA fragments were resolved on agarose (Boehringer Mannheim) gels. The DNA solution was mixed 6: 1 with loading buffer (6 x) and loaded into agarose gel containing 1 x TAE (40 mM Tris/ AcOH (pH 8.2), 20 mM NaOAc, 1 mM EDTA). A range of agarose concentrations (0.5 -2 % (w/v)) were used depending on the size of the fragments to be resolved, with higher concentrations being used to separate out smaller fragments (< 500 bp) [130]. The gels were run from 1 hr. to overnight at voltages between 25 mV to 140 mV (Gels were run on a lower voltage and for a longer time to prevent a sample from being separated, or to separate out very similar sized fragments). DNA size markers used were either 100 bp or 1 kb (NEB) depending on expected fragments.

2.4.4 Visualisation and photography of gels

DNA was visualised using either a short wave (254 nm) or long wave (365 nm) transilluminator after staining the agarose gels in ethidium bromide (EtBr, 0.5 µg/ml) in a shaker tray for 40 mins. Gels were photographed using a Polaroid camera loaded with Polaroid 667 film and fitted with a Kodak Wratten filter No. 23A or the Uvidoc gel documentation system (Uvitec, Cambridge).

2.4.5 Gel extraction of DNA fragments

DNA fragments were excised from agarose gels using a clean razor blade, on either the long wave (365 nm) UV transilluminator or Uvidoc system. Purification of the DNA fragment from agarose, salt and enzymes was carried out with Qiaquick gel extraction kit (QIAGEN), or QIAEXII Gel extraction kit (for fragments > 10 kb) following manufacturer's instructions.

2.4.6 Subcloning DNA into plasmid vectors

The insert to be subcloned, and the target plasmid vector (Table 2-5) were digested with compatible restriction enzymes, as described in 2.4.1. To confirm the digests, resulting DNA was run out on an agarose gel, and gel extracted as described in 2.4.3 and 2.4.5.

2.4.7 Ligation

Purified inserts and vectors were mixed at a ratio of 3: 1 (insert: vector) for cohesive-end ligation. The DNA mix was combined with 1 µl (1 Weiss Unit) of T4 DNA ligase and 4 µl of 5 x T4 DNA ligase buffer (NEB) in a total volume of 20 µl. The ligation mixture was incubated overnight at 4 °C and transformed as described in 2.3.4 and 2.3.4.2.

2.4.8 Ligation of plasmid containing large inserts

2.4.9 Southern blot

Southern blot was performed using a biotin label approach. Biotin is a non-radioactive Southern blot option circumventing the safety risks associated with

radioactivity, traditionally used for Southern blot. Biotin is also more specific and sensitive [131]. Most relevantly, biotin probes can be re-used for up to three months (NEB, personal communication).

2.4.9.1 Purification of genomic DNA

Fly genomic DNA (gDNA) was prepared using a modified Hamilton and Zinn protocol [132]. 0.04 g of flies (approx. 30 flies) were frozen on dry ice and homogenised with a micro-pestle in 400 µl grinding buffer (5 % sucrose, 80 mM NaCl, 100 mM Tris, pH 8.5, 0.5 % SDS, 50 mM EDTA). The homogenate was then incubated at 70 ˚C for 30 mins, 70 µl 8 M potassium acetate added, and centrifuged at 12,000 g for 5 min. The supernatant was removed, and two phenol chloroform extractions were performed (Equal volumes of phenol chloroform were added, vortexed for 15 secs, spun for 2 mins, and the aqueous layer retained). The aqueous layer was incubated on ice for 30 mins and the sample centrifuged for 10 mins at 4 ˚C. The supernatant was removed to a fresh Eppindorf and 300 µl isopropanol added and mixed. The sample was incubated at room temperature for 5 mins, centrifuged for 10 mins and rinsed in 70 % ethanol. The sample was dried and resuspended in 30 µl TE buffer/ RNase A (2 µl RNase in 1 ml TE) and dissolved at 55- 60 ˚C for 1- 3 hrs.

2.4.9.2 Digestion, gel electrophoresis and blot

The DNA samples were digested all day with an appropriate restriction enzyme at 37 ˚C, and run on a 0.8 % agarose gel overnight at 25 V. The Southern blot was set up: The gel was placed face down on Whatmann paper, which acted as a wick to draw 10x SSC from a reservoir. On to the gel was placed a piece of Hybond-NX (Amersham Biosciences) and three further sheets of Whatmann

paper (all cut to the size of the gel, and pre-soaked in 10x SSC). A stack of paper towel was placed on top to draw through the 10x SSC solution, which was weighed down, and Parafilm lined around the gel to prevent solutions from being drawn away from the membrane. Between each layer it was important to ensure no bubbles were present which would prevent accurate transfer of the DNA. The stack was left overnight to allow the DNA to be transferred from the gel to the membrane.

2.4.9.3 Making the probe, and determining the level of biotinylation

Primers spanning the area to be probed were used in a standard PCR amplification. Two 25 µl reactions were required for sufficient DNA. The DNA was run on a 1.5 % agarose gel at 80 V for 1½ hrs. using wide lanes from which the DNA was cut and gel extracted using the Qiagen Gel Extraction Kit (Qiagen). The DNA was quantified on an agarose gel, and using a Nanodrop (Thermo Scientific). 10- 500 µg of DNA was denatured in boiling water for 5 mins, incubated on ice for 5 mins and centrifuged briefly at 4 °C. To each sample the following reagents were added in the given order: 10 µl 5 x labelling mix; 5 µl dNTP mix (biotinylated dATP); 1 µl Klenow. The samples were incubated for 6 hrs at 37 °C, and the reaction terminated with the addition of 5 µl 0.2 M EDTA, pH 8.0. At this point the probe can be stored at -20 °C.

The probe was cleaned using Qiagen Nuclease Removal Kit (Qiagen) following the manufacturer's instructions.

To determine the level of biotinylation of the probe, serial dilutions for each probe were made (1: 10) in 0.1 M NaOH of biotinylated probe, control DNA and 2-log DNA ladder. 1 µl samples were spotted onto a membrane from full concentration

to 10^{-6} (7 spots per sample, total) and allowed to dry for 15 mins. The DNA was covalently bound to the membrane in a UV stratalinker, on the autocrosslink setting. The membrane was washed in a series of solutions at room temperature (Phototope®-STAR detection kit, NEB.): detection solution A (5 % SDS, 25 mM Sodium Phosphate, pH 7.2, 125 mM NaCl), 20 ml for 5 mins; Streptavidin incubation (20 µl Streptavidin in detection solution A), 20 ml for 5 mins; 2 washes of detection solution B (1: 10 dilution of Detection Solution A), 50 ml for 5 mins; biotinylated alkaline phosphatase incubation (20 µl alkaline phosphatase in detection solution A), 20 ml for 5 mins; detection solution A, 50 ml for 5 mins; 2 washes of detection solution C (100 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, pH 9.5), 50 ml for 5 mins; detect DNA (40 µl STAR reagent in 1 X STAR dilution buffer), 30 ml for 5 mins.

The membrane was exposed for a series of time points (1, 5, 10 mins). The probes should be of similar exposure to the control DNA (10^{-5} -10^{-6}), the 2-log ladder should be visible to 10^{-3}.

2.4.9.4 Hybridisation

15 ml of pre-hybridisation buffer (1 % SDS, 5 x SSPE, 5 x Denhardts, 50 mM phosphate buffer, 100 µg/ml sheared herring sperm DNA) was warmed to 42 °C in a hybridisation oven. The DNA was crosslinked using a UV Stratalinker on setting 'auto-crosslink'. The membrane was added to the pre-hybridisation buffer ensuring full contact with the side of the hybridisation cylinder, bubbles will prevent clean exposure of the membrane. The cylinder was returned to the oven and incubated for a minimum of 3 hr. at 42 °C.

The probe was boiled for 5 mins to denature and incubate on ice for 5 mins. The sample was briefly centrifuged to collate the sample, and the full sample was added to the pre-hybridisation buffer. The membrane was added to the probe solution overnight at 45 °C.

2.4.9.5 Washing the membrane, detection and exposure

The probe solution was poured off into a Falcon tube, labelled clearly, and frozen at -20 °C for future use. The membrane was washed to remove all traces of unbound probe: Initial wash (2 x SSC, 0.1 % SDS) for 20 mins, twice at room temperature; wash 1 (0.5 x SSC, 0.5 % SDS) incubated at 68 °C for 30 mins, twice; Rinsed in Maleic Acid Buffer (100 mM Maleic Acid, 150 mM NaCl, pH 7.5), shaking for 5 mins at room temperature.

Detection was the series of washes as described for detection of the probe in 1.4.8.4.

2.4.10 Quantification

Estimation of DNA concentration in an agarose gel was done by comparing a given band with the 3 kb band of the 1 kb ladder (NEB), which contains approximately 125 ng of DNA when the ladder is prepared according to the manufacturer's instructions.

Where more accurate Quantification was required, samples were measured using the Nanodrop 1000 (Thermo Scientific) which gives an accurate reading for samples containing a concentration of DNA between 2 – 3700 ng/µl double stranded DNA (dsDNA).

2.4.11 Polymerase chain reaction (PCR)

Taq polymerase (NEB) was used for diagnostic PCR. The PCR mix was set according to manufacturer's instructions. PCR was carried out using either a Robocycler Gradient 40 thermal cycler (Stratagene) or a T personal Thermocycler (Biometra). Standard conditions were: initial denaturation (5 min at 95 °C); denaturation (50 sec, 94 °C); annealing (40 sec 50- 60 °C); extension (1 min, 72 °C), where denaturation through extension were repeated up to 29 times. The cycle was completed with a final extension of 5 min at 72 °C.

2.4.11.1 Primer design

Primers used in this study were designed using MacVector 7.2.2 (Oxford Molecular) software purchased from Invitrogen. The Tm of primers was also calculated using this software. Primers were ordered from Sigma-Aldrich (Dorset, UK). A description of the primers used in this study can be found in Table 2-6 Primers used in this study.

2.4.12 DNA sequencing

Automated sequencing was carried out by Source Bioscience using Sanger sequencing SpeedREAD™ technology. Primers used for sequencing can be found in Table 2-6.

2.4.12.1 Primers

Primers used in this thesis are listed in Table 2-6.

Primer	Sequence	Tm (℃)
FruA-rt-rev1	GGTCGTGCTTAATGTTAAAACA	63.0
FruB-rt-rev1	GTTAGCCGCATGTGCCAGGCCTTGGA	80.2
FruB-rt-rev2	GCCGCACAGCTTGCAACGGT	74.9
FruC-rt-rev1	ATTGCCCGATCCCTGATTGC	70.2
5Ccom-for	CCAACTTCAAGTCGAGTCCC	63.6
HRFRT1R1	CAACTGAAGGCGGACATTG	56.7
HR_START	TATAATACATCCCATTTTCAT	48.1
7AF	CCCACTCTATACCGAGCCGTTAGTGCCTTGTTGTAGCACAGG	82.5
7AR	CCTGTGCTAGAACAAGGCACTAACGGCTCGGTATAGAGTGGG	82.0
7BF	CCGCAGAGTCATGAGGCCCCCGTTGCCGGGCAGAAATTCACTCGC	93.8
7BR	GCGAGTGAATTTCTGCCCGGCAACGGGGGCCTCATGACTCTGCGG	93.8
dREPCRAfor	GCCAAGTGATTTGATCCCTATTCC	58.4
dREPCRArev	GGTGGGTGTTTTGCCAAAG	56.7
REPCRB2for	TTCCCATCCCGTTCAAAAGTG	57.9
REPCRB2rev	GCCGACGAGGACATTAGG	58.2
EA_HRAFor	CGAAATGGAAATGTCCCGAATC	58.4
EA_HRARev1	TGGAGTTGGTGGTTTGCGTG	59.4
EA_HRARev2	GCTATTGAAACCGAGAGACTGAACCC	62.7
Adt	CTAATACGACTCACTATAGGGCAGCGTGGTCGCCGAGGA	82.8
Adb	TCCTCGGCCG	29.0
AdR-PCR	GGTCGCGGCCGAGGATC	71.9
EA_DIGA_FOR1	CGAAATGGAAATGTCCCGAATC	68.2
EA_DIG_REV1	TGGCGTTGGTGGTTTGCGTG	69.4
EA_DIGB_REV2	GCTATTGAAACCGAGACTGAACCC	67.6
EA_DIGB_FOR1	CAGATAGCGTCCAATTTCAAGCC	67.5
EA_DIGB_REV1	GGGGTTTTTCTTTTGAGGAGGTG	67.5
EA_DIGB_FOR2	CAGAGCCCCAGAGATAAATATCGAC	66.6
EA_DIGB_REV2	GCCAAAAAAACGCTGATGCGG	67.6
ZNF_7	TTCGTGAGATGGAAGCGGGGAA	75.3
ZFR1	CGCCCTTAAATGGATGAGTTCA	67.4
1ZF1_FOR	CGGGCACCATGTGGCGGGCCCGCTCCGCCGGCAAGGAGGTGACC	99.6
1ZF1_REV	GGTCACCTCCTTGCCGGCGGAGCGGGCCCGCCACATGGTGCCCG	99.6
1ZF2_FOR	CCGCCCAGCGGTCCATGGCTCCCTACGCCCCGGCCACCTACAGC	95.9
1ZF2_REV	GCTGTAGGTGGCCGGGGCGTAGGGAGCCATGGACCGCTGGGCGG	95.9
25ZF1_FOR	GCTGCAACCTGCACCGTGCCAAGCTGGCCGGCAAGGTCGTCACGC	96.6
25ZF1_REV	GCGTGACGACCTTGCCGGCCAGCTTGGCACGGTGCAGGTTGCAGC	96.6
25ZF2_FOR	CCGGGCCGCTTCGAGGCTCCGCTGGCCCGGGCCACCTATACGC	96.3
25ZF2_REV	GCGTATAGGTGGCCCGGCCAGCGGAGCCTCGAAGCGGCCCGG	96.3

Table 2-6 Primers used in this study.

2.4.13 Multi-site directed Mutagenesis (mSDM)

In 10 x reaction buffer, 5-50 ng template dsDNA, 125 ng of primers and 1 µl dNTP mix was made to a total volume of 50 µl. 2.5 U *Pfu* turbo DNA polymerase was added and the sample PCR amplified according to Table 2-7.

Cycles	Temperature (°C)	Time (secs)
1	95	30
	95	30
	55	60
12-18	68	60/ kb plasmid length

Table 2-7 PCR cycling programme for SDM

2.4.14 Buffers

2.4.14.1 TAE buffer

Combine 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA. This solution is made at a concentration of 50 X and diluted to 1 X as required.

2.4.14.2 DamID buffer

Combine 0.5 ml 0.5 M EDTA, 5 ml 1 M Tris-HCl pH 8.5, and 2.74 ml 2.5 M NaCl. Make to 50 ml with distilled water and autoclave.

2.4.14.3 PBS (Phosphate buffered saline)

Combine 8 g NaCl, 0.2 mg KCl, 1.44 mg $NaHPO_4$ and make to 100 ml with distilled water. Autoclave.

2.5 Bioinformatics

2.5.1 Corrections

2.5.1.1 Familywise error (FWE)

Accounting for a cumulative type I error, FWE represents the probability that any one comparison contains a type I error. As more comparisons are compared, the cumulative effect is taken into account:

$$\alpha_{FWE} \leq 1 - (1 - \alpha_{EC})^C$$

where α_{FWE} is the familywise error rate, α_{EC} is the error rate for an individual test (in this study, 0.05), and *C* is the number of comparisons (datasets).

2.5.1.2 Bonferroni

A highly flexible correction which maintains a specific value for the familywise alpha value (usually 0.05):

$$\alpha_B = \frac{(\alpha FWE)}{c}$$

Where α_B is the new alpha (based on the Bonferroni correction), αFWE is the familywise error , and *C* is the number of comparisons (datasets).

Bonferroni assumes that the null hypothesis is true and that all tests are independent. This can lead to an overcorrection for type I errors, and so there are several additions, of which in some instances of this study the Holm-Bonferroni method has been used.

2.5.1.3 Holm-Bonferroni

This method has greater power than the standard Bonferroni while retaining the flexibility of the correction. By using a series of steps which correct for the previous steps, rather than the whole sequence it provides high power correction without confidence intervals.

2.5.2 R

R (derived from S) was used to create intensity tracks for each DamID dataset which were uploaded to UCSC genome browser (http://genome.ucsc.edu/). The following script converts moving average files (.gff) to wiggle files (.wig):

```
#file format of original file must be seqname_start_end_score
t=as.data.frame(read.table ("MAadult_initial.tabular",
     skip=1))
length(t$V3)         # To find length of data set
a=t$V3 [-2032590]    # Edit figure in square brackets to
                      length of dataset - output of previous
                      line
b=0
t$V5<-c(b,a)# Bind Dataset Sizes to column V5



 ## FOR LOOP ITSELF

result<-c()
for(i in 1:length(t$V2))
{if(t$V5[i] < t$V2[i]){result[i]<-t$V2[i]}
else {result[i]<-t$V5[i]+1}}
 #Result should now be populated with values

t2<-cbind(t,result)
write.table(t2, file="MAadult_export.txt", sep="\t",
col.names=TRUE)
t3=as.data.frame(read.table ("MAadult_export.txt", skip=1,
colClasses=c("character", "character", "numeric", "numeric",
"numeric", "numeric", "numeric")))
t4=cbind(t3$V2,t3$V7,t3$V4,t3$V5) #Re-orders columns
```

```
write.table(t4, file = "MAadult_export_final.txt", sep =
"\t", col.names=TRUE, row.names = FALSE)

 #In text edit, remove " - find and replace with blank space

t5=read.table("MAadult_export_final.txt", skip=1,
colClasses=c("character", "numeric","numeric","numeric"))
t6=t5[with(t5,order(t5$V1, t5$V2)),]
 # To order the columns first by chromosome, then by location

write.table(t6, file = "MAadult_exportfinalfix.txt", sep =
"\t", col.names=TRUE, row.names = FALSE)
t7=as.data.frame(read.table ("MAadult_exportfinalfix.txt",
skip=1))
t7$V4[which(t7$V4<0.3)]<-0
 # To remove e-X values and change to 0

write.table(t7, file = "MA_final.txt", sep = "\t",
col.names=TRUE, row.names = FALSE)
 # Open file in text edit, search for ends of chromosome and
  manually edit falsely corrected values
 # Remove "v" first line and introduce wiggle descriptor
 # Manually check for exponential values in start/end columns
  and correct to whole numbers
 # Change file format to .wig
 # Upload to UCSC
```

A further script was required to fix an error in the Fru^{MC} pupae dataset. The Fru^{MC} pupae file originally provided by Nimblegen incorrectly had the same value for both the start and end probe values. To fix this the following script was used:

```
t=as.data.frame(read.table("MCpupae_finalfix.txt", skip=1))
t$V3<-t$V3+5
write.table(t, file="MCpupae_columnrepair.txt", sep="\t",
col.names=TRUE, row.names=FALSE)
 # Some probe values now overlap, manual editing required.
```

R was also used for creating heat maps to analyse the datasets. The datasets were amalgamated into one file using excel. Following this, the Ringo script was edited to fit requirements [133].

```
GO<-read.delim("genelist_heatmap.txt",header=TRUE)
 # Loads File
GO
names(GO)
GO[,-c(1)]
row.names(GO)<-GO$GENE
```

```
source("http://www.bioconductor.org/biocLite.R")
biocLite("ALL")
library("Ringo")
biocLite("limma")
biocLite("marray") # Loads Source Files

GO_matrix<-data.matrix(GO[,-c(1)])

library(limma)
library(gplots)
library(marray) # Loads Relevant Libraries

by <- maPalette(low="green", high="red", mid="black")
 # Identifies Colour Scheme
heatmap.2(GO_matrix,dend=c("both"),col=by,
margins=c(10,10),xlab="DATASET",ylab="GENE",
key=TRUE,keysize=1,symkey=FALSE,density.info="none",
trace="none",cexCol=0.5,cexRow=0.05)
 # Draws Heat map
```

2.6 Behavioural analysis

2.6.1 Complementation

Single male mutant flies were crossed to two virgin females of another mutant line at 25 °C. The F1 progeny were scored for phenotypes and compared to the expected genetic outcome. A decrease in homozygous flies showed non-complementation between the mutations.

$$\text{Complemenation} = \frac{\left(\frac{\text{Homozygotes}}{\text{Total flies per cross}}\right)}{\text{Expected number of homozygotes}} \text{ x } 100$$

2.6.2 Viability

Single male mutant flies were crossed to two female virgin females of the same mutant line at 25 °C. The number of F1 progeny were scored and compared to expected genetic outcomes:

$$\text{Viability} = \frac{\left(\frac{\text{Homozygotes}}{\text{Total flies per cross}}\right)}{\text{Expected number of homozygotes}} \text{ x } 100$$

2.6.3 Fertility

Mutant flies were crossed to flies of the opposite sex of the same genotype, deficiencies of *fru*, mutations in *fru* locus, or WT. The single homozygous offspring of these flies were crossed to two virgin WT flies and left for 10 days at 25 °C. If the mutation affects fertility, it should register as a failure to produce offspring when combined with deficiencies or mutations in the *fru* locus affecting the sex-specific transcripts of *fru*. Lines were scored as either fertile or infertile (whether progeny can be seen) and a percentage fertility calculated:

$$\text{Fertility} = \frac{\text{Number of fertile crosses}}{\text{Total number of crosses}} \text{ x } 100$$

Both males and females are tested for fertility issues to identify whether one or both sexes are affected.

2.6.4 Courtship behaviour

Male and female flies were separated immediately post-eclosion and aged under standard conditions (5- 7 days for males, 3- 5 days for females). Females were stored in small groups in vials; males were kept individually in vials. One naïve male fly was placed with a virgin female in a small chamber and recorded for a period of one hour. The following observations were made.

Courtship initiation:

The time in seconds it takes for the male to show three continuous seconds of courtship behaviour towards the female.

Time to copulation:

The time in minutes from initiation of courtship to copulation.

Copulation duration:

Time in minutes that the flies spend copulating.

Courtship index (CI):

Total time in a ten-minute period (starting at courtship initiation) that a male fly spends courting the female. If the fly copulates in less than ten minutes, the time recorded is multiplied up to be equivalent to ten minutes.

Wing extension index (WEI):

Total time in a ten-minute period (starting at courtship initiation) that a male fly spends extending its wing (45 – 90 °) towards the female. Time is recorded as a percentage of the total time observation period.

Line crossing:

Number of times a fly crosses the midpoint of the chamber in a ten-minute period.

3 Homologous Recombination

3.1 Introduction

We have learnt a lot about the role of the *fru*^{C} isoform through studies of the EMS-generated isoform-specific mutant *fru*$^{\Delta C}$ [2]. These studies have shown that certain aspects of *fru* function are isoform-specific, such as the formation of the MOL, which is controlled specifically by *fru*^{C}. Combining deficiency and mutant lines of the *fru* locus have allowed a further level of understanding, connecting the role of specific transcripts to particular promoters [78, 79, 82, 107, 108]. However, we are lacking the tools to properly characterise the functions and effects specific to the remaining *fru* isoforms: *fru*^{A} and *fru*^{B}.

Considering the existence of a specific deletion mutation for the *fru*^{C} isoform of *fru* (*fru*$^{\Delta C}$), why haven't other isoform's been generated by a similar method? In short, it was somewhat unlikely to achieve even one mutation by means of EMS mutagenesis – often EMS screening leads to larger mutations that wholly abolish or greatly reduce the effect of the gene in its entirety [34, 116, 117, 134, 135]. It is rare that a small specific mutation is made without further larger lesions (major deletions, truncations) or multiple small mutations rendering the resultant fly inviable [36]. Another means of creating a mutation at the *fru* locus has been X-ray induced chromosomal inversion, which was used to create the *fru*^{1} allele, that causes complete male sterility [78, 135-137]. Further mutants have been generated by *P*-element insertion, many of which broadly disrupt *fru* splicing, leading to mutants with hugely differing transcripts (e.g. *fru*^{sat} and *fru*^{4}) which show little courtship [86]. These mutants have been hugely informative, enabling the characterisation of many aspects of *fru* function, however the phenotypic defects identified cannot necessarily be attributed to specific isoforms. An

alternative mutagenic approach is therefore necessary to characterise the roles of additional individual *fru* isoforms.

Our knowledge and understanding of *fru* and its interactors would be greatly increased if we had access to mutants similar to $fru^{\Delta C}$ for both fru^{A} and fru^{B}. To achieve mutants like these there are a number of possibilities, including non-targeted mutation such as EMS used to create the $Fru^{\Delta C}$ mutant. However, a targeted approach is the most efficient method. This can be achieved through ends-in HR, a targeted approach designed to incorporate small specific mutations, ideal for introducing null mutations. The resultant mutants would increase our understanding of *fru*, including enabling further insight into the downstream targets of specific Fru proteins.

3.2 Aims and objectives

To create fru^{A} and fru^{B} isoform mutants using HR, by inserting small nonsense mutations in the *fru* A and B exons, rendering these exons non-functional and if possible, create a $fru^{\Delta A/ B}$ isoform double mutant. Further, to characterise the changes of Fru expression in these mutants using antibody staining, to log their effects on viability, fertility and basic courtship behaviour.

3.3 Understanding *wild-type* Fru expression

To be able to successfully characterise mutant flies, it is important to have a clear understanding of the wild-type phenotype. Fru expression is very specific and distinct, being expressed in approximately 2,000 neurons in the adult fly CNS, this is in the region of 2 % of the total CNS [81]. Its expression changes clearly throughout development, both in the number of cells in which it is expressed and

the intensity of that expression [81]. It was prudent to confirm this expression at the developmental time points considered in this body of work, and further confirm the accuracy and quality of the Fru isoform-specific antibodies available (J. Walker, unpublished, [80]).

3.3.1 Antibody stainings identify expression of Fru^M in wild-type CNS

The brain and VNC were dissected as per 2.1.5 at three distinct time points: L3 'wandering' larvae (onset of Fru expression); 46 hr. pupae (peak of Fru expression); and 48 hr. adults (the fully formed adult fly, at which point Fru is no longer required for developmental purposes, and potentially performs a physiological role).

Following standard Fru staining protocols (2.2.1), CNS preparations were co-stained with Fru^M antibody [2], and one of three isoform specific antibodies: Fru^A [80], Fru^B or Fru^C (J. Walker, unpublished). The Fru^B and Fru^C antibodies have not been affinity-purified, and so may potentially have higher levels of background signal (J. Walker, unpublished) – however for the requirements of this work, a reasonable Fru^B antibody would be helpful for characterising any potential mutants. The Fru^A isoform antibody had previously been shown to work and should be useful for characterising any Fru^A mutants derived by HR. These antibodies will confirm successful HR mutants by showing a loss of Fru expression, compared to wild-type expression.

3.3.1.1 The larval time point

The larval time point is known to show only a small amount of Fru expression, from the L3 'wandering' time point onwards [81]. Most notable at this time point

was non-specific staining seen in both the male and female VNC, Figure 3-1. This very distinct patterning down the lateral edges of the VNC is slightly below the top layer of cells, and shows potential segmental patterning [138]. The patterning is most likely non-specific, as it is non-sex-specific, and is seen only with this one Fru^M antibody preparation (compared to a previously generated Fru^M mouse antibody), suggesting it is an artefact of this specific antibody.

Sex-specific Fru staining in the larval time point is seen solely in the brain, and appears to be inconsistent between samples. These observations mirror those of Lee et al. in 2000 who, through in-situ hybridisation, characterised between six and eight clusters in the brain, and one in the VNC [81]. This inconsistency is indicative of the quickly changing morphology at this developmental stage, which is superseded by more clearly defined cell clusters later in development (from the pupae onward). In the samples shown (and in those considered) no cluster was seen in the VNC using Fru^M.

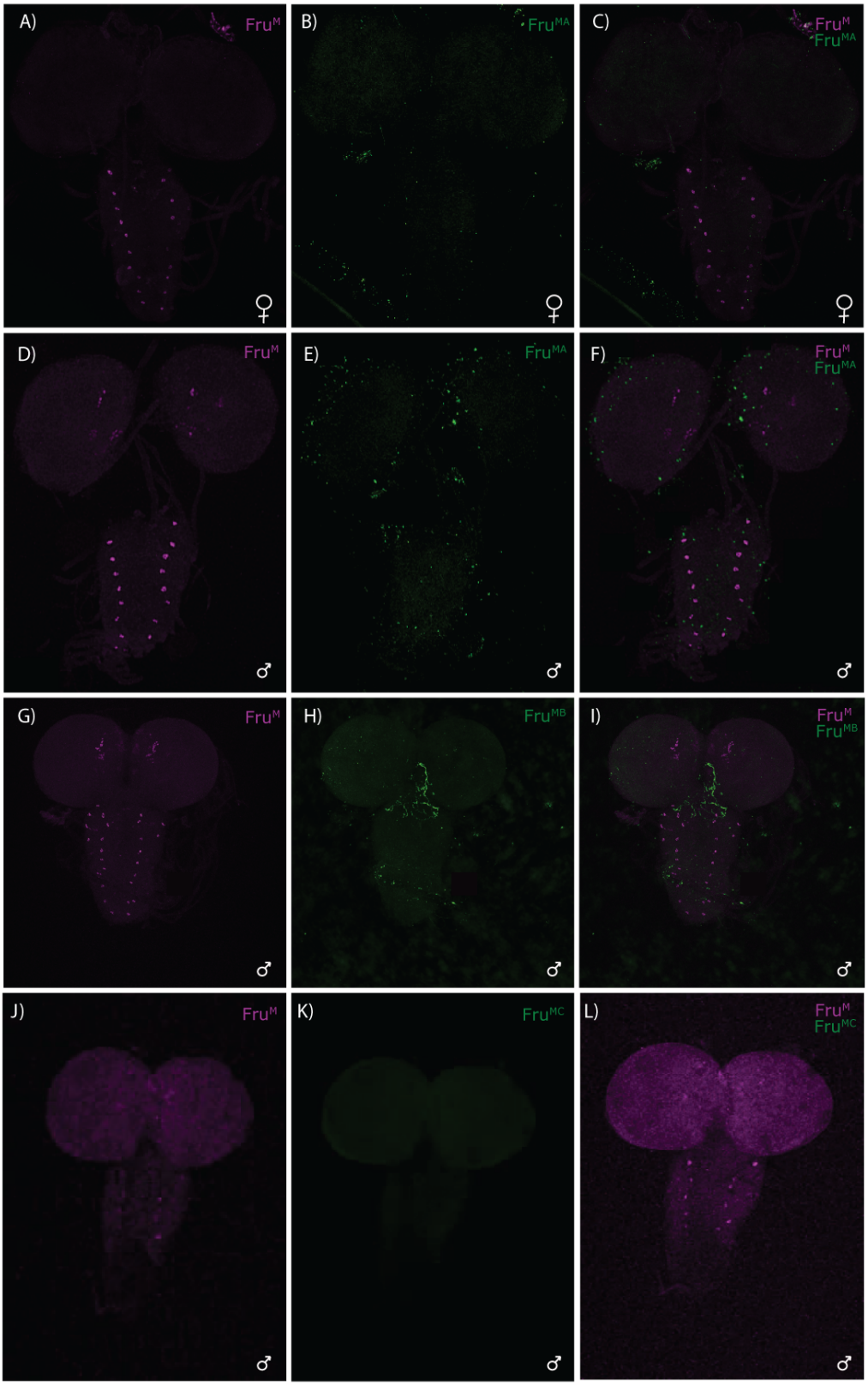


Figure 3-1 Fru^M expression in wild-type larval CNS. Fru^M expression has been documented as starting at L3 'wandering' larvae. (A-L) show WT L3 'wandering' larvae stained with Fru^M antibody (magenta) and isoform specific antibodies (green). (A-C) Female L3 larvae stained with Fru^B antibody and Fru^M antibody. Other isoform specific antibodies were used, but not shown. (A) Fru^M antibody, showing previously unseen staining pattern. (B) Fru^B antibody showing no staining. (C) Fru^M and Fru^B merge showing no overlap of staining. (D-L) Male L3 larvae stained with Fru^M and either Fru^A antibody (E-F), Fru^B antibody (H-I) or Fru^C antibody (K-L). In each images F, I and L, Fru^M antibody staining is merged with that of the isoform specific antibody to observe overlap between the two. All images are maximal Z stacks of approximately one-µm slices, anterior view.

At this time point, the isoform specific antibodies showed no staining. This seems to be at odds with the results for Fru^{M} (a combination of the staining of each of the individual isoforms), and may be explained in two ways. Either, the isoform specific antibodies are too insensitive (having not been affinity purified), and are incapable of either staining low levels of expression or identifying occasional cells above the background signal. All isoforms could potentially contribute very low-level expression, which individually are not detected, but collectively produce a detectable signal with the Fru^{M} antibody (as it detects all isoforms). Given the small clustering seen with the Fru^{M} antibody along with the Lee et al. *in situ* images, either of these possibilities may be the case, however given later data in this body of work, it seems more likely that the second scenario, where too little expression is present to be detected by the individual isoform specific antibodies, is more likely [81].

3.3.1.2 The pupae time point

The pupae time point appears to be the point at which the non-specific staining of Fru^{M} dissipates. The female brain in Figure 3-2A-C is slightly early then the defined time point and still shows some non-specific staining, whereas by 46 hr. proper, the non-specific staining is lost.

Previously, is has been suggested that isoform Fru^{MA} is expressed in fewer cells compared to isoforms Fru^{MB} and Fru^{MC} [2]. This study shows that the staining is not very strong, but that – at least in the brain – the staining is similar to that of Fru^{M}. Clusters eight and nine (Figure 3-3), as defined by Lee et al., are difficult to see against the background due to their wide spacing. In the VNC there may be less staining, particularly in the abdominal ganglion, however the high

background levels make it difficult to definitively identify overlap between these cells.

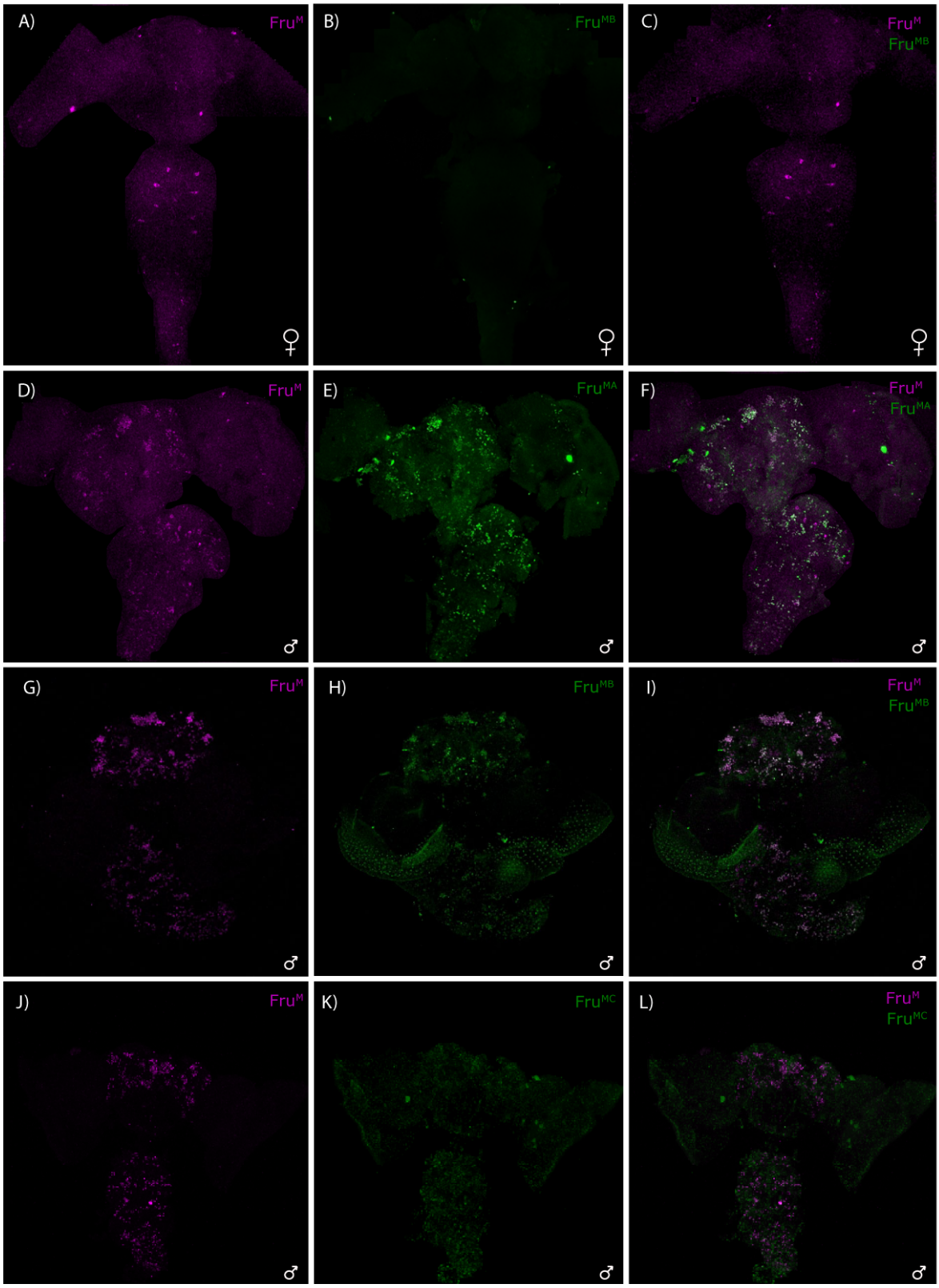


Figure 3-2 Fruitless expression of the pupal wild-type. *fruitless* expression has been documented as reaching peak expression at 46 hr. pupae. (A-L) show WT 46 hr. pupae stained with Fru^M antibody (magenta) and isoform specific antibodies (green) to document this staining. (A-C) Female 46 hr. pupae stained with Fru^B antibody and Fru^M antibody. (A) Fru^M antibody showing no Fru specific staining, but some non-specific staining seen in the larval time point is retained. (B)

Fru^B antibody showing no staining. (C) Fru^M and Fru^B merge showing no overlap of staining. (D-L) Male 46 hr. pupae stained with Fru^M and either Fru^A antibody (E-F), Fru^B antibody (H-I) or Fru^C antibody (K-L). In each of the images F, I and L, Fru^M antibody staining is merged with that of the isoform specific antibody to observe overlap between the two. All images are maximal Z stacks of approximately one-µm slices, anterior view.

Isoform's Fru^{MB} and Fru^{MC} both appear to be congruous to Fru^M, showing strong staining in all 20 clusters associated with Fru^M expression [5, 81]. Being the peak of Fru expression, this ought to be where the clearest images can be achieved, and indeed, good images of isoform's Fru^{MB} and Fru^{MC} were achieved compared to other time points.

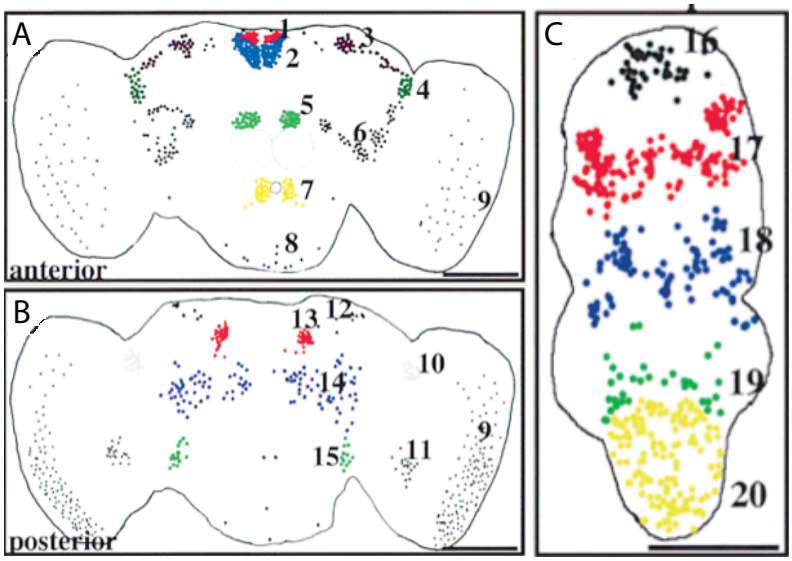


Figure 3-3 Schematic of the location of Fru expressing cells in the CNS of *Drosophila melanogaster*. Originially produced by Lee at al., 2000, Fru^M immunoreactive neurons are shown in the male pupae CNS – the peak of Fru expression. (A) The anterior of the brain showing clusters 1-9 as defined by Lee et al., (B) The posterior of the brain showing clusters 9-15 as defined by Lee et al., (C) The VNC showing clusters 16-20 as defined by Lee et al.. Scale bars = 100 µm [81].

3.3.1.3 The adult time point

By the adult time point, the expression of Fru should represent the mature and fully developed state. Although Fru^M expression peaks in the pupal stage, a large amount of expression carries on into adulthood, with all clusters defined by Lee *et al.* being visible at this time point [81]. The expression of the Fru^{MB} and Fru^{MC} isoforms appears to be very similar, with staining present in all the clusters,

suggesting both isoforms were expressed in most or all Fru expressing cells at this time point. The Fru^{MC} staining for this time point was less clear, due to high levels of background, but similarly appeared to be in all Fru expressing cells

By comparison, Isoform Fru^{MA} appeared to be expressed in fewer cells in the adult time point, following a similar pattern to that seen in the pupae. Again, the brain appeared to express similarly to the other isoforms, however in the VNC, and particularly in the abdominal ganglion there were clearly Fru^{M} expressing cells, which did not express Fru^{MA}. This pattern was also observed by *Billeter et al.,* in 2006, and alludes to a potentially different role for the Fru^{MA} isoform [2].

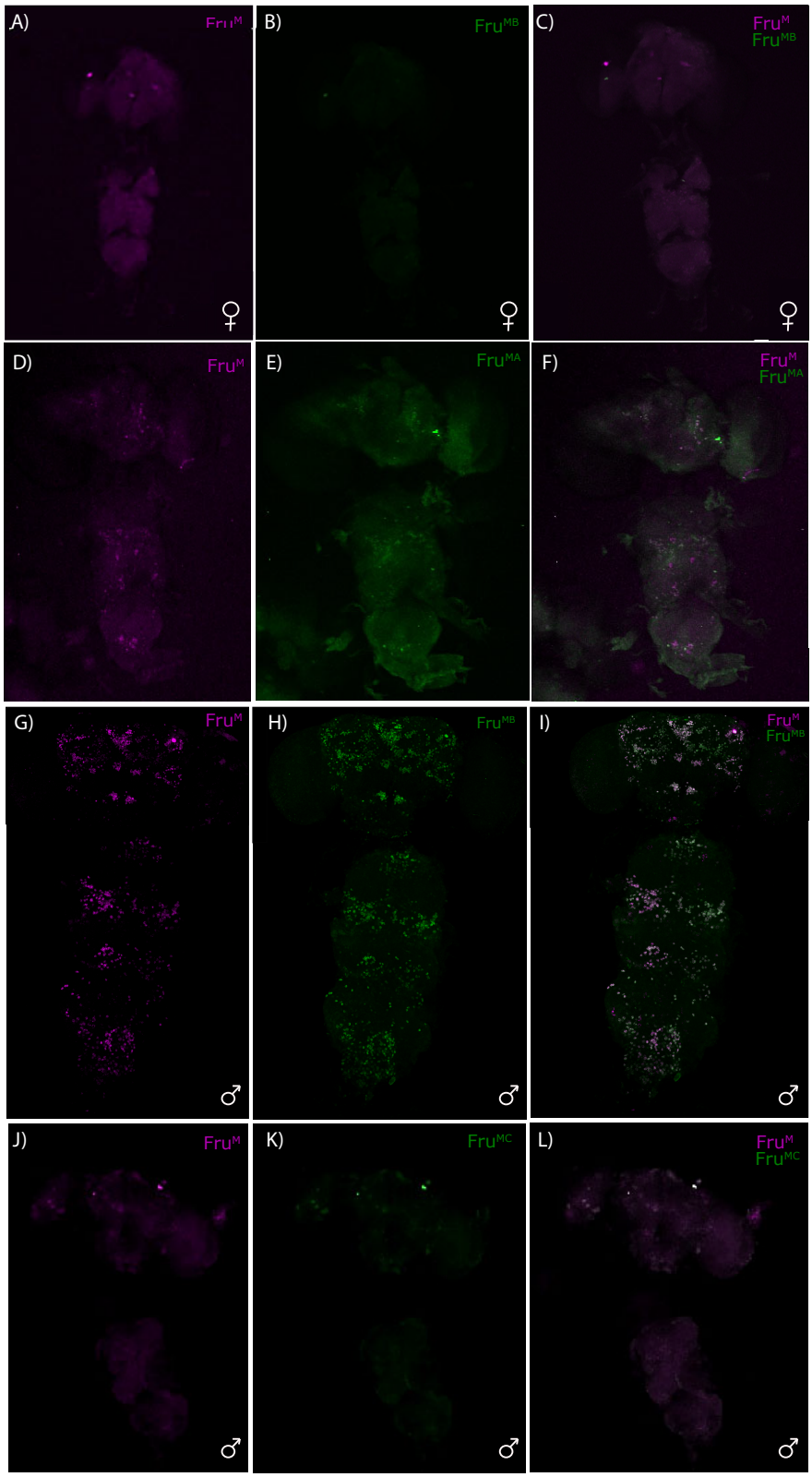


Figure 3-4 Fruitless expression1 of the adult wild-type. *fruitless* expression has been documented as reaching peak expression at 46 hr. pupae, but is maintained through out the adult time point. (A-L) show WT 48 hr. adult stained with Fru^M antibody (magenta) and isoform specific antibodies (green) to document this staining. (A-C) Female 48 hr. adult stained with Fru^B antibody and Fru^M antibody. (A) Fru^M antibody showing no Fru specific staining, but some non-specific staining seen in the larval time point is retained. (B) Fru^B antibody showing no staining. (C Fru^M and Fru^B merge showing no overlap of staining. (D-L) Male 48 hr. adult stained with Fru^M and either Fru^A specific antibody (E-F), Fru^B antibody (H-I) or Fru^C antibody (K-L). In each of the images F, I and L, Fru^M antibody staining is merged with that of the isoform specific antibody to observe overlap between the two. All images are maximal Z stacks of approximately one-µm slices, anterior view.

3.4 Mutating the *fru* locus

Our knowledge of *fru* today stems largely from studies of mutant phenotypes [2, 78, 79]. However, as most of the mutants have arisen from non-targeted mutagenesis, they largely abolish the gene in its entirety, restricting their contribution to the understanding of the isoforms of *fru*. In fact, as discussed in (1.5.1) only the fru^{C} isoform has specific functional attributes identifiable to it, due to work completed using $fru^{\Delta C}$ [5].

HR is a powerful targeting tool for generating targeted genomic mutants, which was first used in Yeast [116], before being modified to function in *Drosophila*. This approach can be used to address the lack of isoform specific mutants in the *fru* locus [139]. Uncontrollable structures within the genome vary the efficiency of this method, such as chromatin modelling [139], and yet previous targeting of the *fru* locus via this method has been successful (J. Walker, unpublished, [83, 124]).

3.4.1 Previous attempts to create null mutations of fru isoforms fru^{A} and fru^{B}

Previously work in the lab undertaken by Dr. John Walker exploited a technique pioneered by Dolezal et al., in 2003 whereby two mutations can be inserted simultaneously using HR [140]. The proximity of the fru^{A} and fru^{B} exons in *fru* allowed both isoforms to be targeted for mutagenesis together, and more interestingly – allowed for the potential to create a double mutant. Dr. Walker initially modified a targeting vector, *p*ED7, adding mutated *fru* sequence and an I-SceI site into the MCS to use in the HR process, Figure 3-5 [127, 140]. The vector comprises transposition sequences at the 5′ and 3′ ends to facilitate integration into the target genome of the donor construct. Within these elements are

directional FRT sites (which allow FLP recombinase to excise the donor DNA as an extrachromosomal circle) and the donor DNA itself (including the two mutated exons with engineered restriction sites for confirming integration of the mutations, separated by an I-SceI recognition site). 3′ to the I-*Cre*I restriction enzyme site, the *mini-white* marker gene is included, allowing the donor DNA to be followed by eye colour during the subsequent genetic crosses. The donor DNA was designed carefully, using strict parameters to increase the likelihood of successful targeting, and thus successful subsequent mutation incorporation. A minimum of 2 kb has been shown to be required for efficient targeting [127, 141, 142]; The position of the I-SceI site is also crucial, as resolution of a DSB can affect up to 1.3 kb on either side of it [143]. Thus, the I-SceI site is equidistant from both mutations, and there is 4 kb homology flanking DNA, creating a fragment of 12 kb that was inserted into the *p*ED7 vector, creating *p*ED7-*fru*$^{A/B}$ (Figure 3-5), which increases the likelihood of equal frequencies of recombination events for each mutation. This work proved to be unsuccessful due to a design flaw in the I-*Cre*I restriction enzyme site in the original *p*ED7 donor vector. The I-*Cre*I site is critical for the final step of the HR process, when the duplication incurred by the ends-in method is resolved, therefore in this instance the duplicated *fru* locus could not be resolved.

By comparing the sequence of the I-*Cre*I site in the *p*ED7 construct to those found in a previously published ends-in HR construct, we found that additional sequences were included on either side of the I-*Cre*I site (see Figure 3-6). We therefore designed primers that were annealed to incorporate an additional functional I-*Cre*I site into a unique *BamHI* site, coincidentally available adjacent to the non-functional site. In fact, three I-*Cre*I sites were added at this single *BamHI* site in the *p*ED7-*fru*$^{A/B}$ vector. These multiple copies however do not interfere with

the resolution process and thus a functional construct was successfully engineered.

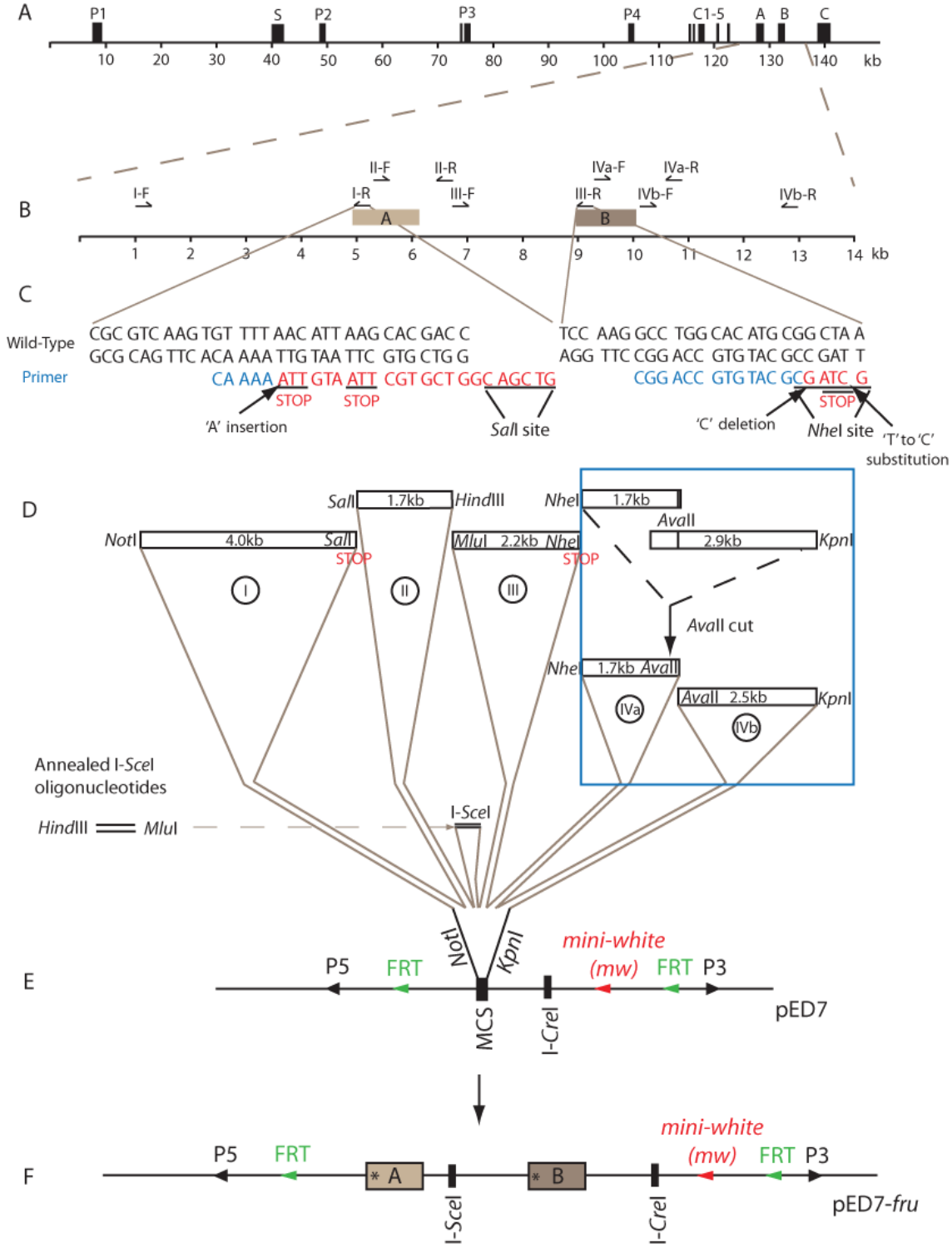


Figure 3-5 Schematic of donor construct for gene targeting of *fru* locus. (A) Entire *fru* locus. Promoters and exons are represented by black boxes. Size of locus is indicated in kb. (B) Enlarged view of target locus showing PCR primer positions, forward (F) and reverse (R) indicated by half arrows. Numbers on arrows indicate the fragment (I-IV) amplified by a particular primer set (fragment IV being a composite of IVa and IVb). Exons A and B are represented by beige boxes. (C) Nucleotide sequence of primers used to introduce point mutations into coding sequences of exons A and B, compared to wild-type sequence. Induced point mutations and resultant stop codons are indicated as are the novel diagnostic restriction sites engineered into the primers. (D) Sizes of PCR fragments used to assemble the donor construct. Area inside the blue box represents

the amalgamation of fragments IVa and IVb. Restriction sites used to clong PCR framgements and annealed oligonucleotides into *p*ED7 vector. (E) Restriction sites of *p*ED7 vector used to clone donor construct and relative positions of elements within the vector. (F) Final assembled gene targeting constuct, *p*ED7-*fru*$^{A/B}$. Introduced exon mutations are represented by an asterix (*). (J. Walker, unpublished).

```
Original I-CreI AAAACGTCGTGAGACAGTTT
                ....................
Additonal I-CreI  ACCAAAACGTCGTGAGACAGTTTGCC
```

Figure 3-6 I-CreI site. The original I-CreI site found in the *p*ED-7 construct was incorrect. The site was repaired by the insertion of the correct sequence at a convenient unique *BamHI* site.

3.4.2 Ends-in homologous recombination

There are two types of HR that have been shown to work in *Drosophila*, these are referred to as ends-in and ends-out (Figures 1-6 and 1-7) [84]. Ends-in HR has been shown to be more efficient in yeast and mammalian cells [117, 144, 145], while Bellaiche et al. were initially unsuccessful with the ends-out approach in *Drosophila* [117]. The ends-out approach has now been shown to be successful, showing similar levels of efficiency as the ends-in approach in *Drosophila* [117], however features of the ends-in approach make it more beneficial for precise manipulations of the genome [84].

Ends-out HR replaces can be used to generate truncated copy of a gene, entire deletions, and even specific mutations, however a “footprint” of the donor vector is always left behind, which could have further exogenous repercussions, whereas the ends-in HR method can introduce accurate mutations which can modify the gene product leaving no changes to the morphology of the locus [127]. Additionally, Ends-in HR has been successfully used to modify two genes via the same donor DNA construct, as a result of the nature of the duplication resolution during the I-*Cre*I step. This step also removes the HR “footprints” left *in-situ* with the ends-out approach [140]. This was a crucial requirement in modifying the *fru* locus as the aim was to generate novel mutants that targeted

individual isoforms, leaving the rest of the locus unchanged, thus ends-in was the ideal method.

3.5 Targeting the *fru* locus

The *p*ED7-*fru*$^{A/\ B}$ construct was microinjected into w^- fly embryos (Genetic Services Inc., Cambridge, MA). Adult flies were mated to w^{1118} line and transformants identified from their progeny by eye colour (red) conferred by the *mW*+ marker in the targeting construct, Figure 3-8. Only one independent transformant lines was isolated, and a series of crosses to balancer chromosome lines was used to determine the chromosome location of the insert. This line had a strong red eye colour and the *p*ED7-*fru*$^{A/\ B}$ insertion was localised to the second chromosome. The second chromosomal location was fortuitous as we were targeting the *fru* gene on the third chromosome. This allows the easy distinction in later steps between targeted insertions into the *fru* locus, versus the original *p*ED7-*fru*$^{A/\ B}$ insertion. This *p*ED7-*fru*$^{A/\ B}$ line was thus used to integrate the mutations into the *fru* locus.

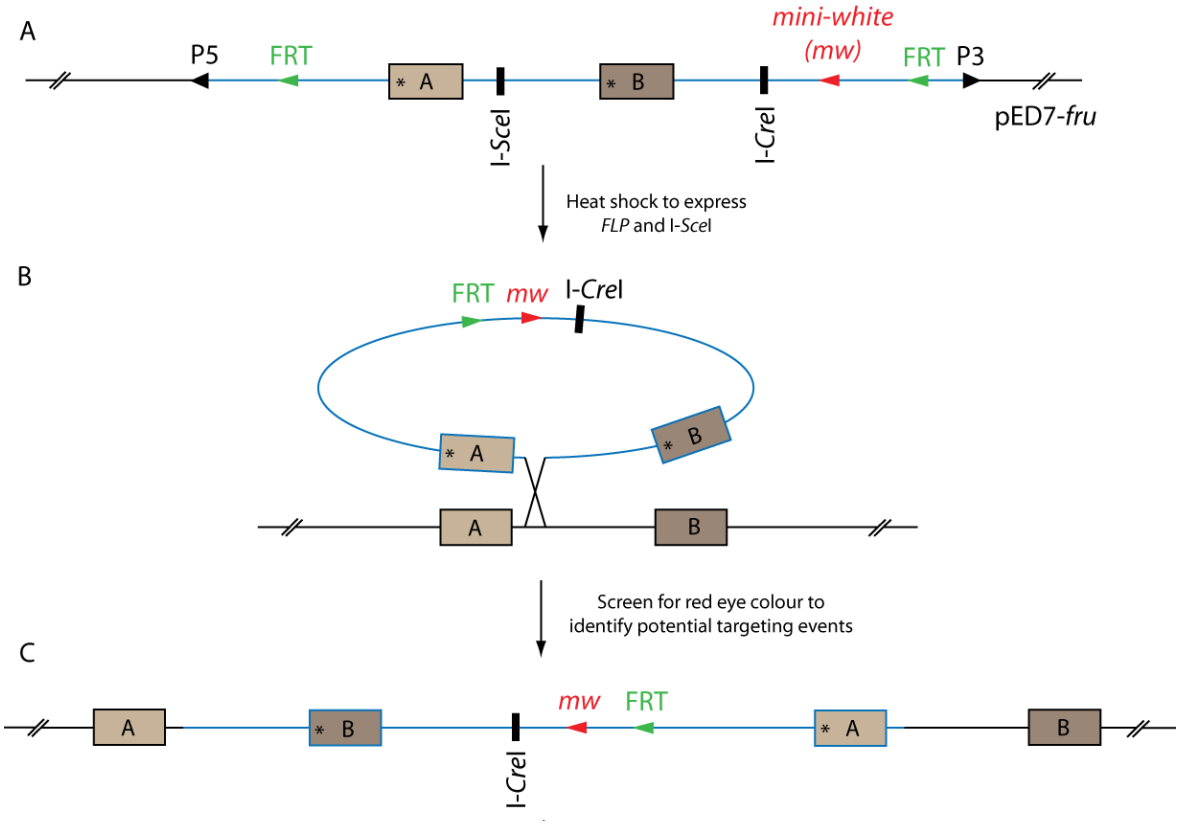


Figure 3-7 Molecular representation of gene targeting process at the *fru* locus. (A) Donor construct integrated at random location on second chromosome. Endogenous DNA is represented by a black line, DNA of targeting construct shown by blue line. Arrowheads indicate components of targeting construct; *P*-element ends (black), FRT sequences (green) and *mW*+ marker gene (red). I-*Sce*I and I-CreI endonuclease sites are represented by thick black lines. (B) Excision and circularisation of donor construct through the action of FLP recombinase. Followed by induction of DSB by I-*Sce*I endonuclease and insertion of donor construct. (C) Red-eyed re-integrants are isolated and mapped to chromosome of insertion.

3.5.1 Integration at the target locus

The first step was to cross virgin *p*ED7-*fru*$^{A/\ B}$ containing females to males from the M690 line (see Cross 1). The M690 line carries heat shock hid (*hs-hid)* on the Y chromosome, and heat shock FLP recombinase (hs-FLP) and I-SceI endonuclease on the second chromosome. *hid* is a cell-death regulator on the Y chromosome, under the control of a heat shock inducible promoter which should be inactive at temperatures under 25 °C, however was propagated and maintained at 18 °C [146]. When raised to 37 °C *hid* was induced in males, cell death occurred preventing male progeny, thus enabling easy large-scale virgin female collections that were necessary for HR. The second chromosome contained the heat-inducible FLP (*hs-FLP),* which allows the ends-in construct to

be excised as a circular molecule, and heat-inducible *I-SceI* restriction enzyme *(hs-I-SceI*), which creates a double-stranded DNA break. The resulting "free-arms" of the targeting construct are thus able to target regions of homology in the genome. The adults of Cross 1 (150 females of pED7-*fru*$^{A/B}$, 50 males of M690 per bottle) were maintained at 25 °C for 2 days, upon which time they were transferred to a new bottle and the original bottle was heat shocked at 37 °C for 1½ hrs., followed by 45 mins of heat shock the following day. Ten bottles were originally set-up, and turned ten times each. Approximately 1500 females were used in total for these crosses.

$$\frac{w}{w};\frac{\mathrm{p}[>w^{+}>]}{\mathrm{p}[>w^{+}>]};\frac{+}{+} \times \frac{y,w}{w,hs-hid};\frac{\mathrm{Sco},hs-I-SceI,hs-\mathrm{FLP}}{\mathrm{CyO}};\frac{+}{+}$$

Cross 1 *p*ED7-*fru*$^{A/B}$ donor line crossed to M690. *PED7-fru*$^{A/B}$ is a homozygous donor line on the second chromosome. M690 contains *hs-hid* on the Y chromosome, which kills males at 25 ˚C allowing females to be easily selected. The second chromosome contains *hs-I-SceI* and *hs-FLP* allowing the creation of a DSB and targeting of the donor DNA to the homologous genomic locus.

Virgin female progeny from Cross 1 were collected with either solidly white eyes (indicative of complete loss of the targeting construct from the somatic cells of the eye) or red-white mosaic eyes (indicative of partial loss of the targeting construct in the soma). These females were used for Cross 2, where 200 female progeny from Cross 1 and 100 males of the M699 genotype were crossed in bottles. After three days the flies were turned into a fresh bottle. This cross was turned as necessary until enough progeny were collected. The progeny required from this cross had red eyes, indicating the integration of the targeting construct. This occurred in the gametes of the females heat shocked in the previous cross, and was rare in occurrence. In total 41 flies were identified with red eyes.

$$\frac{w}{y,w};\frac{\mathrm{p}[>w^{+}>]}{\mathrm{Sco},hs-I-SceI,hs-\mathrm{FLP}};\frac{+}{+}\ \mathrm{x}\ \frac{y,w}{y};\frac{ey-\mathrm{FLP}}{ey-\mathrm{FLP}};\frac{+}{+}$$

Cross 2 Progeny from Cross 1 crossed to ey-FLP line. The virgin female progeny of Cross 1 contain the target vector, as well as *hs-I-SceI* and *hs-FLP* on one chromosome, giving the ability to create a DSB and using FLP recombinase, jump in the vector DNA to the genome at the homologous site, under the influence of heat shock therapy. The line to which it is crossed, M699, contains two copies of ey-FLP to drive this process.

Each red eyed fly obtained from the second cross, was crossed to M700 and M701 lines to identify the location of integration, Cross 3A and 3B, from which those found to be on the third chromosome (location of *fru*) were taken on to be resolved.

$$\frac{y,w,ey-\mathrm{FLP}}{y,w,ey-\mathrm{FLP}};\frac{\mathrm{Pin}}{\mathrm{CyO}};\frac{+}{+}\ \mathrm{x}\ \frac{(y),w}{\mathrm{Y}};\frac{\mathrm{ey}-\mathrm{FLP}}{?};\frac{>w^{+}}{+}$$

Cross 3A Identification of the location of the insertion. Using Balancer Chromosomes, the mutated line was crossed to balancer lines sequentially: confirming locations on the second and third chromosomes. Line M700 confirms the location on the second chromosome, by crossing the mutant line to a second chromosome balancer, containing *Pin* and *Curly of Oyster (CyO)* markers. A positive identification of the location of the chromosome is conferred as a solid red eye.

$$\frac{\mathrm{y,w,ey}-\mathrm{FLP}}{\mathrm{y,w,ey}-\mathrm{FLP}};\frac{\mathrm{Ly}}{\mathrm{TM3,sb}};\frac{+}{+}\ \mathrm{x}\ \frac{(\mathrm{y}),\mathrm{w}}{\mathrm{Y}};\frac{\mathrm{ey}-\mathrm{FLP}}{?};\frac{>\mathrm{w}^{+}}{+}$$

Cross 3B Identification of the location of the insertion. Using Balancer Chromosomes, the mutated line was crossed to balancer lines sequentially: confirming locations on the second and third chromosomes. Line M701 confirms the location on the third chromosome, by crossing the mutant line to a third chromosome balancer, containing *Lyra (Ly)*, and *Stubble (Sb)* markers. A positive identification of the location of the chromosome is conferred as a solid red eye.

3.5.2 Analysis of initial integration events

To identify successful re-integration of the donor DNA, verification by PCR was employed prior to resolution. Primers designed on either side of the integrated FRT site were used to confirm re-integration. If the FLP recombinase had successfully recombined the FRT sites, joining them together, the primers would bind and produce a 600 bp fragment, Figure 3-8.

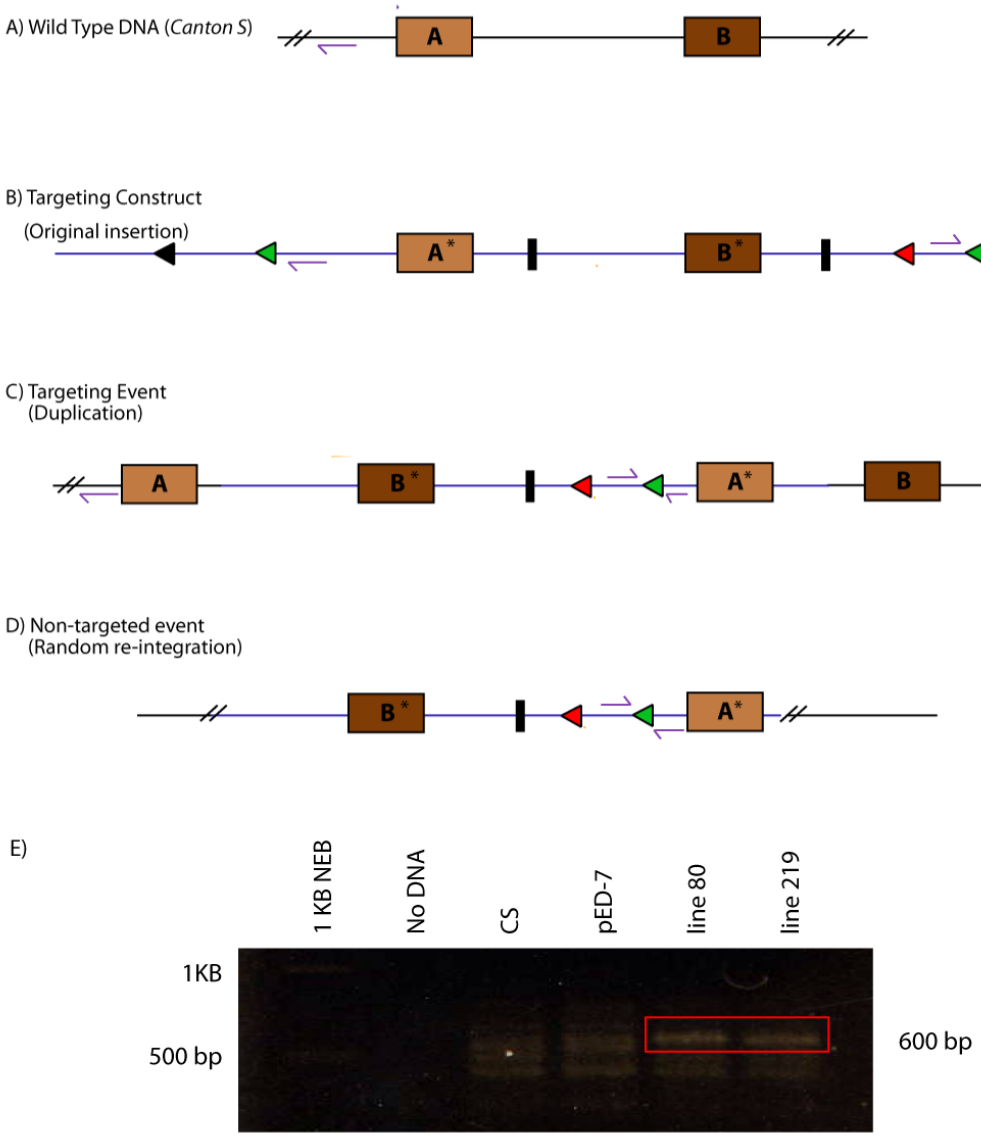


Figure 3-8 Verification of integration of the donor DNA. To confirm the re-integration of the donor DNA into the targeted location on the genome, primers designed around the FRT sites (green arrow) were used to confirm integration. The red arrow represents the *mW+* gene. Primer sites are shown with a purple arrow. (A) WT shows a single primer location upstream of A. (B) Targeting construct showing the same primer location upstream of A, as well as another upstream of the FRT site. (C) Targeted event – by excising the donor DNA as an extrachromosonal circle, the FRT sites come together allowing a primer product to be amplified. (D) Non-targeted event – also shows a primer product able to be amplified, as excision has occurred and so the FRT sites have been brought together. (E) Gel showing primer products after amplification. No product is seen for either CS or the original targeting construct (*p*ED-7). 600 bp primer product is seen in the lanes for $Fru^{80.1}$ and Fru^{219} showing both have the donor DNA integrated into the genome (red box).

Only two were found to be located on the third chromosome. Both of these lines proved to contain re-integrated material (seen as a 600 bp fragment), Figure 3-8. The controls were CS (which had no integration event and so showed no fragment) and the original targeting construct which contains the donor construct, but has not been heat shocked (and so primer sites are not adjacent) also shows

no fragment. Therefore, in the case of the two lines which both show an amplified 600 bp product, the donor DNA had been excised from the donor plasmid as an extrachromosomal circle and re-integrated with a single recombined FRT site. These lines were designated $Fru^{80.1}$ and Fru^{219} and were both taken on to be resolved, with the intention of creating mutants: $fru^{\Delta A}$, $fru^{\Delta B}$, and if possible a $fru^{\Delta A, B}$ combined mutant.

3.5.3 Resolution of the duplication

Ends-In HR causes a duplication through the integration of the donor DNA, Figure 3-7C [84]. To remove this duplication a further series of crosses were initiated, starting with a heat-shock step to initiate expression of the I-*Cre*I endonuclease which induces a DSB, Resolution Step 1, Figure 3-6 and 3-8. Flies were crossed according to Resolution Step 1, and heat shocked at 35 °C for 1 hour on day three after crossing, and again at the same temperature for 45 mins the following day:

$$w^{118};;\frac{\mathrm{P\{v^{+}701.CreI\}}\,Sb}{\mathrm{TM6},Ubx}\ \mathrm{x}\ \frac{fru^{HR^{*}}}{\mathrm{TM3}}$$

* Where fru^{HR} represents one of the four following options:

y^{w}eyFLP2;+;$\frac{<w^{+}80.1;}{\mathrm{TM3}}$

y^{w}eyFLP2;+;$\frac{<w^{+}80.1;}{\mathrm{TM3}}$

y^{w}eyFLP2;+;$\frac{<w^{+}219.1;}{\mathrm{TM3}}$

y^{w}eyFLP2;+;$\frac{<w^{+}219.1;}{\mathrm{TM3}}$

Resolution Step 1 *I-CreI* endonuclease present in the vector in the resolution line is used to induce a DSB, which allows for the possibility that the duplication is resolved.

The resultant flies were either a double balancer (TM3, *Sb*, *e*/ TM6, Ubx, *e*) exhibiting the *ebony* phenotype; the TM3, *Sb*, *e* balancer over the *I-CreI* marked

with *Stubble* (lethal); *fru*^{HR} balanced with TM6, Ubx, *e* (*I-CreI* is not present, red eyes); or *fru*^{HR} over *I-CreI* marked with *Sb* which exhibit mosaic red/white eyes. It is these mosaic eyed flies which have the potential to resolve the duplication and so are separated at eclosion and crossed to virgin third chromosome balancer and their offspring screened for white eyes (showing the loss of the *mW+* and potentially the duplication). Resultant white eyes offspring were re-crossed to a third chromosome balancer, whose progeny were used to establish a balanced stock.

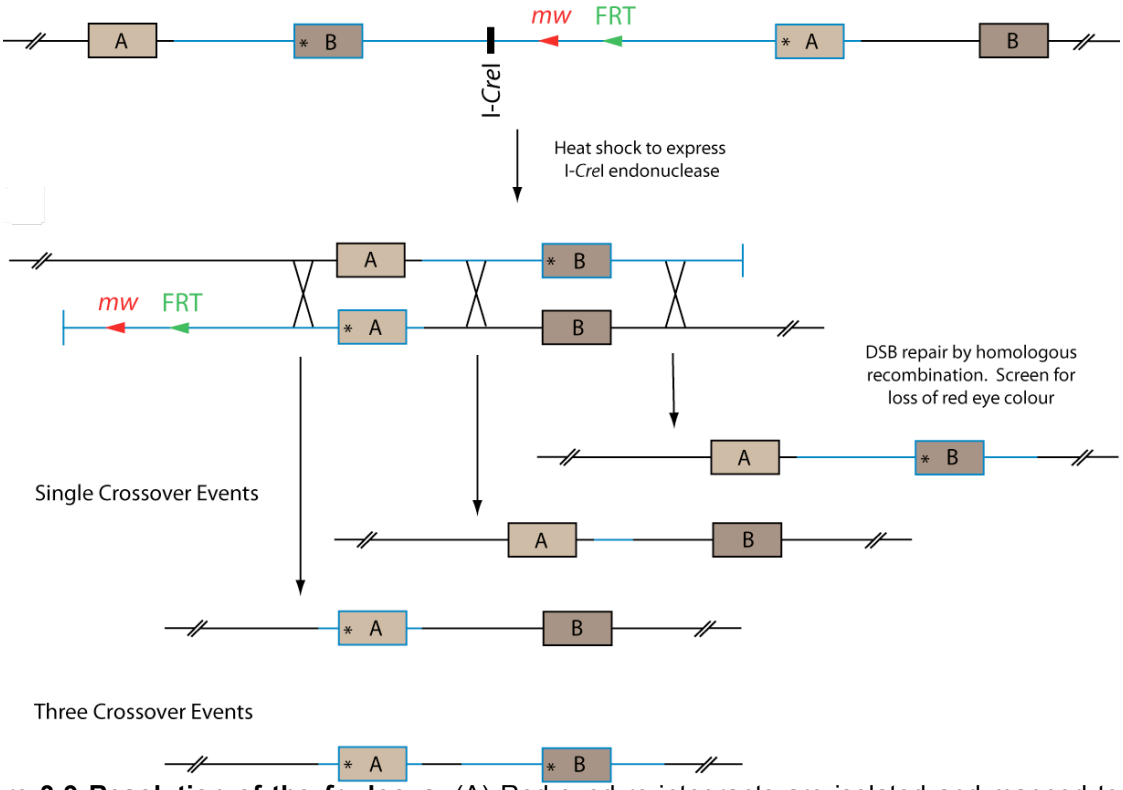


Figure 3-9 Resolution of the *fru* locus. (A) Red-eyed re-integrants are isolated and mapped to chromosome of insertion. (B) Formation of DSB by I-CreI endonuclease, inducing resolution of targeted duplication by homologous recombination. (C) Single crossover events potentially resulting in a single mutated exon. (D) Triple crossover events resulting in a double mutation of both the A and B exons. (J. Walker, unpublished). Arrowheads indicate components of targeting construct: FRT sequence (green) and *mW+* marker gene (red). I-CreI endonuclease sites are represented by thick black lines.

There are several potential outcomes from this resolution step (See Figure 3-8): one of three single crossover events or the more rare triple crossover event. In the event of a single crossover (Figure 3-8C) the resulting line would contain one of

three potential outcomes: a mutation in the B exon of *fru*; a WT copy of *fru*; a mutation in the A exon of *fru.* In the rare event of a triple crossover event, the genetic background of the resultant line would be mutant for both the A exon and the B exon of *fru*. To verify the mutants, and identify their incorporated mutation ($fru^{\Delta A}$ or $fru^{\Delta B}$, or $fru^{\Delta A,B}$), genomic DNA was made from each line, followed by PCR and restriction digestion analysis.

Both HR capable lines isolated ($Fru^{80.1/\ 219}$) were resolved producing 30 white-eyed lines with the potential to be mutants. To characterise their genotypes, DNA from each line was PCR amplified and cut with restriction digests. Each HR mutation was designed to incorporate a unique restriction site (*SalI* for exon A, and *NheI* for exon B), see Figure 3-5C. By digesting DNA from each line with either *SalI* or *NheI*, lines containing mutations of either exon can be identified, Figure 3-10. Fru^{219} was found to be capable of producing only $Fru^{\Delta A}$ mutants. From the lines resolved, five were found to be mutant in the A exon (indicated by the 594 bp and 682 bp fragments seen in Figure 3-10F). Of these five lines, four were found to be homozygous lethal and only one, Fru^{219F}, was found to be homozygous viable, Table 3-1. Upon further analysis it was found that the parent line Fru^{219} did not contain the mutation in the B exon of *fru*, suggesting that the original donor DNA was incompletely inserted during HR. Without sequencing, it is unclear how much of the *fru* gene is present, and what aspects of *fru* are associated with the A exon mutation and which are due to this incomplete targeting process. This was confirmed with basic characterisation of the resolved lines derived from Fru^{219} which were dis-similar to each other (four being inviable, and one being viable). Further, the viable line, Fru^{219F}, lost its viability through a number of generations suggesting this effect may not have been associated with *fru* itself. For these

reasons, these lines were discarded as being potentially non-representative of *fru* and Fru^{219} was discarded for future use.

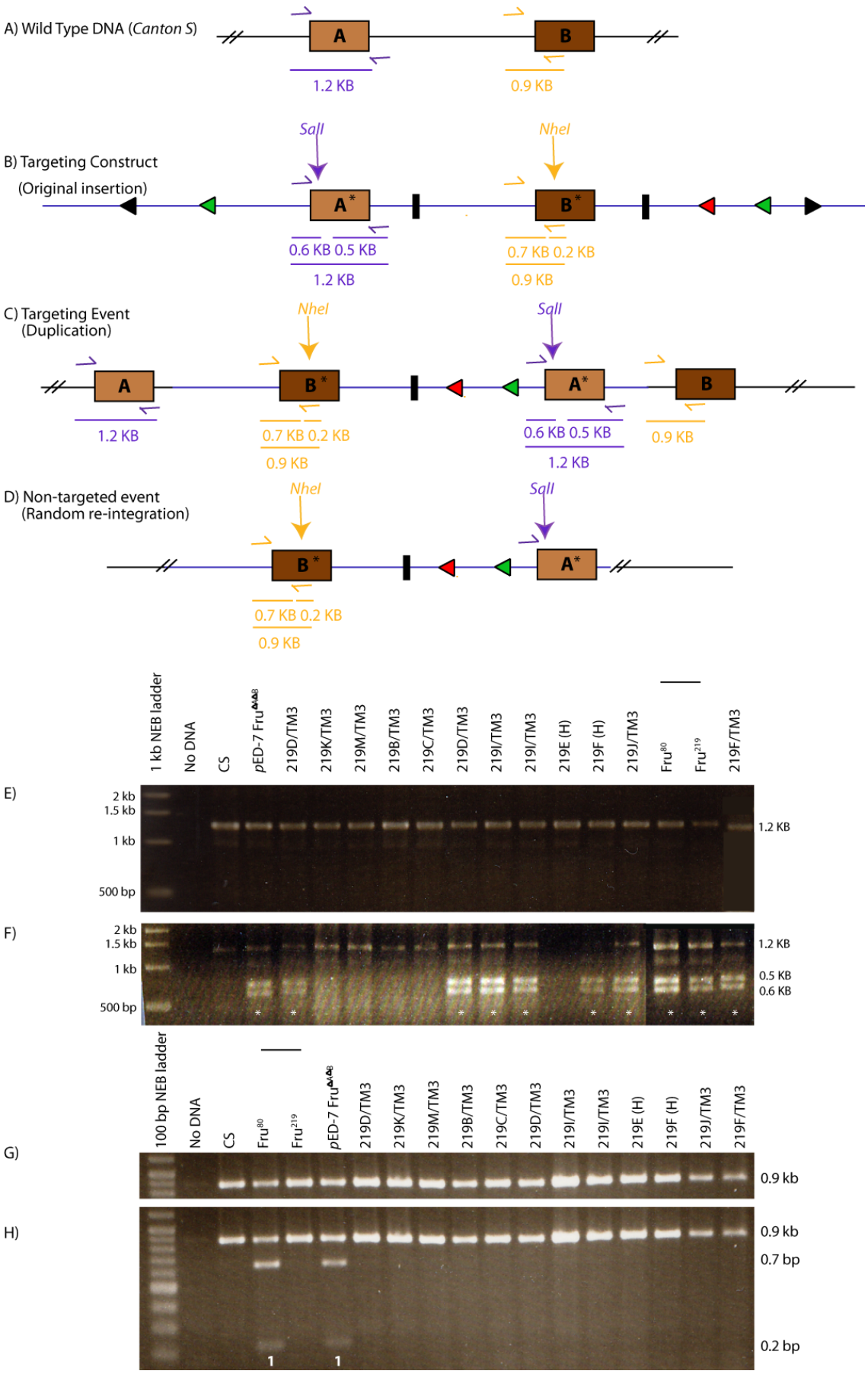


Figure 3-10 Verification of targeted events and PCR verification of HR mutants derived from Fru^{219}. The design of the homologous recombination targeting construct leads to a very specific insertion. (A) The wild-type gene showing the two relevant exons, A and B, and the location of primers to confirm the wild type. The distance between the primers is shown in kb, and reflects the resultant PCR product. (B) The original targeting construct, showing the mutated versions of exons A and B (signified by *) and depicting the schematic location of the restriction sites engineered into the mutant copies of the exons, and their effects on the fragments produced from restriction digests compared to those of the wild type. Arrowheads indicate components of targeting construct; *P*-element ends (black), FRT sequences (green) and *mW+* marker gene (red). I-*Sce*I and I-CreI endonuclease sites are represented by thick black lines, as seen in Figure 3-6. (C) The targeted event: Insertion of the mutated copies of the exons occurs between the two wild type exons, at which point gene products would show both the fragments engineered into the mutants, as well as the wild type original fragments. (D) A non-targeted event where integration has occurred randomly – in this instance only the B mutant would produce fragments, as one of the primers for the A mutant is outside the exon, and thus would not be integrated. The B mutant would show only the mutated copy fragments, with no wild type fragment. (E) Proof of the wild type A exon fragment (prior to digestion) (F) DNA fragments produced from the mutants, post digestion with *SalI* producing, in this instance, four mutants for the A exon, one of which can be seen to be homozygous (no wild type copy visible). (G) Proof of the wild type B exon fragment (prior to digestion). (H) DNA fragments produced from the mutants, post digestion with *NheI* showing no successful mutations of the B exon. [1] signifies the correct fragments expected for a B exon mutation. The black line identifies the un-resolved lines.

Viability of mutant genotype (% of expected number of progeny observed)					
Genotype	D	I	F	J	F (H)
D	0***	3***	0***	0***	*
I	0***	0***	0***	45***	*
F	0***	0***	0***	0***	*
J	0***	0***	0***	0***	*
F (H)	*	*	*	*	*
Fru$^{4-40}$	192	109	85***	145	76***
Fru^{sat15}	90**	90**	89***	174	41[1]***
Fru^{w24}	40***	87***	99	101	90[2]**
Drosdel	92**	139	125	117	*
CS	123	104	114	100	*

Table 3-1 Viability of Fru^{219} derived mutations. The viability crosses shown depict data for the five HR mutated lines derived from Fru^{219}, in the first resolution when crossed to themselves, each other and other known deficiency lines for Fru, as detailed in 2.6.2. The homozygous aspect of Fru^{219F} was lost during the production of this data and so '*' defines a cross that was not achieved due to a lack of flies of a suitable genotype, '[1]' denotes a an n of only 58, '[2]' denotes an n of only 86. Unless otherwise stated n > 100. Significant ratings are: *** p-value <0.001, ** p-value 0.001-0.01, other values were not significant to a p-value of 0.05 (Fisher's exact test).

Fru$^{80.1}$ was identified as being capable of resolving both Fru^{A} and Fru^{B} mutations, by PCR and restriction digest, Figure 3-10H. However, no mutants were achieved when it was resolved. This was not altogether surprising, as the rate of mutation

may be fairly low. Due to this lines ability to resolve both Fru^{A} and Fru^{B} mutants (thus suggesting proper incoporation of the donor DNA during HR) this line was taken on to be re-resolved.

Eleven vials were populated with the resolution line and $Fru^{80.1}$ and turned every third day, the original vial containing resultant larvae was heat shocked at 35 °C on the third day for 1 hr, and for 30 mins the following day. From this resolution 26 lines were derived of which nine mutants were resolved, four being verified as $Fru^{\Delta B}$ mutants, and five that were mutant in the $Fru^{\Delta A}$ exon, Figure 3-11. No double mutants were verified, however given the rate of mutation seen for $Fru^{\Delta A}$ and $Fru^{\Delta B}$ (0.15 and 0.1° respectively) the resulting double mutant would only be expected at a rate of 0.01, which is relatively low. For the purposes of this study, although a double mutant would be advantageous, it is not necessary. Further, it is unknown whether the double mutant would be seriously ill given *fru's* involvement in viability, or potentially lethal.

These mutants were again verified using the same restriction digests as for the first batch of resolved lines to confirm the presence of a mutated exon. The results of these digests are seen in Figure 3-11. From this point on, as all resolved lines are derived from the same original line, mutant lines will be referred to by their exon of mutation and resolved line number, i.e. 80.131/TM3 (Figure 3-11) will become Fru^{B131}/TM3, as it contains a mutation in exon B, and is line no 131 over a TM3 balancer chromosome.

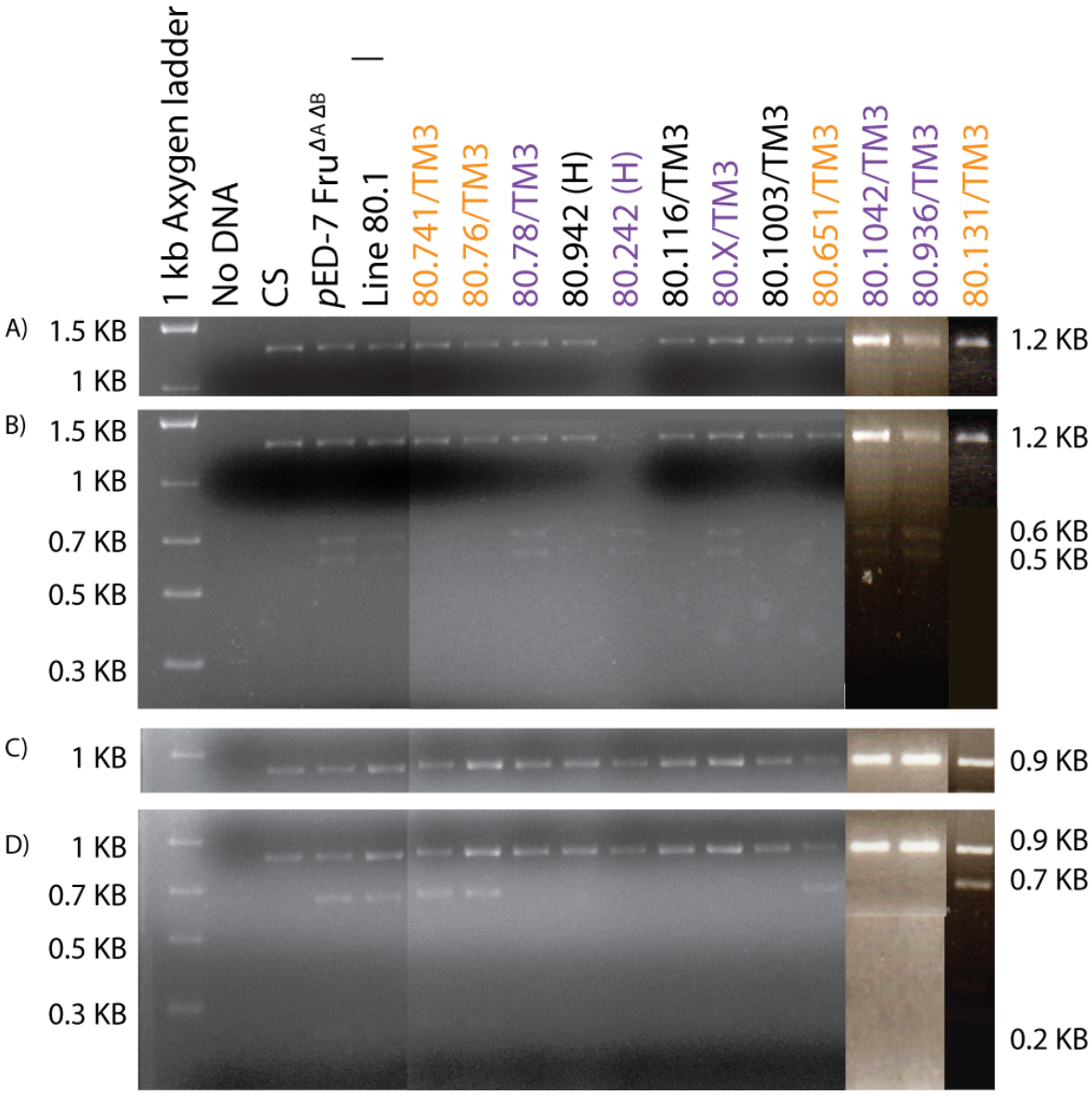


Figure 3-11 PCR verification of HR mutants derived from Fru$^{80.1}$. DNA was extracted from each mutant line and digested twice: once with *SalI* whose restriction site is engineered into the mutant Fru^{A} isoform, and once with *NheI* whose restriction site is engineered into the mutant B isoform. CS was used as a WT control, and the original targeting vector, *p*ED-7 as a positive control. (A – B) Digests run out on 1.5 % agarose showing presence of the *SalI* restriction site, and thus the presence of a mutated A exon. (A) Original WT band, pre-digestion, confirming successful extraction and amplification of the DNA. (B) Post-digestion results, showing five Fru$^{\Delta A}$ mutants (double band at 0.6 and 0.5 kb). (C – D) Digests run out on 1.5 % agarose showing the presence of the *NheI* restriction site, and thus the presence of a mutated B exon. (C) Original WT band, pre-digestion, confirming successful extraction and amplification of the DNA. (D) Post-digestion results, showing 4 Fru$^{\Delta B}$ mutants (bands at 0.7 kb, and 0.2 kb – barely visible due to strength of staining). No double mutants were identified. Fru$^{\Delta A}$ mutants are highlighted in orange font, Fru$^{\Delta B}$ mutants are highlighted in purple font. Axygen ladder was used for these gels. Black line represents un-resolved line.

To further verify the mutants and confirm the *fru* locus is otherwise intact, genomic Southern Blotting was performed. Biotin labelling was used as a non-radioactive alternative, which is reported to be more sensitive than standard radioactive methods [147]. Previous attempts to use radioactive Southern blotting

for *fru* could not detect the B exon successfully, leading to attempting non-radioactive alternatives. However, biotin labelling proved to be unsuccessful as the genomic DNA was not detected by standard methods. Within the timescale of this work, Southern blot was not successful. *fru* has previously proven difficult in some diagnostic procedures, such as Southern blots, therefore although a Southern blot will be desirable in the long term, it was not achieved in this time frame. An alternative approach would be to PCR and sequence the entire genomic region.

3.5.4 Characterising novel fru isoform mutants

Having successfully isolated several versions of both $fru^{\Delta A}$ and $fru^{\Delta B}$ mutants, the lines could now be used to identify the traits specific to each individual isoform. Initially, the compatibility of the mutants both with themselves and each other were assessed to identify any inviabilities associated with the mutant genotypes.

3.5.4.1 Viability

Flies were crossed (single freshly eclosed males to two virgin females) and incubated at 25 °C until the progeny eclosed. The progeny of each vial was scored according to their phenotypes to assess whether the flies produces offspring unexpected according to Mendelian inheritance. The outcome of these crosses is seen in Table 3-2.

$fru^{\Delta A}$ isoform mutants derived from line $Fru^{80.1}$ all proved to be homozygous viable. Generally, all lines derived from $Fru^{80.1}$ were found to be consistent in their viability, and were not only homozygous viable but largely showed no viability defects when crossed to themselves and each other. Low-level loss of viability (<25 %) was seen when they were crossed to $Fru^{\Delta B}$ mutants, however this wasn't

reflected in crosses to *fru*^{sat15} (representative of a *fru* null) and therefore was attributed to the B mutants which overall showed a more substantial drop in viability.

Viability of mutant genotype (% of expected number of progeny observed)														
Genotype	936	X	78	242	1042	936 (H)	X (H)	78 (H)	242 (H)	1042 (H)	131	651	76	741
936	99	96	107	98	123	100	137	108	113	93	73***	157	130	85
X	89**	117	106	91**	81***	104	110	100	87***	90**	110	137	57***	92
78	78***	97	101	120	102	146	85***	122	98	103	83**	110	155	101
242	113	98	82**	140	145	104	112	113	108	105	174	81***	132	96
1042	127	95	110	100	117	101	103	127	113	78***	66***	76***	180	81*
936 (H)	124	179	104	118	88***	100	100	100	100	100	62***	123	103	99
X (H)	125	85***	99	98	104	102	100	100	102	100	81***	119	51***	88
78 (H)	102	86***	86*	184	99	100	100	100	100	100	89*	98	84***	76**
242 (H)	105	122	102	94*	119	100	100	100	100	100	113	106	83***	38***
1042 (H)	99	95	92	109	80***	100	100	100	100	100	54***	129	99	68***
131	91*	95	99	32***	27***	87***	96	117	107	90**	20***	48***	77***	8***
651	102	81***	122	77***	86***	94*	111	98	88***	106	119	67***	106	45***
76	108	74***	79***	73***	123	88***	165	112	95	92**	55***	48***	42***	95
741	89**	108	75***	76***	90**	52***	92*	73***	89***	121	51***	31***	113	63***
Fru$^{4-40}$	100	106	51***	91**	115	118	105	109	106	110	12***	92	85***	102
Fru^{sat15}	76***	96	125	119	194	110	93*	73***	117	90**	10***	60***	90**	63***
Fru^{w24}	111	108	100	109	107	97	129	95	100	116	90	77***	91**	96
Drosdel	69***	74***	88	144	90**	94*	100	100	100	100	154	140	85***	88
CS	99	104	96	126	118	100	155	100	100	100	97	126	149	93

Table 3-2 Viability crosses for the second resolution mutants. Mutant lines derived from the HR derived Fru$^{80.1}$ were crossed to themselves, each other and known deficiencies in the *fru* locus to characterise the viability phenotype incurred by the mutation. Yellow shading identifies homozygous progeny, and green shading represents deficiency crosses and the control, CS. n > 100. Significant values are: *** p-value<0.001, ** p-value 0.01-0.001, * p-value 0.05- 0.01, remaining values are not significant to a p-value of 0.05 (Fisher's exact test).

fru$^{\Delta B}$ mutants derived from line Fru$^{80.1}$ were found to be less consistent across the board than the *fru*$^{\Delta A}$ mutants, however it is clear that although the lines were between 40- 60 % viable as homozygotes, they are sick and do not survive well as a homozygous stock. These lines were best kept over a balancer chromosome. Crossing the *fru*$^{\Delta B}$ mutants to each other similarly give 40- 60 % viability supporting the hypothesis that they mutate the same region of the genome. When crossed to *fru*$^{\Delta A}$ mutants, overall viability improves (to around 80

%) suggesting some redundancy between common elements of the two exons. The *fru*$^{\Delta B}$ lines, which appear to be most consistent, and therefore most likely to be showing the most accurate effects of loss of isoform B, are *fru*$^{\Delta B651}$ and *fru*$^{\Delta B741}$.

Crossing the mutant lines to known deficiencies of the *fru* locus can also assess viability. For the purposes of these crosses, Df(3R) *fru*$^{4-40}$ deletes the sex-specific promoter, P2 and their transcripts; Df(3R) *fru*^{sat15} deletes the common promoters and their transcripts; and Df(3R) *fru*^{w24} largely deletes both sex-specific and common transcripts. Therefore, for the mutation to be affecting the common transcripts of *fru*, a decrease in viability would be expected with Df(3R) *fru*^{sat15} and Df(3R) *fru*^{w24} but not Df(3R) *fru*$^{4-40}$. More specifically, the decrease in viability in Df(3R) *fru*^{sat15} should mirror any decrease in viability when the mutant is crossed to itself. All *fru*$^{\Delta A}$ mutants were fully viable against all the deletion lines showing, as would be expected given earlier results, that the *fru*^{A} isoform on its own does not have an effect on the viability of the fly. This agrees with known details of the *fru*^{A} isoform: its limited expression and distinct function compared to the other isoforms [2]. The *fru*$^{\Delta B}$ mutant lines as previously discussed were not so consistent. Fru$^{\Delta B131}$ clearly differs from the other *fru*$^{\Delta B}$ lines, being practically inviable in its homozygous state (20 %), which was mirrored in the Df(3R) *fru*^{sat15} cross (10 %). Further support that this mutation wasn't only affecting the *fru*^{B} isoform was that the line was also heavily inviable when crossed to Df(3R) *fru*$^{4-40}$ (12 %), suggesting other mutations may have been incurred alongside the incorporation of the donor DNA, leading to the line being discarded from further investigation. Of the remaining three-*fru*$^{\Delta B}$ lines, *fru*$^{\Delta B76}$ was also slightly unusual, not as severe as *fru*$^{\Delta B131}$, however showed lower viability in its homozygous state (40 %) and was therefore perhaps not as representative. The final two mutations *fru*$^{\Delta B651}$ and *fru*$^{\Delta B741}$ show consistency (~ 60 % viability – which was mirrored in the

homozygous cross) and so were taken on for further investigation. Further confirmation of their associate with the *fru*^{B} isoform was the slight drop in viability when combined with the Df(3R) *fru*^{w24} deficiency, consistent with Df(3R) *fru*^{w24} deleting both the sex-specific and common transcripts. This preliminary data showed isoform *fru*^{B} to have similar characteristics to the *fru*^{C} isoform, as characterised using the *fru*$^{\Delta C}$ mutation by Billeter et al., in 2006. The drosdel mutant, which deletes P1 and P2 and possibly P3 (the deletion end point is very close to the P3 promoter, so it is uncertain as to whether the transcripts derived from this promoter are fully functional) was slightly affected, both in the *fru*$^{\Delta A}$ and *fru*$^{\Delta B}$ mutants – however, due to the scope of this deletion, it is not particularly informative. From these results, further characterisation of known *fru* phenotypes was undertaken.

3.5.4.2 Fertility

fru is known to have an effect on male-specific fertility, therefore to characterise the contribution of individual isoforms, fertility tests were carried out on the new mutants. Crosses were set up as per 2.6.3, results of which can be seen in Table 3-3. Briefly, single flies (virgin) were crossed to two males (naïve) and incubated at 25 °C for 10 days. Their ability to produce viable offspring was assessed. The contribution of the sex-specific isoform (from P1 promoter) was specifically assessed by combining mutants with the Df(3R) *fru*$^{4\text{-}40}$ and Drosdel deficiencies.

Fertility (% of individuals of each genotype able to produce offspring)												
Genotype	/*fru*$^{4-40}$		/*fru*^{sat15}		/*fru*^{w24}		/Drosdel		/wild-type (+)		Homozygote	
Sex of deficiency	M	F	M	F	M	F	M	F	M	F	M	F
A936	93	97	100	100	96	100	92	100	97	100	98	100
A242	100	100	100	100	100	100	93	95	100	95	100	100
A1042	100	96	91	100	97	100	100	100	100	95	100	100
A78	100	100	95	100	96	100	97	100	100	100	100	100
AX	100	100	100	100	90	93	100	97	100	100	95	100
B131	100	100	100	100	97	100	96	100	100	100	100	100
B651	93	100	100	100	100	91	100	100	100	100	83	100
B741	80	100	100	96	90	95	100	100	100	100	100	100
B76	85	100	95	100	86	100	95	100	100	95	97	97
Wild Type	100	100	100	95	100	100	92	100	N/A	N/A	100	100

Table 3-3 Fertility crosses depicting the effect of isoform mutations on sucessful reproduction. The effect of each mutation on fertility in each sex is derived through crossing each sex to a number of deficiencies of *fru*. Fertility is associated with the sex-specific transcripts of Fru, and so should be present in *fru*$^{4-40}$ crosses. n > 20. In this table, a value from the box A936/ *fru*$^{4-40}$ male shows the result for the fly line w^{+}; +; *fru*$^{\Delta A936}$/ *fru*$^{4-40}$; + x CS (where the mutant line is the male). In this table, homozygote refers to the mutant crossed to another fly of the same genotype. No results were significant, $p < 0.05$, using Fisher's exact Test.

No significant fertility issues were identified in either the *fru*$^{\Delta A}$ or *fru*$^{\Delta B}$ mutants, suggesting that transcripts from neither isoform contribute to successful fertilisation.

3.5.4.3 Antibody staining shows isoform-specific loss of expression

To unequivocally confirm the mutations, antibodies were used to show that the mutations have completely and specifically lost expression of the appropriate *fru* isoform. At this point only two *fru*$^{\Delta A}$ mutants (*fru*$^{\Delta A242}$ and *fru*$^{\Delta A936}$) and the two most reliable *fru*$^{\Delta B}$ mutants (*fru*$^{\Delta B651}$ and *fru*$^{\Delta B741}$) were tested. To confirm mutagenesis, antibodies that stain for Fru male-specific proteins (Fru^{M}) and antibodies that stain individual isoforms of Fru (Fru$^{A/\ B/\ C}$) were used to co-stain both wild-type and mutant adult CNSs which had been staged to two-days post eclosion. The results of these stainings are shown in Figure 3-12 and Figure 3-13 (anti-Fru^{C} not shown).

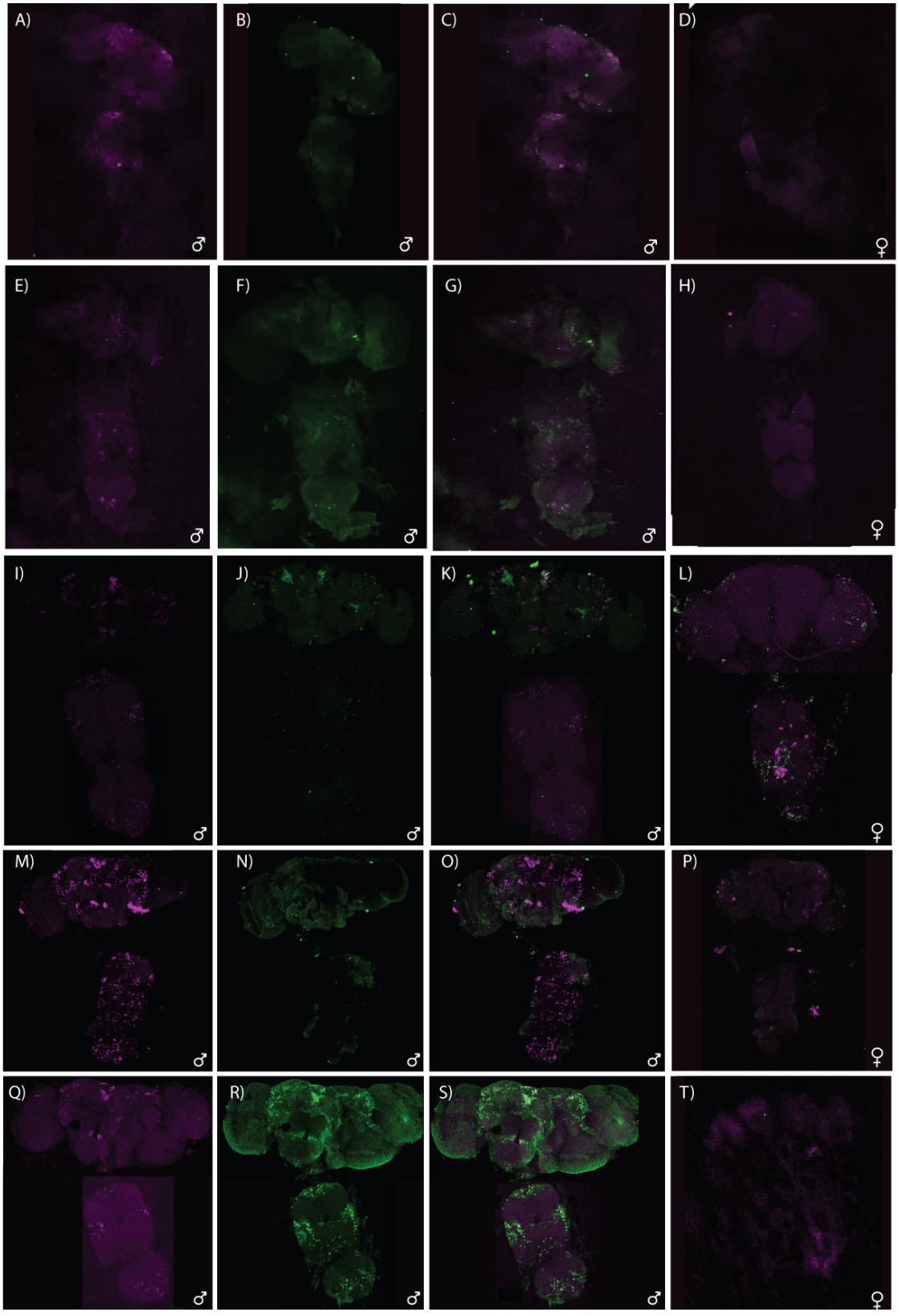


Figure 3-12 Expression of Fru^{A} in CNS of HR mutants. The expression of Fru^{A} protein in the HR mutants of *fru* derived from this study was assessed. 2 day adult CNS were stained with Fru^{M} (magenta) and Fru^{A} (green) and maximal Z-stack images taken on a confocal microscope. (A-C) *fru*$^{\Delta A242}$ male, (D) *fru*$^{\Delta A242}$ female, (E-G) *fru*^{A936} male, (H) *fru*$^{\Delta A936}$ female, (I-K) *fru*$^{\Delta B651}$ male, (L) *fru*$^{\Delta B651}$ female, (M-O) *fru*$^{\Delta B741}$ male, (P) *fru*$^{\Delta B741}$ female, (Q-S) CS male, (T) CS female.

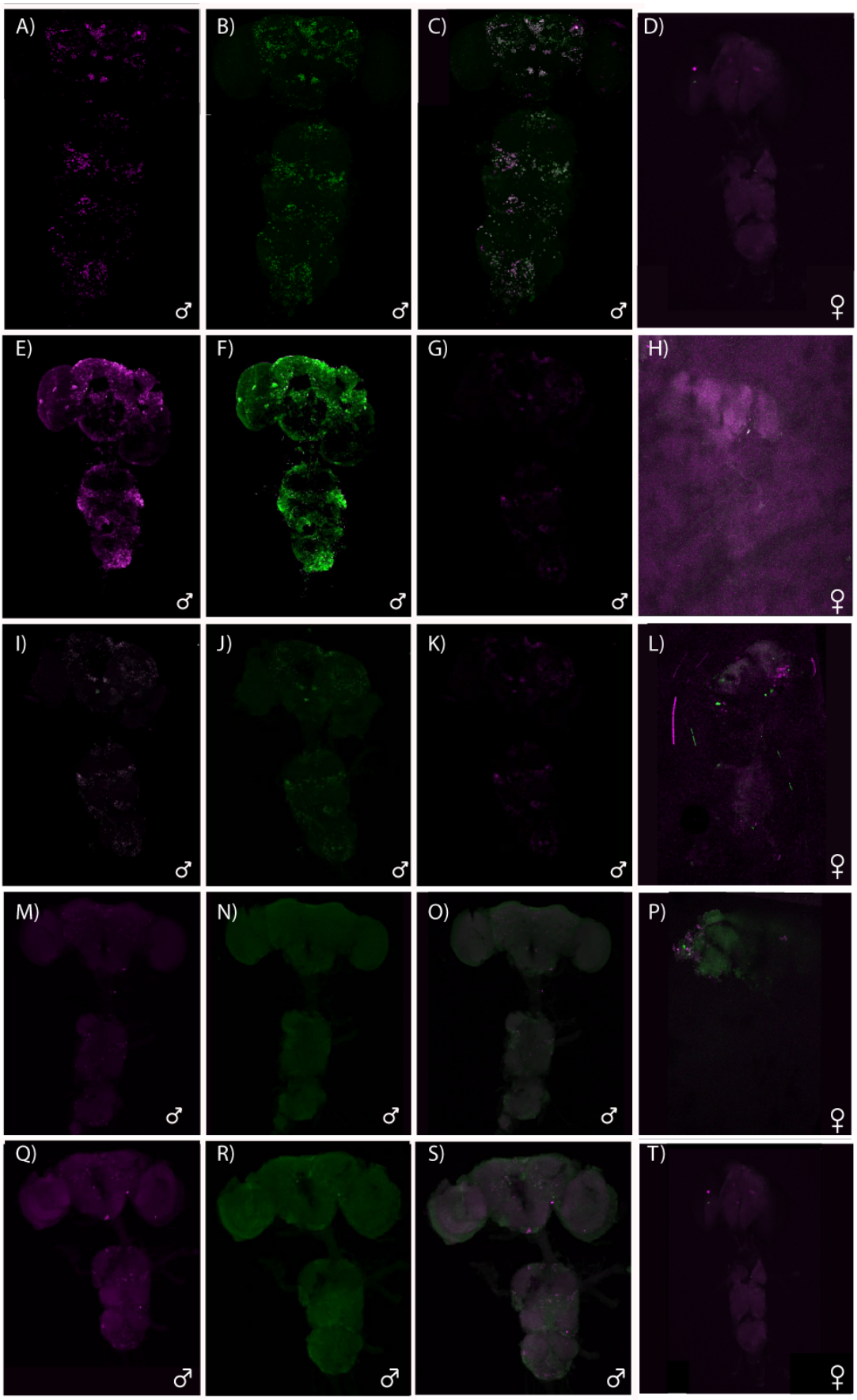


Figure 3-13 Expression of Fru^{B} in CNS of HR mutants. The expression of Fru^{B} protein in the HR mutants of *fru* derived from this study was assessed. 2 day adult CNS were stained with Fru^{M} (magenta) and Fru^{B} (green) and maximal Z-stack images taken on a confocal microscope (A-C) CS male, (D) CS female, (E-G) *fru*$^{\Delta A242}$ male, (H) *fru*$^{\Delta A242}$ female, (I-K) *fru*$^{\Delta A936}$ male, (L) *fru*$^{\Delta A936}$ female, *fru*$^{\Delta B651}$ male, (L) *fru*$^{\Delta B651}$ female, (M-O) *fru*$^{\Delta B741}$ male, (P) *fru*$^{\Delta B741}$ female,

They showed in each instance that the mutant male in question is missing only the protein for the exon that has been mutated (Fru^{A} protein for $fru^{\Delta A}$, Fru^{B} protein for $fru^{\Delta B}$) and that all other Fru proteins are expressed as normal (both express all Fru^{C} proteins, data not shown). Further, no sex-specific Fru expression is seen in females. Now that these mutants have been confirmed by a number of methods, the phenotypic consequences of their loss of function can be investigated.

3.5.4.4 Generating a MOL

One of the distinct functions derived from the isolation of $fru^{\Delta C}$ was the identification that Fru^{MC} is necessary for successful formation of the MOL. It is thought that only the Fru^{MC} isoform is required for MOL formation, however due to a lack of $fru^{\Delta A}$ and $fru^{\Delta B}$ isoform mutants, this hasn't been confirmed. To resolve this issue, two $fru^{\Delta A}$ mutant lines, and two $fru^{\Delta B}$ mutant lines were dissected to isolate the MOL and stained as per (2.1.5.4). The results of these dissections are seen in Figure 3-14, and shows as predicted that neither the $Fru^{\Delta A}$ mutant or $Fru^{\Delta B}$ mutant is lacking a MOL, thus confirming the prediction that this developmental role of Fru in males is solely derived from the sex-specific Fru^{MC} isoform [2]. This also confirms these lines contain a functional Fru^{MC} isoform.

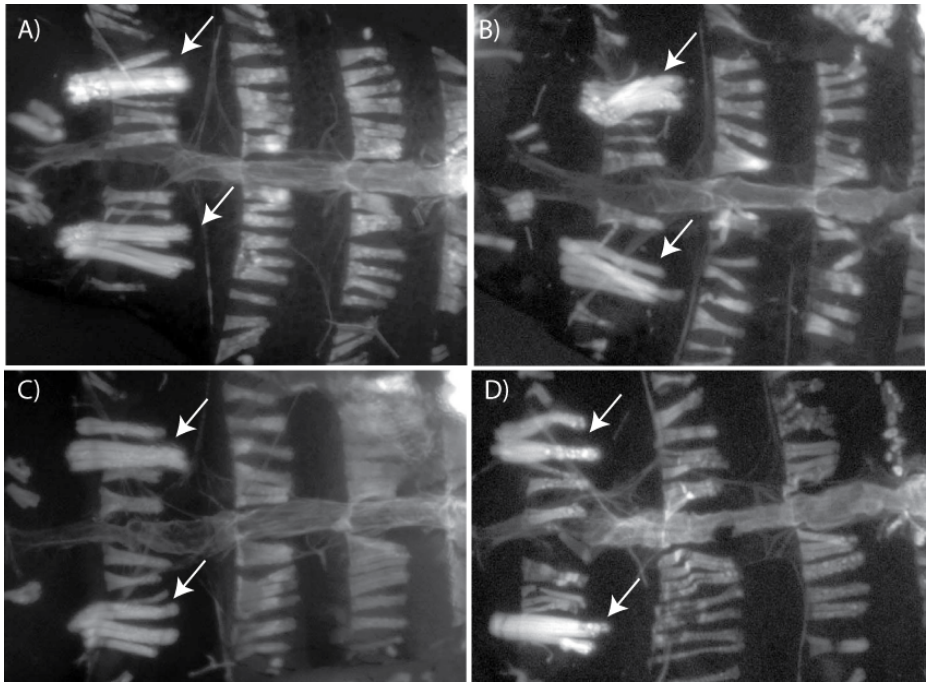


Figure 3-14 Evidence of a MOL in *fru*$^{\Delta A}$ and *fru*$^{\Delta B}$ HR mutants. MOL production is thought to be solely the function of the Fru^{MC} isoform. This was confirmed by staining for MOL in two mutants of both isoform A and isoform B males. (A) *fru*$^{\Delta B651}$ (B) *fru*$^{\Delta B741}$ (C) *fru*$^{\Delta A936}$ (D) *fru*$^{\Delta A242}$. In each instance, a fully formed MOL is identified (indicated by a white arrow). Female 48 hr. adult were stained as a control, and showed no MOL formation (data not shown). n > 20.

To further investigate the mutants derived by HR, the mutant lines were cantonised by Mr Darren Parker, by crossing individual mutants to CS. Cantonising the mutants moves the mutants into a genetic background common to the lab, making any subsequent behaviour analysed comparable to other data generated. Further, it removes any secondary or *de* novo mutations which may have occurred during mutagenesis [148]. Progeny from these flies were crossed to CS, their offspring crossed to CS, and the flies from the second cross tested for the relevant mutation (using the original PCR assay) once their progeny were established. Lines which had not retained the mutation were discarded. This was repeated six times before balancing the mutants over a third chromosome balancer [149]. This intricate method of cantonising is required with the mutants resulting from ends-in HR, as the mutation cannot be followed using the *mW+* marker, which is resolved out in the last step. Therefore the mutation must be molecularly assayed after each cantonising step. From the resultant cantonised

lines, one $fru^{\Delta A}$ mutant ($fru^{\Delta A936}$), and one $fru^{\Delta B}$ mutant ($fru^{\Delta B651}$) were taken on to be further characterised.

3.5.5 Characterising the cantonised fru mutants

Most strikingly, after characterisation, the $fru^{\Delta B}$ mutants were found to be 100 % inviable. This suggests that the viability incurred over time in the un-cantonised lines was in fact due to a second-site mutation generated during the HR process [149]. Such mutations have previously been documented, and further investigated by Roy et al., who found that HR by the ends-in approach had the potential to incur lethal mutations at both the initial random integration of the donor construct, or at the resolution step. They defined a spontaneous lethal mutation rate of 2.95 % which was higher than 0.6 % previously identified [150, 151]. They proposed that this increase in base mutation could be linked to the heat shock process. Beyond this, they identified a 3.5 fold increased mutation rate through the remaining processes of ends-in HR, significant to nearly 99 %, clearly identifying the possibility that non site-specific mutations occur at an increased frequency, as the cantonising of the $fru^{\Delta B}$ alludes to. However, as this second-site mutation is present in all the fru^{B} mutants, it is possible that the mutation exists in the genetic background line used for the HR process. In both cases, post-cantonising, the mutation should be removed and no longer be affecting the phenotype displayed.

To re-characterise the lines, post-cantonisation, viability was repeated for $fru^{\Delta A936}$ and $fru^{\Delta B651}$, Table 3-4. The results of the viability were the same as those pre-cantonisation for $Fru^{\Delta A936}$, confirming that fru^{A} is not involved in the viability of the fly. $fru^{\Delta B651}$ was shown to be completely inviable, both as a homozygote and over fru^{sat15} which is deficient for the common *fru* promoters, which control viability of the fly. It is viable over the deficiency $fru^{4\text{-}40}$, which does not remove the common

fru transcripts from promoter P3 and P4. Further, when balanced against TM6 (whose *Tb* phenotype is visible in the pupae time point), it was clear that the homozygotes die at the pupae time point, which is the time that loss of *fru* common expression leads to lethality [34, 78]. When these terminated homozygous pupae were dissected from their pupae case, malformations indicative of a *fru* mutation were observed, Figure 3-14. The wings fail to expand fully in both the sexes, and the metathoracic (third) leg showed malformation in the proximal tarsal which appeared to have a partially formed extra joint. Although not identical, this is similar to phenotypes previously documented in *fru* mutants [34].

Percentage viability of mutant genotype				
Genotype	$fru^{\Delta A936}$ male	$fru^{\Delta A936}$ female	$fru^{\Delta B651}$ male	$fru^{\Delta B651}$ female
$fru^{\Delta A936}$	117 (78)	109 (94)	115 (163)	107 (131)
$fru^{\Delta B651}$	107 (131)	115 (163)	0 (56) ***	0 (54) ***
fru^{4-40}	113 (105)	99 (121)	100 (101)	110 (136) ***
fru^{sat15}	98 (125)	96 (146)	0 (90) ***	9[1] (98) ***
CS	97 (105)	100 (132)	100 (121)	100 (101)

Table 3-4 Viability of cantonised mutants. Post-cantonisation, *fru* mutants were re-assessed for viability. Figures in red represent a drop of over 20 % in viability, n is in brackets. [1] represents 3 escapers. All escapers died shortly after eclosion. Significant values are: *** p-value<0.001, remaining values are not significant to a p-value of 0.05 (Fisher's exact test).

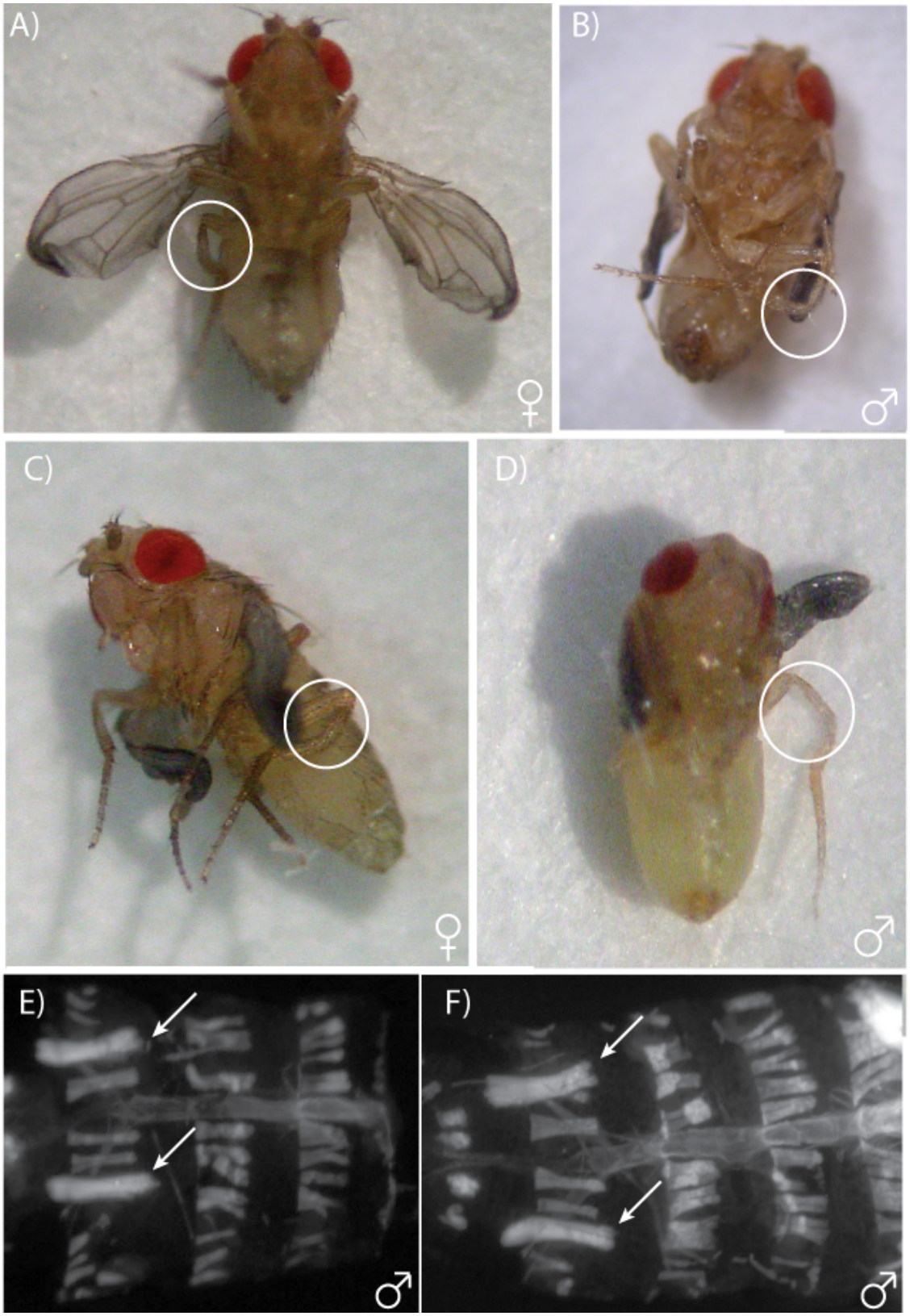


Figure 3-15 Malformations associated with *fru*$^{\Delta B}$. Homozygous *fru*$^{\Delta B}$ mutant and *fru*$^{\Delta B}$ over the *fru*^{sat15} deficiency were removed from the pupae case, as they are incapable of escaping. Phenotypes were observed are associated with loss of the *fru*^{B} exon. (A, B) Female and male (pharate) flies, respectively, of the *fru*$^{\Delta B651}$ / *fru*^{sat15} phenotype showing wings that are only partially expanded and leg defects in the rear (metathoracic) leg second tarsal. Three females successfully escaped the pupae casing for this phenotype, of which the female in A is one. (C, D) Pharate female and male flies, respectively, of the *fru*$^{\Delta B651}$/ *fru*$^{\Delta B651}$ phenotype showing no attempt to expand the wings, and leg defects in the rear (metathoracic) leg, second tarsal. (E) Male +/ *fru*$^{4-40}$ phalloidin-stained MOL. Both MOL are present and fully formed, indicated by the white arrows. (F) Male *fru*$^{\Delta B651}$ / *fru*$^{4-40}$

Similarly, fertility was re-assessed, however as was originally seen in the uncantonised data, neither mutant showed any significant loss of fertility, Table 3-5.

Percentage fertility				
Genotype	fru^{4-40}		fru^{sat15}	
Sex of deficiency	Male	Female	Male	Female
A242	100 (10)	100 (10)	81 (11)	100 (12)
B651	100 (8)	100 (13)	(lethal)	(lethal)
Wild Type	100 (6)	100 (20)	96 (24)	100 (22)

Table 3-5 Fertility of cantonised HR mutants. Cantonised mutants were crossed to Df(3R) fru^{4-40} and Df(3R) fru^{sat15} to identify any drop in fertility that they induce. No significant fertility issues were identified, $p < 0.05$, using Fisher's exact Test. Number of samples in each test is in parenthesis.

To observe the effects of the mutants on the courtship behaviour of the fly, several basic courtship indices were tested, Figure 3-16. Initially, courtship initiation, time to copulation, copulation duration and courtship index (CI) were assessed. Briefly, courtship initiation is the time in seconds it takes for the fly to show three seconds of courtship behaviour; Time to copulation is the time in minutes from courtship initiation to copulation; copulation duration is the time (in minutes) that the flies spend copulating; and CI is the total time (in 10 mins) that the male fly spends displaying courtship behaviour towards the female (or until copulation, whichever occurs first).

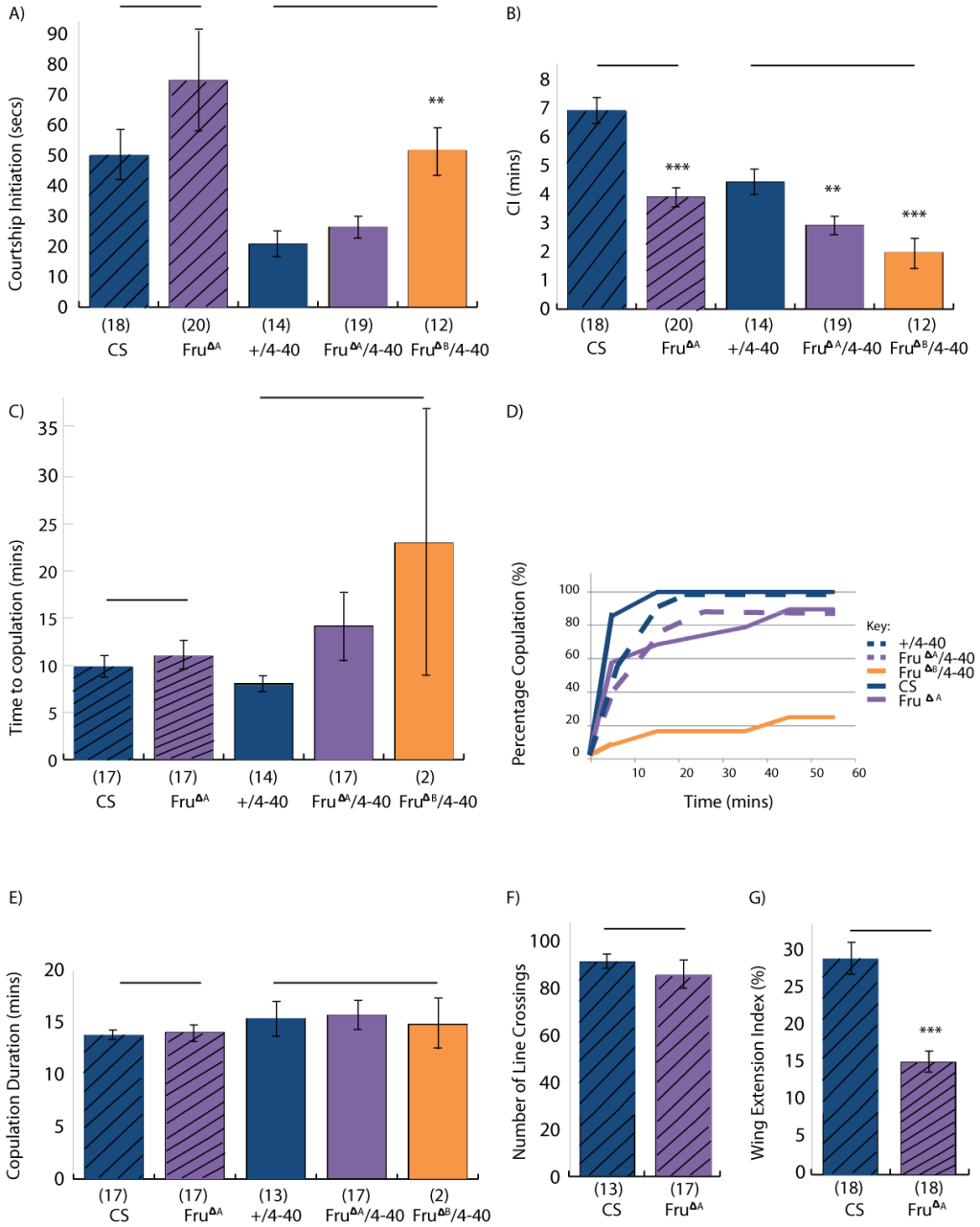


Figure 3-16 Courtship Behaviour Analysis of HR mutants. The two cantonised mutants, *fru*$^{\Delta A242}$ and *fru*$^{\Delta B651}$ were analysed for various aspects of courtship to identify the idiosyncrasies of the individual isoforms. The mutants were analysed in two groups: initially homozygous *fru*$^{\Delta A242}$ were analysed directly with CS (patterned columns/ dashed lines), then both mutants were crossed to Df(3R) *fru*$^{4-40}$ and analysed to WT/ Df(3R) *fru*$^{4-40}$, as *fru*$^{\Delta B}$ is homozygous inviable (solid coloured columns/ solid lines). Naïve males were aged for three to five days, and single virgin females aged for five to ten days. Single naïve males and single virgin females were combined in small courtship chambers and recorded for an hour. From these recordings the following analyses were taken: (A) Time taken for males to initiate three seconds of courtship, (B) Courtship Index (CI) which records the amount of time in a ten minute period that the male spends performing any behaviours associated with courtship, (C) Time taken for copulation to be initiated, and (D) Percentage copulation over a period of one hour, and finally, (E) the duration of courtship in minutes. Further analyses of Fru$^{\Delta A242}$ were undertaken to attempt to identify the point during courtship behaviour at which A mutants are deficient: (F) Number of line crossings in a ten minute period, identifying any locomotion issues, and (G) wing extension index (WEI) to identify a lack of wing extension during courtship. Error bars represent ± 1 standard error. Homozygous *fru*$^{\Delta A242}$ was analyses using the Mann Whitney unpaired test (assuming non-parametric data with equal variance). Behaviour crossed to Df(3R) *fru*$^{4-40}$ was analysed using Kruskal-Wallis one-way ANOVA with Dunn's pairwise comparison (non-parametric data). Solid black lines above the data columns identify which datasets were statistically analysed against which controls. Numbers in parentheses indicate sample size. In all instances (irrespective of statistical analysis invoked) p-values are as follows: ***

p-value<0.001, ** p-value 0.01-0.001, * p-value 0.05- 0.01, remaining values are not significant to a p-value of 0.05.

From this data it appeared that *fru*$^{\Delta A}$ mutants displayed a slight delay in courtship initiation, however this was not found to be statistically significant, Figure 3-16B. More significantly was a drop in CI (just under four minutes in comparison to just under seven minutes for the WT) representing almost a 50 % drop, which was highly significant, Figure 3-16B. To rule out a drop in overall locomotor activity, a line crossing assay was used to check that the mutants' movement was not impaired. No significant difference between the *fru*$^{\Delta A}$ mutant and WT was identified, Figure 3-16F. Upon closer inspection of the mutant it appeared that it spends less time with its wing extended compared to the WT, and so wing extension index (WEI) was considered– whereby percentage courtship in a ten minute period spent with wing 45 – 90 °C was timed, Figure 3-16G [107]. Here, a highly significant change was observed: the *fru*$^{\Delta A}$ mutant spent on average only 15 % of courtship with its wing extended, compared to 28 % observed in the WT. This suggests the potential for courtship song to be interrupted, however in the scope of this study there was not sufficient time to investigate this.

As *fru*$^{\Delta B}$ is inviable as a homozygote, courtship behaviour for both mutants was assessed in combination with the *fru* deficiency Df(3R) *fru*$^{4\text{-}40}$ (indicative of loss of the sex-specific transcripts, while not effecting *fru* related viability). No significant change was observed in time taken to initiate copulation or copulation duration of those flies that successfully copulated (as was also the case for the Fru$^{\Delta A}$ homozygote) for either mutant, however most (83 %) *fru*$^{\Delta B}$ mutants failed to copulated within one hr., indicating a major disruption of courtship behaviour, Figure 3-16C, D. Fertility tests have shown that this is not due to an inability to

copulate or due to physical sterility, suggesting courtship is delayed due to inaccurate/ incomplete performance of early courtship steps.

To attempt to identify where in courtship behavioural sequence a loss of fru^{B} transcripts is having its effect, the courtship sequence was broken down. A significant increase in time taken to initiate courtship was identified for $fru^{\Delta B}$/ Df(3R) fru^{4-40}, Figure 3-16A, which was not seen in $fru^{\Delta A}$/ Df(3R) fru^{4-40} mutants, indicating that this accounted for at least part of this extended period of courting. However, given the highly significant drop in CI, Figure 3-16B, which was considerably more than that of $fru^{\Delta A}$, it is unlikely that this is the only factor. Time constraints have prevented this from being further investigated within the scope of this study, but clearly further investigation is required.

3.6 Discussion

Creating mutants which inhibit specific isoforms allows more freedom in defining the roles of *fru*. Further, once characterised, such mutants can be used for characterising the downstream targets of individual Fru isoforms. The experiments undertaken in this chapter successfully produced both $fru^{\Delta A}$ isoform and $fru^{\Delta B}$ isoform mutants. The coveted $fru^{\Delta A/ B}$ double mutant was not achieved. Although this double mutant would further elude to the relationship between the individual isoforms of *fru*, the potential downsides of pursuing this mutant meant that it wasn't of great enough importance at this point in time. Those downsides being: its potential occurrence at a very low frequency, at least as low as 0.01 (and likely lower), given the number of lines created in these experiments; its likelihood of being either severely ill or lethal, considering the known functions of *fru* [34, 78, 79].

Five *fru*$^{\Delta A}$ mutants were defined: *fru*$^{\Delta A1042}$, *fru*$^{\Delta A242}$, *fru*$^{\Delta A78}$, *fru*$^{\Delta A936}$ and *fru*$^{\Delta AX}$. These mutants were all found to be homozygous viable, and showed similar behavioural attributes suggesting they all contained successful targeted mutations in the *fru*^{A} exon. They showed little evidence of any viability concerns, the only exception being when crossed to the un-cantonised *fru*$^{\Delta B}$ mutants – which was attributed to the *fru*$^{\Delta B}$ mutant rather than the *fru*$^{\Delta A}$ mutant through deficiency crosses, furthermore this was not seen with the cantonised mutants.

Four *fru*$^{\Delta B}$ mutants were defined: *fru*$^{\Delta B131}$, *fru*$^{\Delta B651}$, *fru*$^{\Delta B741}$ and *fru*$^{\Delta B76}$. Initially, these mutants were not homozygous viable. As they were maintained, homozygous flies started to occur. This eventual occurrence of homozygous viable flies was initially thought to suggest a low level of fitness that required many generations to clearly manifest itself. However, post-cantonisation, this viability as a homozygote was lost, and the viability previously attained attributed to a second-site suppressing mutation. Upon further investigation it was found that other labs had had similar experiences of second-site mutations during ends-in HR, thus highlighting the need to ensure mutants generated by HR are sufficiently backcrossed into a known genetic background (beyond the requirement for comparisons during behaviour analysis) [149].

Through viability crosses, it was seen that there is a viability phenotype associated with the *fru*^{B} isoform. In the cantonised mutant this was 0 %, which was mirrored when crossed to the Df(3R) *fru*^{sat15} deficiency, which carries a deletion of the common promoters of *fru* (the common promoters being those associated with viability). The exception to this were three female escapers of phenotype *fru*$^{\Delta B651}$/ Df(3R) *fru*^{sat15} which successfully eclosed, however lived for less than 24 hrs., and exhibited deformities associated with *fru*: deformation of

the metathoracic leg and partially expanded wings. When pharate adults of homozygote genotype and $fru^{\Delta B651}$/ Df(3R) fru^{sat15} were dissected out of the pupae casing, these flies were also found to exhibit deformed metathoracic legs and partially expanded wings. Deformed metathoracic legs have been associated with loss of the $fru^{\Delta C}$ isoform, suggesting an overlap in function between these two isoforms ($fru^{\Delta C}$ also showing a loss of viability, albeit less severe) [2]. Partially expanded wings have also previously been associated with *fru*, but not associated with a specific isoform [81]. Once Drosophila emerge from their pupal case they must expand their wings, a behaviourally driven process that involves forcing hemolymph into the wings to unfold them. Interestingly, a recent abstract from Benjamin White's lab (NIH) has found a link between 13 of the 28 known BTB containing genes in Drosophila (of which *fru* is one) and the wing expansion response [152]. This study suggests that it is the Fru^{B} isoform that plays a role in wing-expansion, further characterisation of the $fru^{\Delta B}$ mutant should lead to a greater understanding of the role of *fru* in this process.

Fertility was tested in each mutation, but was shown to be unaffected by either a loss of fru^{A} or fru^{B}. In light of this, to confirm that the mutants were indeed mutated in the appropriate exons (and to confirm that no protein was produced), antibody stainings of two of each mutant were completed. $fru^{\Delta A242}$ and $fru^{\Delta A936}$ of the $fru^{\Delta A}$ mutants, and $fru^{\Delta B651}$ and $fru^{\Delta B741}$ of the $fru^{\Delta B}$ mutants were co-stained with anti-Fru^{M} and one of the Fru isoform specific antibodies. This confirmed that in each instance the appropriate isoform did not express protein product, and that the remaining isoforms remained unaffected. Further, Phalloidin staining of the abdominal musculature showed these mutants have fully formed MOLs. The implications of this result are two-fold: the formation of the MOL is dependent on

the Fru^{C} isoform alone, and therefore as the MOLs are present Fru^{C} expression is normal in these mutant backgrounds.

Partial analysis of courtship behaviour showed $fru^{\Delta A}$ to have a significant drop in CI, which was correlated with a drop in wing extension. This will result in a drop in the amount of courtship song, however the ability to produce song is currently being investigated by a collaborator (D. Parker and M. Ritchie, St.Andrews). $Fru^{\Delta B}$ was shown to take significantly longer to initiate courtship and to have a highly significant drop in CI, which has not currently been attributed to a decrease in performance of specific behaviours. The most dramatic behavioural phenotype was seen in the percentage copulation: only two of 12 $fru^{\Delta B}$ males copulated within one hr., although fertility assays showed these flies did eventually mate, and were fertile.

Overall, successful $fru^{\Delta A}$ and $fru^{\Delta B}$ isoform null mutants were shown to have been isolated by HR. Analysis of these mutants showed that fru^{B} has some characteristics in common with isoform fru^{C}: a common loss of viability and common deformities of the metathoracic leg. However there are now clear differences between the isoforms: $fru^{\Delta B}$ shows only certain morphological deformations and does not affect fertility. Potentially most distinctively a fully formed MOL is exhibited in $fru^{\Delta B}$, which is clearly absent in $fru^{\Delta C}$ mutants [2]. All this data suggests that although there may be some basic similarities between fru^{B} and fru^{C}, and perhaps a level of redundancy, there are clearly distinct roles they each play.

Although there are clearly other characteristics still to be investigated with these mutants (analysis of song, further investigations into the viability aspect of *fru*

function and a more detailed analysis of courtship behaviour), it is clear that the tools by which *fru* can be investigated have been increased through the addition of these novel isoform mutants, which will be required in the future to tease apart the idiosyncrasies of the individual *fru* isoforms, and the targets they govern.

4 Identifying Putative Downstream Targets of Fru using DamID

4.1 Introduction

We now have a good understanding of the basic functions of *fru*, and have started to correlate those functions with specific isoforms in the previous chapter (and by Billeter et al., previously) [2, 5, 34, 37, 39, 78, 79, 81, 110]. To this end, we now have the tools to tease apart the individual contributions of the *fru* isoforms, and begin to understand how they coordinate to produce the wide-ranging phenotypes associated with *fru*. The next step is to identify targets of the Fru proteins, the binding site(s) used in this targeting process, and the way in which these targets contribute to the *fru* phenotypes.

To identify downstream targets of Fru throughout the genome, a robust whole-genome approach must be used. There are several methods currently in use which differ dramatically in methodology, cost, and robustness of results. These methods fall into two common categories: tiling array (or chip) analysis, and high-throughput sequencing analysis [15, 16, 153]. Although high-throughput sequence analysis provides more robust data, its more recent development makes it more expensive, and at the time of this work was untenable for our studies. For this study, tiling array methodology was the more amenable option for which there are two commonly used methods: **Ch**romatin **I**mmuno**p**recipitation, analysed using a genomic tiling array (ChIP-chip), and DNA adenine methyltransferase identification (DamID) [16, 126, 154].

The most widely used method of identifying the occupancy of a transcription factor throughout the genome is ChIP-chip [154-156]. Using this method, DNA is cross-linked with formaldehyde to fix the proteins to their DNA targets. The cross-linked DNA is then sonicated to fracture it into small fragments which are isolated

using an antibody specific to the protein of interest (POI). The isolated fragments are labelled and washed over a tiling array to identify their location in the genome [155, 157, 158]. Unfortunately, our extant Fru-antibodies work well for immunohistochemical analysis, however have thus far been unsuccessful for immunoprecipitation experiments. As this method relies heavily on a functioning antibody, currently this method is not an option for Fru.

An alternative to using a Fru-specific antibody is the introduction of a protein-fusion tag with the Fru protein(s). These tags are recognised by antibodies and can therefore be used for precipitation. For Fru, this method was explored using the Tandem Affinity Purification (TAP)-tagging method first in described in 1999 [159]: whereby *fru* cDNAs were fused to the TAP tag (consisting of Protein A sequence, a **c**almodulin **b**inding **p**eptide (CBP) sequence and a **T**obacco **E**tch **V**irus (TEV) protease recognition site) which could be bound on a column of IgG beads which bind Protein A. Post isolation of fragments identified by precipitation, the TAP tag could be cleaved via the engineered TEV site, a step that has been shown to decrease the background in such experiments [115]. An alternative approach to ChIP is **D**NA **a**denine **m**ethyltransferase **id**entification (DamID) [14], which utilises a lack of native methylation in *Drosophila* to circumnavigate the requirement of an antibody and has been shown to be highly reproducibility in previous screens [1, 14, 160]. Both approaches were initially explored in the laboratory for Fru, and DamID was found to be the most immediately effective and was therefore adopted for future experimentation.

4.1.1 DamID v ChIP-chip

The process of DamID leads to some specific differences between its output and that of ChIP-chip. Importantly, DamID is much more reproducible – allowing replicates to be easily amalgamated, and thus nuances of individual datasets to be discounted at an early stage. Comparatively, ChIP-chip gives a more defined peak (Figure 4-1). This again is due to technical differences in the two approaches – whereas ChIP-chip produces small generically sheared fragments of target DNA through sonification, which are identified by the cross-linking of the POI to its target and subsequently pulled down with the use of an antibody; DamID produces wider peaks, as methylation of the target sequence can occur at any GATC sequence (spaced 200- 300 bp apart on average), therefore peaks can be large, given that methylation has been shown to spread up to 5 kb surrounding the location of the gene [14]. Another fundamental difference is that ChIP assays where the protein of interest *is* at a given time (snapshot), whereas DamID assays where it *has been* (footprint). The reason for this is that the methylation stays on the DNA after the Dam fusion protein goes away.

ChIP was first pioneered in 1988, but has only recently been used in conjunction with a tiling array – leading to the creation of a powerful tool, ChIP-chip [161, 162]. ChIP-chip has the potential to identify a binding site to within a few hundred base pairs, however it requires a highly specific antibody and can incur false positives due to the cross linking process [126]. Comparatively, in the case of Fru, DamID's obvious advantage is its lack of required antibody, and its high reproducibility. Further, it successfully identifies indirect binding events, which may be relevant for a number of Fru's genomic associations.

DamID detects protein-DNA interactions *in vivo,* and adenine methylation has limited effects on the topology of the DNA, further adding to the strength of this sensitive tool [163]. The major restriction of DamID is the addition of the Dam protein itself to the *Drosophila* system – which is not native. Although the molecule itself is small and therefore not likely to interfere with the cellular mechanisms, it is toxic in high levels and has the potential to accumulate methylation in the system – potentially making older flies appear to have more targets than they actually do – albeit this concept is speculative. ChIP-chip is probably the more robust system – however given that there is currently no suitable antibody for Fru, DamID is a very viable alternative (Figure 4-1).

fru has been shown to have differing roles throughout development [2, 79, 81]. Therefore, to fully understand *fru* function, it is necessary to identify Fru genomic targets in the male NS at various developmental time points, as these targets are likely to change throughout development as the roles of Fru vary. Three specific developmental time points were chosen: L3 'wandering' larvae (onset of Fru expression); 46 hr. pupae (peak of Fru expression); and 48 hr. adults (once development is complete, to observe the effects of the gene in the fully formed fly). Further, the individual isoforms are known to have individual phenotypes [2], therefore each developmental time point needs to be considered separately for each sex-specific isoform, Fru^{MA}, Fru^{MB} and Fru^{MC}. To address the specificity of Fru binding, DNA binding mutants can be created by mutating the cysteine residues in the C_2H_2 ZnF of the DNA-binding domains. DamID screening at all these time points, as well as DamID analysis of DNA binding mutant lines, will provide data required to start elucidating the downstream targets of Fru.

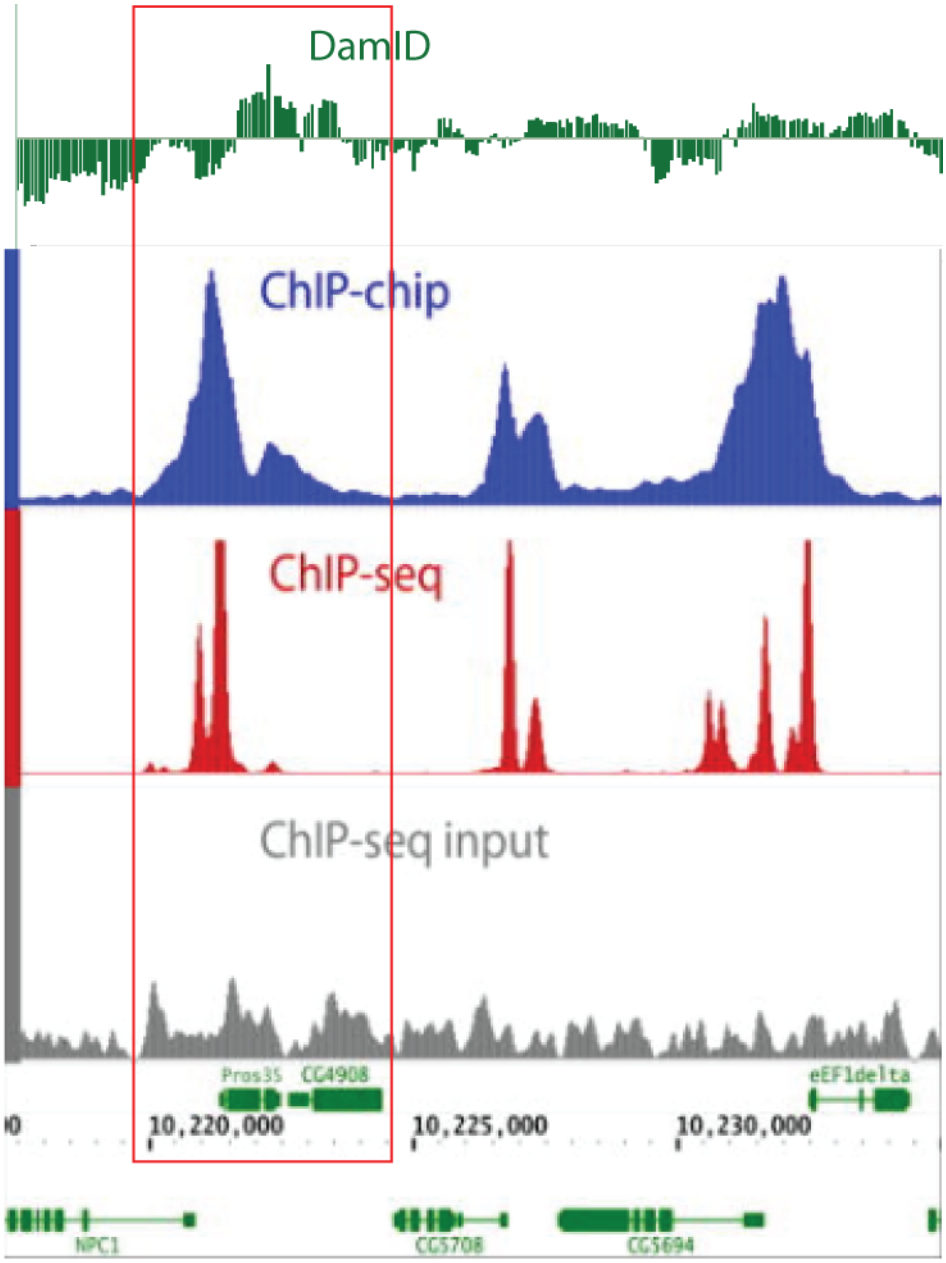


Figure 4-1 ChIP-chip v DamID. The varying methods for identifying putative targets of a transcription factor produce different qualities of output data. The graph above (modified from Park, 2009) shows the different peaks formed from different input data. The DamID peak (green) is wider, and doesn't necessarily sit directly above the binding site. The ChIP-chip data (blue) produces a much more specific peak than the DamID data, which is superseded further by the very specific peak formed by ChIP-seq data (red) which in itself is discernable from the original input data of ChIP-seq (grey). The data shown shows a significant peak for *prospero* (*pros*) which is highlighted in the red box [153].

4.2 Aims and objectives

Produce datasets representative of the genes targeted by Fru isoforms Fru^{MA}, Fru^{MB} and Fru^{MC} at three specific time points: L3 'wandering' larvae; 46 hr. pupae; 48 hr. adult. Complement these datasets with DNA binding mutant datasets to

show the effects of DNA-binding specificity and potentially uncover interactions due to co-factor binding. The resultant datasets should be normalised, with the help of collaborators, to account for false positives, and to compensate for the idiosyncrasies of the method so as to accurately portray the downstream genomic targets of Fru.

4.3 DNA binding mutants

fru has various levels of complexity built into it: alternative splicing allows multiple isoforms to be produced, both sex-specific and common to the sexes [34, 78, 79, 81]; the C-terminal BTB protein-protein interaction domain allows the formation of homodimers of Fru and potentially heterodimers with co-factors [89, 92, 121, 125, M. Neville, unpublished]; and three alternative N-terminal C_2H_2-ZnF DNA binding domains [102]. To separate out the different types of binding of Fru to DNA (direct binding, or indirect binding via dimerisation), DNA binding mutants were created using multi site-directed mutagenesis (SDM) to target the cysteines of the zinc fingers. DNA binding mutants were created for each of the sex-specific isoforms: Fru^{MA}, Fru^{MB} and Fru^{MC} (2.4.13).

4.3.1 Site-directed mutagenesis primer designs

Site-directed mutagenesis uses primers to incorporate mutations into *fru* cDNA-containing plasmids. To mutate the individual *fru* isoforms, two sets of primers were created for each isoform. As each isoform has two C_2H_2 ZnF, two sets of primers were designed for each isoform in order to target the ZnF independently (listed in Table 4-1). Due to the close proximity of the cysteines resides within the *fru* sequence, each primer was designed to incorporate two independent

mutations. The Quikchange (Stratagene, #200514) protocol used for the mutagenesis requires primers between 25 and 45 bp in length with a melting temperature of more than 75 °C. It is also necessary for the mutation to be near the centre of the primer, with 10- 15 bp of complementary sequence on either side for successful complementation with the native strand.

The introduced mutations are non-conservative substitutions of cysteines in the Zn-fingers into alanines. As the structure and therefore DNA-binding ability of a C_2H_2 ZnF depends entirely on $Zinc^{++}$ ion coordination by the cysteine and histidine resides in the fingers, non-conservative substitution of these residues disrupt Zn^{++} ion binding and therefore DNA binding [103]. It is necessary only to mutate at least one cysteine in each ZnF to prevent successful DNA binding [102].

Isoform	Primer Name	Sequence	Tm (°C)
A	7AF	CCCACTCTATACCGA**GCC**GTTAGT**GCC**TTGTTGTAGCACAGG	82.5
	7AR	CCTGTGCTAGAACAA**GGC**ACTAAC**GGC**TCGGTATAGAGTGGG	82.0
	7BF	CCGCAGAGTCATGAG**GCC**CCCGTT**GCC**GGGCAGAAATTCACTCG	93.8
	7BR	GCGAGTGAATTTCTGCCC**GGC**AACGGG**GGC**CTCATGACTCTGCG	93.8
B	25zf1for	GCTGCAACCTGCACCGT**GCC**AAGCTG**GCC**GGCAAGGTCGTCACGC	96.6
	25zf1rev	GCGTGACGACCTTGCC**GGC**CAGCTT**GGC**ACGGTGCAGGTTGCAGC	96.6
	25zf2for	CCGGGCCGCTTCGAG**GCT**CCGCTG**GCC**CGGGCCACCTATACGC	96.3
	25zf2rev	GCGTATAGGTGGCCCG**GGC**CAGCGG**AGC**CTCGAAGCGGCCCGG	96.3
C	1zf1for	CGGGCACCATGTGGCGG**GCC**CGCTCC**GCC**GGCAAGGAGGTGACC	99.6
	1zf1rev	GGTCACCTCCTTGCCGGCGGAGCG**GGC**CCGCCACATGGTGCCCG	99.6
	1zf2for	CCGCCCAGCGGTCCATG**GCT**CCCTAC**GCC**CCGGCCACCTACAGC	95.9
	1zf2rev	GCTGTAGGTGGCCCGG**GGC**GTAGGG**AGC**CATGGACCGCTGGGCGG	95.9

Table 4-1 Primers required for site directed mutagenesis. Each isoform requires four primers, two forward and two reverse, to cover the two Zn-fingers. Mutated cysteines are underlined.

4.3.2 Sub-cloning the fru fragment

For DamID analysis, the *fru* cDNA sequences were inserted into the 9.9 kb *p*UAST-NDam plasmid [22]. This plasmid size is too large to use with the Quikchange Protocol, whose limit is an 8 kb plasmid, and so the *fru* fragment

must be sub-cloned into a smaller plasmid. The isoform-specific *fru* fragments were sub-cloned into either *p*Zero-2 (3.3 kb, Invitrogen) or *p*SC-B (3.5 kb, Agilent Technologies) as seen in Figure 4-2. fru^{MB} and fru^{MC} were cloned into *p*Zero-2 which contains kanamycin resistance, *lac-Z* and a MCS, whereas fru^{MA} was cloned into *p*SC-B which contains ampicillin resistance, *lac-Z* and a MCS. The *fru* fragments were excised from *p*UAST-NDam-$fru^{MA/\ MB/\ MC}$ constructs using the appropriate restriction enzymes (described in Table 4-2) and ligated into either *p*Zero-2 or *p*SC-B that had been digested with the complementary restriction enzymes (Figure 4-2). Restriction enzyme digests confirmed the correct insertions and these plasmids were subjected to multi SDM.

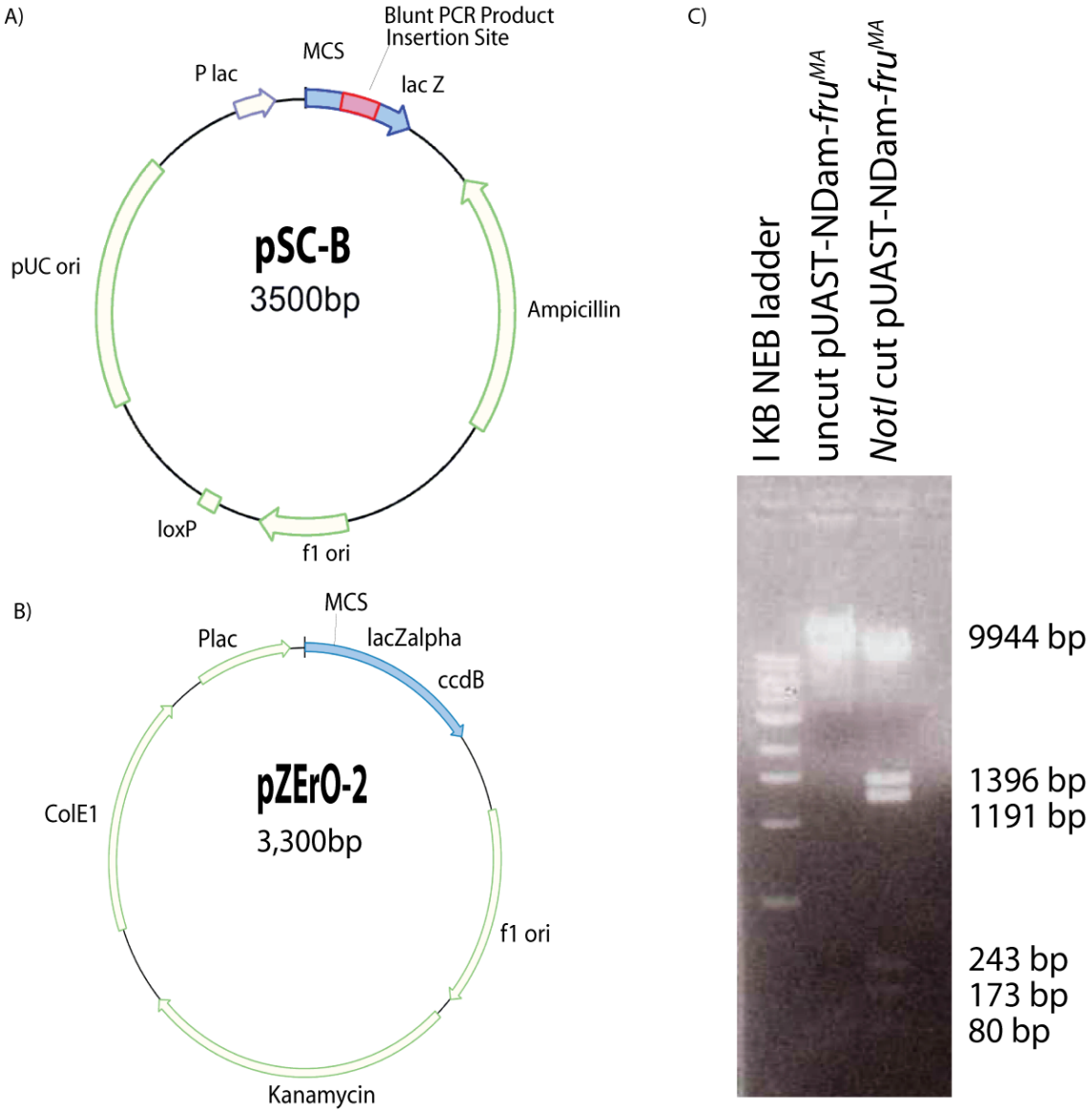


Figure 4-2 Sub-cloning the *fru* fragment. The limitations of the Quikchange protocol require *fru* to be sub-cloned into a smaller vector to be mutated. Two smaller plasmids were used to mutate the three isoforms of *fru*. (A) *p*SC-B, a 3500 bp plasmid used to mutate the fru^{MA} fragment. (B) pZerO-2, a 3300 bp plasmid used to mutate the fru^{MB} and fru^{MC} fragments. (C) An example of restriction digest fragments (using fru^{MA} as an example) confirming the correct ligation of fragments into the *p*UAST-NDam plasmid, using a *NotI* restriction digest.

fru isoform	Restriction enzymes
MA	*NotI, KpnI*
MB	*BglII, PvuI*
MC	*PvuI, KpnI*

Table 4-2 Restriction enzymes used to excise *fru* fragments from pUAST-NDam-$fru^{MA/\ MB/\ MC}$. Restriction enzymes required to excise the Fru fragment from the NDam plasmid for insertion into a smaller plasmid.

4.3.3 Multi site-directed mutagenesis

The Quikchange multi SDM kit (Stratagene, #200514) was used to generate targeted mutations of the four cysteines in each *fru* isoform (described in Figure 4-3). The Quickchange kit uses the natural methylation that occurs in *E. coli* bacteria (in which the plasmid of interest is propagated) along with the lack of methylation incorporated into DNA when it is PCR amplified (using mutagenic primers) to identify the mutated versus non-mutated DNA. The plasmids to be mutated were therefore grown in a methylating bacteria strain (DH5α). Primers incorporating the required mutations were used to PCR amplify the entire plasmid, generating a non-methylated version of the plasmid. The original methylated, non-mutated template DNA, was then restriction enzyme digested with *DpnI* which targets methylated GATC sequences. As GATC sequences are found approximately every 256 bp, the methylated template DNA will no longer form a functional plasmid. The remaining non-methylated (and hopefully mutated) DNA is transformed back into *E. coli*, where it is amplified, purified and then subjected to sequence analysis.

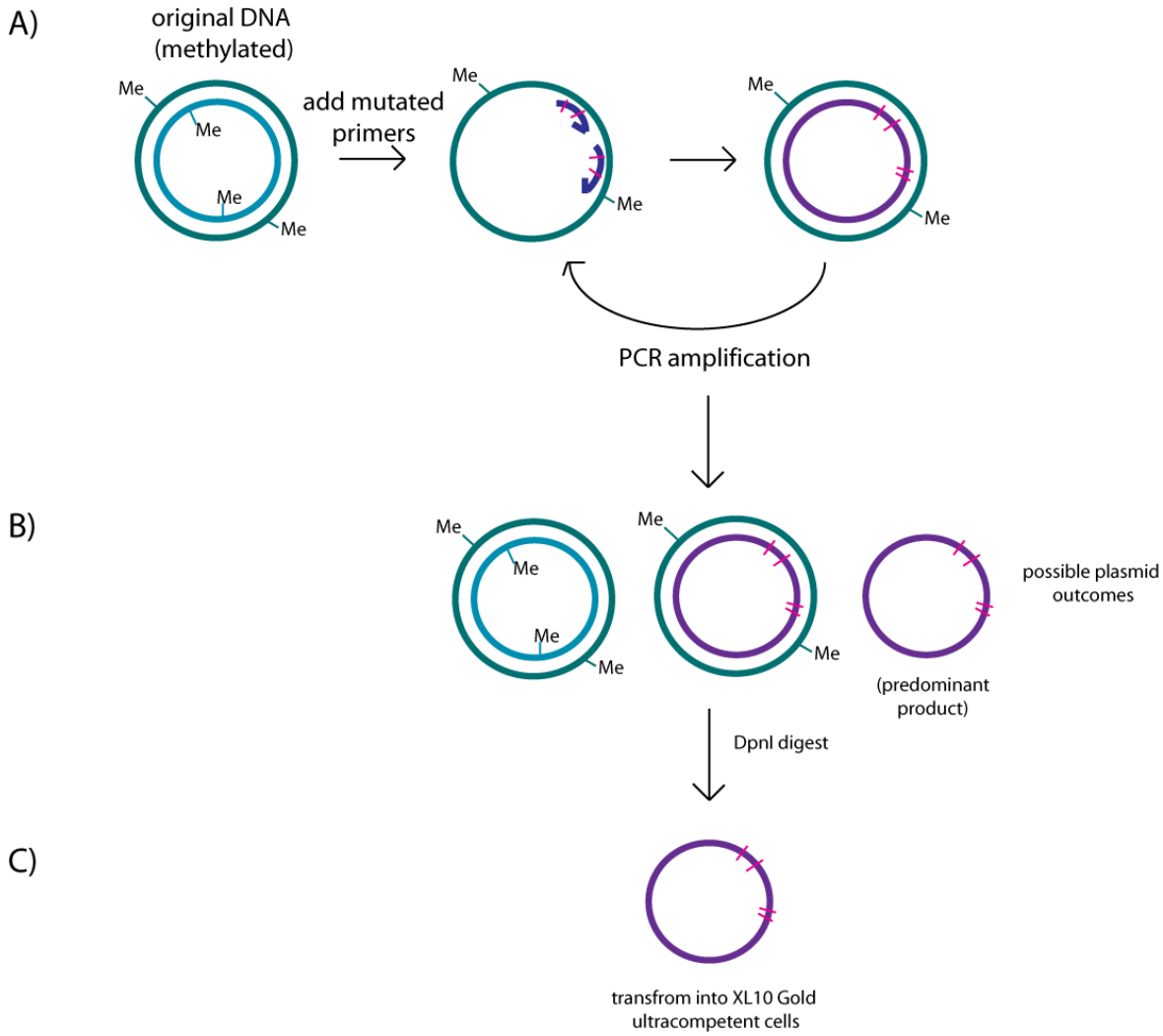


Figure 4-3 Quikchange multi site-directed mutagenesis schematic. The above schematic outlines the steps required to successfully mutate a small plasmid (< 8 kb). (A) Plasmid is grown in a methylating cell to produce methylated DNA. This DNA is copied using PCR incorporating nucleotides, which are non-methylated from a primer incorporating the required mutations. (B) The sample is *DpnI* digested to remove all original DNA. (C) The remaining DNA is transformed into ultracompetent cells. Image modified from Quikchange multi site-directed mutagenesis protocol.

4.3.4 Mutagenesis

Each primer should be added into the reaction in equimolar quantities for efficient mutagenesis. For each reaction a total of four primers were added and the PCR amplification run as per (2.4.11). After the *DpnI* digestion a diagnostic restriction digest was performed to confirm the plasmid is correct, post mutation.

4.3.5 Transformation and sequencing to confirm mutagenesis

Post-mutagenesis the plasmid was transformed into XL-10 Gold® ultracompentent cells 2.3.4. Transformants were selected using the antibiotic resistance present on the mutated plasmid (kanamycin for pZero-2, ampicillin for *p*SC-B), plasmids were then purified and sent for sequencing to confirm the presence of mutations. Sequence analysis confirmed the following successfully incorporated mutations (Figure 3): *fru*^{MA} incorporated three mutations, *fru*^{MB} incorporated three mutations and *fru*^{MC} incorporated four mutations.

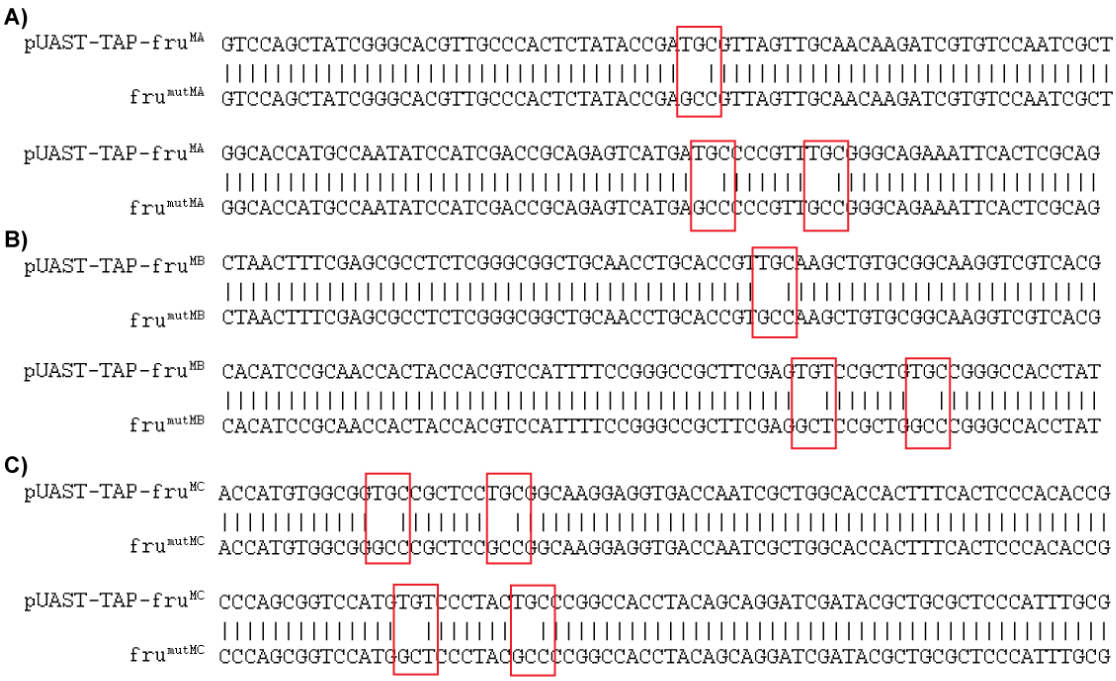


Figure 4-4 Mutations in the DNA binding domain of *fruitless*. Using site directed mutagenesis, up to four cysteines in the C_2H_2 zinc fingers of *fruitless* were mutated to prevent secondary structure formation and therefore DNA binding. The location of the mutations for each of the three sex-specific isoforms are shown within the red boxes. The top strand in each instance is the original plasmid sequence, and the bottom strands the mutated form. A vertical line represents a match between the strands, therefore the mutation is represented by a blank space. The mutations were a first base substitution converting cysteine to alanine. (A - B) Three mutations were achieved in both the A and B exon zinc finger, one in the first finger and both in the second finger. (C) All four mutations were successful in the C exon zinc finger.

4.3.6 3-way ligation, microinjection and balancing

To create the final *p*UAST-NDam-*fru*$^{mutA/\ B/\ C}$ plasmids, 3-way ligations were performed to combine the *p*UAST-NDam plasmid with the common 5′ end of *fru* and the appropriate mutated 3' *fru* fragment. Each DNA fragment was digested with specific restriction enzyme as described in Table 4-3. The *p*UAST-NDam plasmid was linearized and ligated together with the *fru* 5′ end that had been excised from the original *p*UAST-NDam plasmid (this was necessary to enable the use of all the necessary sites) along with the mutated 3' *fru* fragment excised from the successful SDM generated plasmids (*p*Zero-2/ pSC-B). The digested DNA fragments were separated using agarose gel-electrophoresis and the appropriate DNA fragments were excised and gel purified (examples of successfully digested and purified fragments prior to ligation are shown in Figure 4-5A).

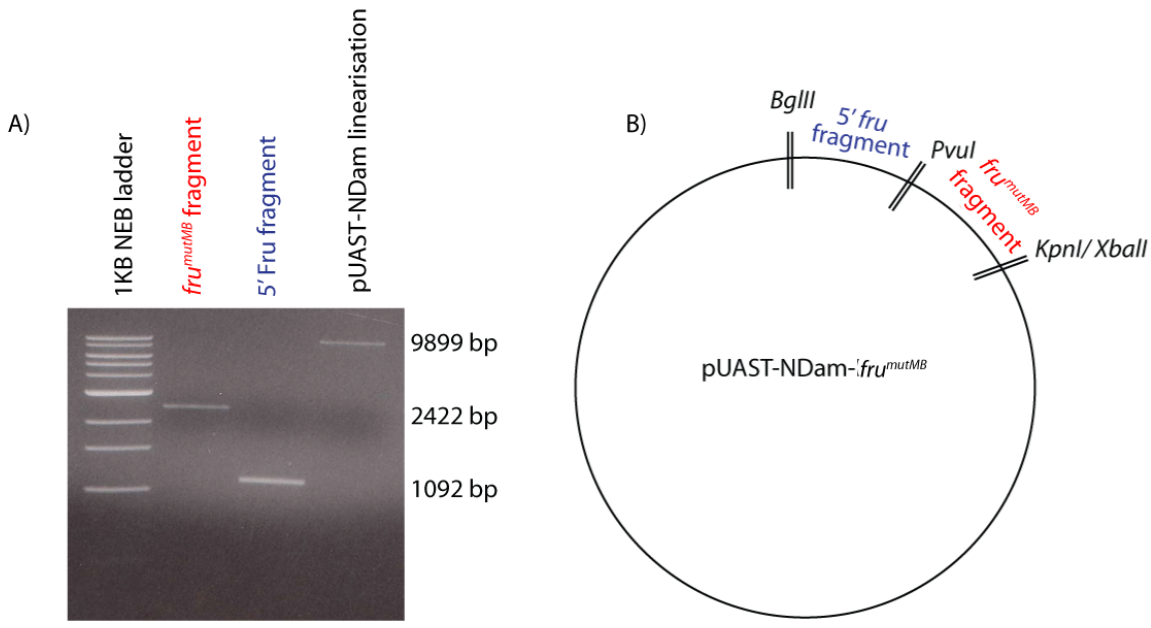


Figure 4-5 3-way ligation of mutated *fru* fragment into the *p*UAST-NDam plasmid. To be able to successfully ligate the mutated fragment (in this example *fru*^{mutMB}) into the original *p*UAST-NDam plasmid, three fragments must be ligated simultaneously. Initially each fragment is digested from its individual plasmid, and gel purified. The mutant fragment is cut out with *PvuI* and *KpnI*, the 5' fragment of the *fru* gene is cut out with *BglII* and *PvuI*, and the *p*UAST-NDam plasmid linearised with *BglII* and *KpnI*. (A) Confirmation of the fragments to be ligated. (B) A schematic of the ligation.

Fragment	Fragment size (bp)	Restriction Enzymes Required
NDam	9899	*KpnI, BglII*
fru 5′	1092	*BglII, PvuI*
Mutated Fragment	2422	*PvuI, KpnI*

Table 4-3 Fragments required for 3-way ligation of the NDam plasmid. To re-create the original NDam plasmid, the individual components were cut out of their respective plasmids with specific restriction enzymes to allow them to recombine together in a specific order.

Three-way ligations were performed, and the resulting plasmid was restriction enzyme digested to confirm correct integration of all DNA fragments (Figure 4-5B). The final *p*UAST-NDam-*fru*$^{mutA/\ B/\ C}$ plasmids were then sent for sequencing to confirm the correct insertion of all fragments. The *p*UAST-NDam plasmid contains *P*-element sequences that enable it to be randomly incorporated into the *Drosophila* genome in the presence of the transposase enzyme, in addition to the *white*$^{+}$ reporter gene enabling its easy identification by red eye colour. The *p*UAST-NDam-*fru*$^{mutA/\ B/\ C}$ plasmids were sent to Genetic Services Inc. (Cambridge, MA.) where they were microinjected into the germ line cells of *Drosophila w*^{1118} (white eye colour) embryos along with DNA encoding the transposase gene. Microinjected embryos were returned and screened for *P*-element incorporation. To identify those flies containing the *P*-element (or 'transformants'), and to identify their chromosomal location, a series of crosses were performed as described in Figure 2-1 [148]. The resulting numbers of transformants and their chromosomal location in the genome are shown in Table 4-4. For consistency, a single line for each *fru* mutant was chosen on the third chromosome for DamID analysis, all having similar shades of solid orange eye colour.

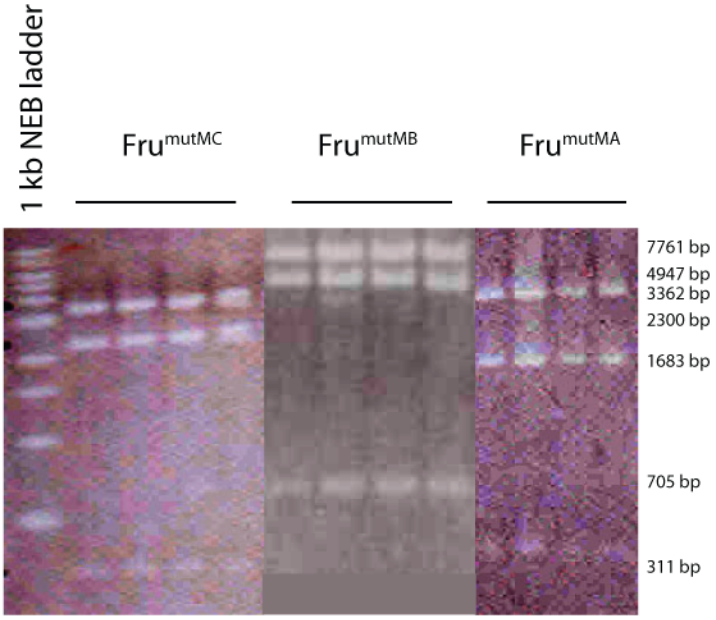


Figure 4-6 Confirmation of 3-way ligation. Once ligated, the mutant plasmids were further digested to confirm the correct integration of all three fragments. A suitable restriction enzyme was identified which is capable of giving a unique digestion pattern to identify correct insertion. For the A (expected sizes 3362 bp, 1683 bp, 311 bp) and C (expected sizes 3362 bp, 2300 bp, 311 bp) mutant plasmids, *XhoI* was used, the B (expected sizes 7761 bp, 4947 bp, 705 bp) mutant was digested with *BamHI*.

Isoform	Total number of lines identified	Number of lines per Chromosome			
		X	2	3	4
Mutation A	26	2	8	14	2
Mutation B	12	0	4	6	1
Mutation C	6	1	2	3	0

Table 4-4 Mutations identified using site directed mutagenesis. Due to the process of microinjection, a random number of transformants are generated per injection. This table records the total number of transformants identified, and their quantity per chromosome.

4.4 DamID

First described by van Steensel and Henikoff [14, 165], DamID avoids the need of a highly specific antibody by fusing the native *E. coli* Dam protein (which methylates $G^{m6}ATC$ sequences) to the protein of interest (Fru) [14]. This system exploits the lack of native methylation in *Drosophila* [165], marking the location the protein binds DNA with methylation which can then be isolated, enabling the identification of genomic regions bound by the gene of interest.

Methylation spreads through the genome, and due to GATC sequences occurring every 200- 300 bp within the genome, regions up to 5 kb are identified [14]. This produces a broad peak to analyse, which must be interpreted differently to that of a ChIP-chip produced peak. Most relevantly, the origin of the binding site cannot be predetermined through the structure of the peak – methylation can occur anywhere near the point of binding dependant on the location of GATC sequences in the genome [14]. For analysis of DamID, a peak is associated with a gene if it occurs within ± 10 kb of a gene [1, 126].

DamID analysis of transcription factor binding has been well characterised in *Drosophila* [1, 22, 166], however most experiments to date have been completed in the embryo [1, 160, 166]. In fact no DamID experiments to date have been completed in pupae, and only a handful have been completed in the larvae or adult time points [165, 167, 168]. For Fru, the embryo is of little interest, as sex-specific Fru protein only begins to be expressed in L3 'wandering' larvae. Ten DamID experiments were completed in total. Experiments for isoforms Fru^{MA}, Fru^{MB} and Fru^{MC} were completed for the larval, pupal and adult time points as well as the DNA binding mutant Fru^{mutB} isoform at the adult time point. Each experiment was completed in triplicate, with one replicate being dye-swapped to eliminate any dye-specific interference.

4.4.1 Creating fusion lines

The NDam-Fru^{MA}, Fru^{MB} and Fru^{MC} lines used for Fru were previously created in the lab. The NDam plasmid consists of five consecutive UAS binding sites, a minimal hsp70 promoter, the Dam methylase (to generate an N-terminal fusion with Fru) and a selection of restriction sites (into which the Dam-Fru sequence

can be inserted). This plasmid was injected into embryos from w^{1118} flies by Genetic Services Ltd (2.1.4), the adults of which were crossed to w^{1118} flies upon return and screened for the presence of red eyes in the progeny. The strength of eye colour relates to the position of the inserted plasmid DNA, ranging from a peach colour through to an orange colour. The same procedure was completed with a plasmid containing Dam only (no *fru* gene) to create an experimental control showing the effect of Dam alone methylation throughout the *Drosophila* genome. The Dam only control was found to be quite toxic, resulting in sickly flies. For this reason, the flies regularly died, and so more than one version of the control line was used during these experiments. In each instance (for both the experimental and control lines) insertions on the third chromosome were chosen.

Each of the three sex-specific *fru* complementary DNAs (cDNA) where fused to the Dam sequence, creating N-terminal Dam fusions (see Figure 4-7). To test for the presence of functional Dam fusion proteins, the fly lines were crossed to an *actin*-GAL4 fly line (expresses GAL4 ubiquitously). The progeny of this cross containing both GAL4 and the Dam-Fru fusions should be lethal (over-expressed Dam causes lethality). All Dam-Fru fusion proteins (Fru^{MA}, Fru^{MB} and Fru^{MC}) were found to be lethal when crossed to actin-GAL4, confirming functional Dam fusions were expressed (Figure 4-7).

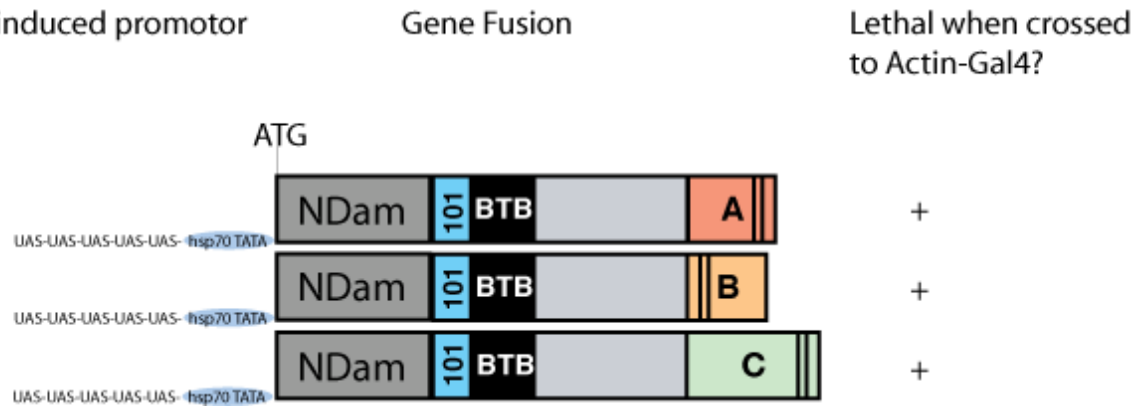


Figure 4-7 Schematic representation of the NDam-Fru^{M} proteins. Three Dam fusion proteins were created representing the three sex-specific transcripts of *fru*. The proteins are shown fused to the Dam protein, which is expressed through the hsp70 promoter. Each product was found to be lethal when crossed to *actin*-GAL4 confirming the presence of the Dam protein.

4.4.1.1 The 'leaky' UAS system

The DamID approach we have taken relies on the 'leakyness' of the UAS system described by Brand and Perrimon in 1993 [22]. These UAS sequences (five in total) upstream of the NDam-*fru* sequences should only be activated in the presence of GAL4. In the case of Dam methylase, activation by GAL4 results in too much methylation by the Dam protein, which causes both excessive non-specific methylation, and toxicity in the fly leading to cell death [14, 160]. A basal hsp70 core *Drosophila* promoter is also included in these constructs to ensure the natural *Drosophila* transcriptional machinery is able to transcribe the sequence down-stream of the UAS. The recruitment of the basal transcriptional machinery to the hsp70 core promoter results in a certain amount of transcriptional 'leakage' in the absence of GAL4, allowing only a small amount of Dam expression, and thus low levels of methylation in the system. This allows just enough methylation to successfully tag the areas where the gene of interest binds [1, 14].

4.4.1.2 Defining the time points

To identify targets of Fru throughout development of the fly the DamID protocol was run at three specific time points. L3 'wandering' male larvae is the definition of the larval time point; this is the onset of Fru expression and is when the larvae have left the food to climb the vial walls to pupate (but before they start to eject spiracles). The peak of Fru expression is 46 hr. male pupae, which is defined by collecting L3 'wandering' larvae, sexing them at this point, and leaving them to incubate at 25 °C for 46 hr. The adult time point defines the period when Fru is maintaining its effects and is defined as 2 days post-eclosion. It should be noted that due to concerns regarding methylation accumulation this point is slightly early in terms of the 'standard' adult time point used for behavioural assays (more often 5- 10 days). Freshly eclosed male flies are incubated at 25 °C for 48 hr. to achieve this time point.

4.4.1.3 Tissue collection

Fru is expressed in both the peripheral and central NS, however as we are interested in the control of behaviour in the CNS, we focused exclusively on this tissue. To ensure the Fru DamID experiments represented only CNS Fru binding, the brain and VNC were dissected for tissue collection as defined in 2.1.5. 45 CNS (brain and VNC) were dissected per replicate for larvae and pupae, and 30 per replicate for adult.

Initial optimisation of the tissue collection was completed is PBS, however DNA was found to be degraded. DamID buffer was optimised, which incorporates EDTA (chelates Mg^{2+}) ensuring the Mg^{2+}-dependent DNase enzyme is unable to

function properly. Subsequently, DamID buffer replaced PBS for dissections, which resolved the degradation problem.

4.4.1.4 DamID protocol

2.5 µg of DNA was required for each experimental replicate which was isolated from pooled dissected CNSs using Qiagen's DNeasy tissue kit following the protocol for cultured cells. The DNA was precipitated by adding two volumes of ethanol and 0.1 volume of 3 M sodium acetate. After spinning the solution at 4 °C for 20 mins at maximum speed, the supernatant was removed and the remaining pellet washed in 70 % ethanol. The pellet was air dried and resuspended in RNase (0.5 µg/ ml in distilled water) to a total volume of 100 µl.

The *DpnI* restriction enzyme digests methylated GATC sequences, cutting out the DNA sequences targeted by Dam or Dam-Fru fusions. 2.5 µg DNA was digested with *DpnI* according to the manufacturer's instructions (NEB) in a total volume of 100 µl. The digest was completed at 37 °C overnight, and a further 0.5 µl *DpnI* enzyme added the following morning for an hour to ensure complete digestion. Following this, *DpnI* was inactivated by heating to 80 °C for 20 mins. For each sample, a *DpnI* negative control sample was also created substituting *DpnI* with distilled water.

The products of the *DpnI* digestion were purified using Qiagen's PCR purification columns and eluted in 30 µl distilled water. Adaptors were ligated to the sticky ends, enabling PCR amplification (see below). The Adaptor (AdR) was created by annealing two primers AdT and AdB. The primers were mixed in equal concentrations, vortexed, placed in just boiled water and left to anneal slowly as the water returns to room temperature.

The purified DNA samples were separated into two Eppindorfs, producing four samples total per replicate: a *DpnI* (+), Ligase (+) sample; *DpnI* (+), Ligase (-) sample; *DpnI* (-), Ligase (+) sample; and a *DpnI* (-), Ligase (-) sample. To each sample 4 % AdR adaptor was added with T4 ligase as described in the manufacturer's instructions (Roche) in a total volume of 20 µl. In the case of ligase (-) samples, T4 ligase was substituted with distilled water. The samples were incubated for 2 hrs. at 16 °C to allow the adaptors to ligate fully. The ligation reaction was terminated by heating the sample to 65 °C for 10 mins. These adapters are required for the amplification stage of the DamID protocol, where individual unknown fragments are amplified using a generic primer set designed to prime from the adapter sequence, allowing the DNA to be amplified to a level suitable for use in conjunction with a tiling array.

Remaining non-ligated DNA was removed by digestion using *DpnII*, which digests non-methylated GATC sequences thus removing non-targeted DNA. *DpnII* was added to all samples according to manufacturer's instructions in a total volume of 80 µl (NEB). The samples were digested at 37 °C for 1 hr.

At this point, the only DNA remaining was that defined as being targeted by Dam or Dam-Fru fusions. Prior to sending samples to Nimblegen for processing, the DNA was PCR amplified using primer Adr-PCR as defined in 2.4.11. Adr-PCR was added to 20 µl of each sample along with PCR advantage enzyme mix according to the manufacturer's instructions in a total volume of 80 µl (Clontech, Paris). The samples were cycled according to Table 4-5. Resultant samples were confirmed by gel diagnostic according to 2.4.3. If DamID has been successful only the *DpnI* (+), Ligase (+) sample should contain DNA, which should contain a smear of fragments averaging around 500 bp in length (Figure 4-8). The

amplification stage could be repeated up to 4 times to amass enough DNA to send to Nimblegen.

Finally, PCR samples were pooled and purified using Qiaquick PCR purification kit (Qiagen) and eluted in 25 µl distilled water and quantified using a Nanodrop. Samples must contain 6.5 µg DNA/ replicate, for both the experimental and control sample. These samples must be larger than 200 nucleotides in length, be above 250 ng/ µl in nuclease free water or 1x TE buffer and have an $A_{260}/ A_{280} \geq 1.7$, and an $A_{260}/ A_{230} \geq 1.6$ to be accepted by Nimblegen for further analysis.

Step	Temperature (°C)	Time (mins)
1	68	10
2	94	1
3	65	5
4	68	15
5	94	1
6	65	1
7	68	10
8	Go to step 5	Repeat 3x
9	94	1
10	65	1
11	68	2
12	Go to step 9	Repeat 17x
13	4	Until ready to continue
14	END	

Table 4-5 PCR cycling programme for DamID. To amplify enough DNA to be used for identifying the fragments on a tiling array, the samples were cycled through the above protocol which was designed by Dr. van Steensel.

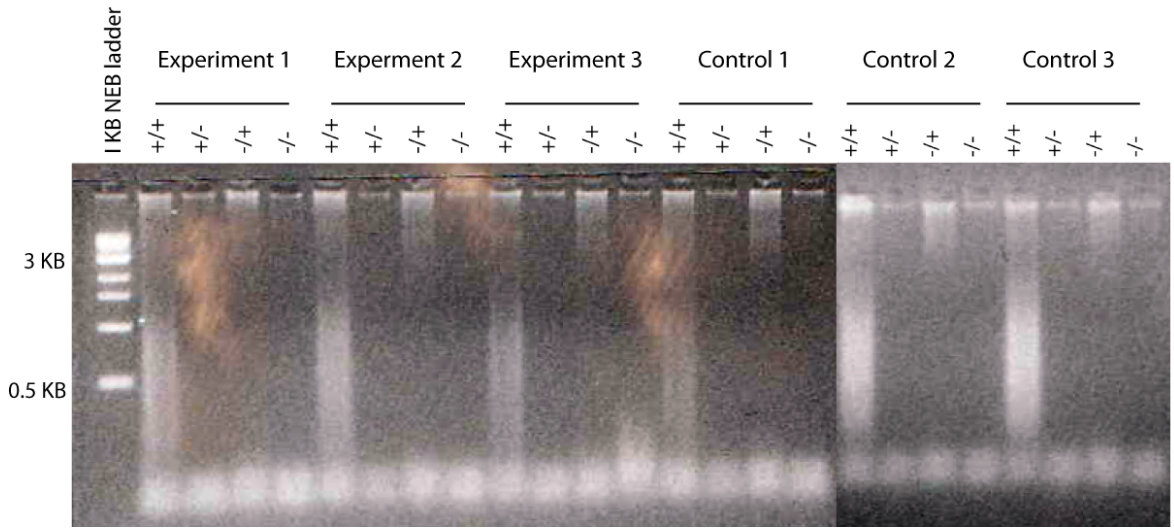


Figure 4-8 Examples of DamID samples prior to analysis by Nimblegen. Prior to this point, samples were digested with *DpnI*, to which adapters were ligated to the resultant fragments. *DpnII* digested non-adapter linked fragments, leaving only 'targeted' DNA whose fragment size was on average 500 bp in length. Upon purification and concentration, the fragments are visualised above on a 2 % agarose gel. +/+ represents the *DpnI*, adapter-ligated sample. The remaining three samples per experiment are controls: +/- represents *DpnI*, but no ligase; -/+ represents no *DpnI* digest, ligase adapted; -/- represents no *DpnI* digest, no ligase.

The samples were posted to Nimblegen where they are labelled with Cy3 or Cy5 fluorophores and washed over a 2.1 M (2.1 million probes per array) tiling array which incorporates the entire *Drosophila* genome. The exception being Fru^{MC} pupae, which was the first experiment to be completed, and was done so on a customised tiling array [14]. For the 2.1 M arrays, probe length was 45 bp, for the custom array they were 60 bp. In each instance, one of the three replicates was dye swapped, i.e. Experiments one and three are labelled so that the experimental sample was labelled with Cy3, whereas experiment two had labelled the control sample with Cy3 – this accounts for potential discrepancies between the floroflours. Again, the exception was the Fru^{MC} pupae samples, which were completed in duplicate, one of which was dye swapped. Where the samples complement the tiling array probes, a florescent intensity was recorded. This data was returned as a set of files, including .gff files for visualisation of the results against the genome – where the control samples had been subtracted from the experimental samples (and the data expressed as a $\log_2$ median normalised ratio

centred and scaled around zero by means of subtracting the bi-weight mean), and .PAIR files which provided intensity data for both the experimental and control samples.

To determine meaningful conclusions from these datasets it was important to carefully normalise the data, removing false results and identifying a reliable set of targets for each isoform at each developmental time point, before proceeding to bioinformatic analysis.

4.5 Normalising the data

Two types of files are provided from Nimblegen: .PAIR files contain the per probe raw data for both the control and experimental arrays for each replicate, and .gff files contain track information showing the merged control and experimental arrays. Our collaborator, Dr. Tony Southall (Brand Lab, Cambridge) performed two initial normalisations. For each experiment, Dr. Southall merged the replicate .gff files to create an averaged file incorporating all three replicates into a single track, and performed moving average smoothing of the track with a window of 800 bp to create clean peaks (Figure 4-9). The second file produced was an excel file containing a list of all peaks and their associated genes with a FDR (four or more consecutive probes of two-fold increase or more, 100 iterations) and location of the peak within the chromosome (where a ‘peak’ is assigned to a gene where six or more consecutive probes, seen in two or more datasets, within five kb up or downstream of the gene is identified. These files form the basis of all further analysis.

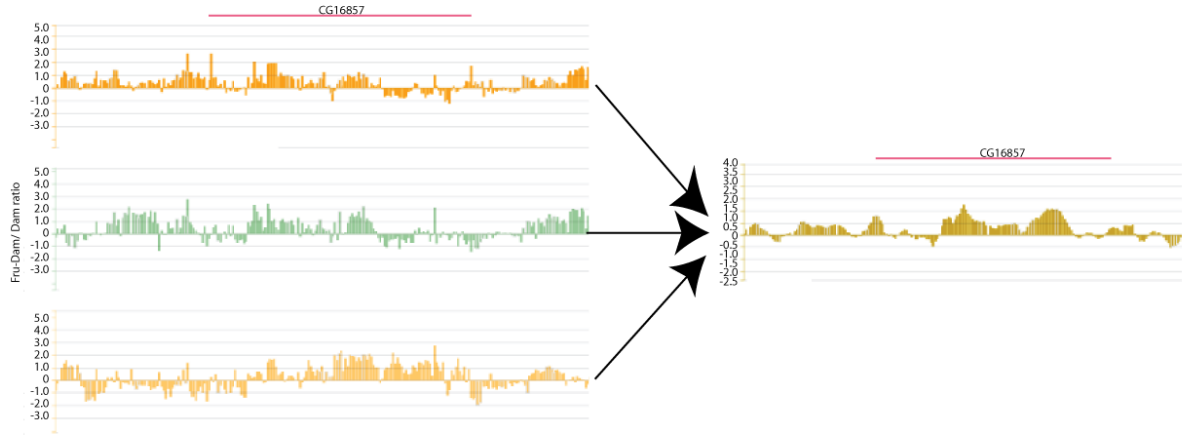


Figure 4-9 Creating a moving average file. Merging the three replicates into an average file, based on individual probe values, created a moving average file. The merged file was then smoothed using an 800 bp window, by taking into account two probes either side of the probe in question, and averaging the values. Comparatively (individual tracks on the left compared to the moving average file on the right), the averaged file shows a much clearer interpretation of the image. (Images are screenshots from Signalmap, perl scripts used to create these files were created and implemented by Dr T. Southall).

It was imperative that these initial normalisation methods were implemented carefully, as overly restricting the data or under-restricting the data at this point can have an adverse effect throughout the subsequent bioinformatics analysis steps. If the data is overly restricted, true targets may be discounted. This would cause increased significance to be put on the remaining targets and further some targets might never be identified. Conversely, by under restricting the datasets, false targets are retained. Though unhelpful in the long-term, over-restricting the data has a higher impact at this point in the analysis, and so the datasets used for subsequent analysis are either derived from the moving average (and thus contains all the original dataset) or are derived from excel files which are restricted by an FDR of 0.01, which accounts for approximately 5 % of each dataset. It should be noted that the natural variation of such a dataset has been shown to be 6.7 % and thus the stringency of analysis taken for this data should therefore be considerably less [169]. Although a significant number of data are discarded by this FDR, it allows a suitable balance between the retention of false

values and the need to restrict the number of targets per dataset to that which is a suitable number to analyse.

4.5.1 Moving average

Each dataset is in triplicate. DamID is generally highly reproducible and so with the exception of two cases, a moving average file was created which takes into account the result for each probe in each replicate, and smoothes the data by considering each averaged probe with relation to its neighbours. This analysis was completed by Dr T. Southall using original Perl scripts. The two exceptions to this were the Fru^{MC} pupae dataset, which as a proof of concept experiment was only completed in duplicate, and the Fru^{MB} larvae for which one replicate was discarded due to a clear discrepancy between it and the remaining two replicates which was attributed to a failure at some experimental stage (see 4.6 below).

4.5.2 Peak finding

Averaged datasets were run through a Peak Finding Perl script (Dr. T. Southall) to identify all peaks. These lists were compared to the genome, and those peaks associated with a known gene identified and isolated. All future analysis performed originated from these gene lists, or the moving average files made alongside them.

4.6 Discarding a replicate

Working in triplicate gives the option to discard a replicate that has at some point being incorrectly processed. In the case of larval Fru^{MB}, one dataset (not the dye swap – which is not viable for discarding, as this would suggest the dyes have an

effect on the experiment) was clearly incorrect in comparison to the other two replicates and, in fact, seemed to be almost a direct inversion (see Figure 4-10). For the purposes of the moving average file, this replicate was discarded to prevent its inaccurate results distorting the averaged files.

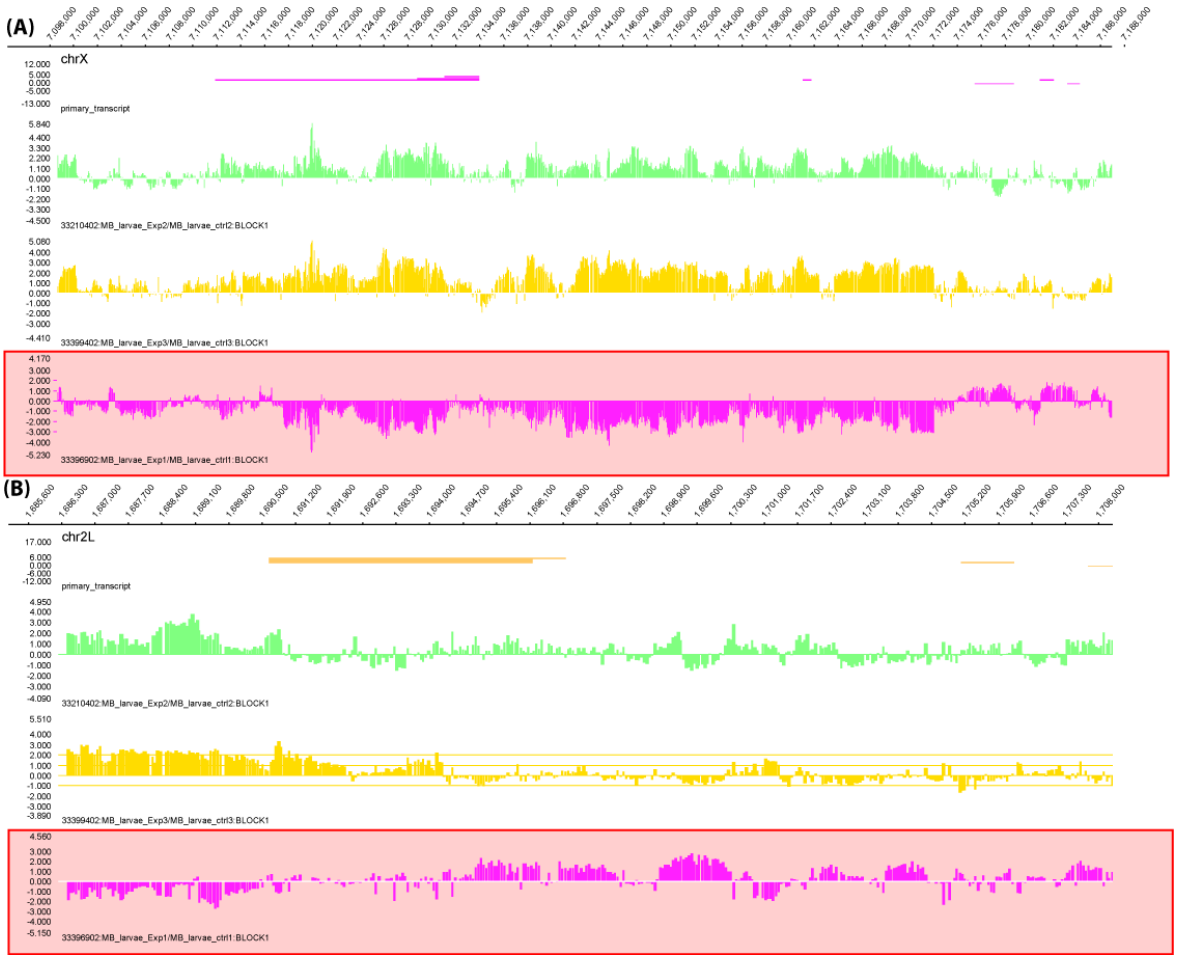


Figure 4-10 Discarding a replicate. DamID is generally highly reproducible, however in some instances one replicate can fail. In one instance, the Fru^{MB} larvae dataset, one of the three replicates was extremely unrepresentative of the other two replicates. As this wasn't the dye swap replicate (which may have indicated an artefact of the dye) this replicate was considered to be rejected. (A) Example of the unrepresentative replicate on chromosome X. The pink shaded box shows the outlier replicate, which appears to be almost a mirror image of the other two replicates. (B) A second example of the unrepresentative replicate on chromosome 2L. Again, the replicate is shaded in pink, and shows very different intensity mapping compared to the other two replicates. The Y scale on this image is Dam-Fusion/ Dam-only control, with the Y scale being distance through the genome in base pairs.

4.7 Discussion

The main objective set out in this chapter was the generation of ten datasets representing the downstream genomic targets of Fru isoforms throughout

development. From just the raw, normalised data discussed in this chapter, it is clear there are isoform and developmental specific differences between datasets– Isoform Fru^{MA} appears to be much more specific, binding fewer targets. This is in keeping with the characterisation of Fru^{MA} expression (Chapter 3.3.1) where Fru^{MA} has a far more restricted expression pattern when compared to Fru^{MB} and Fru^{MC}.

The process of creating site-directed mutations required the sub-cloning of *fru* isoform specific fragments into smaller plasmids due to the restrictions of the Stratagene kit used. However, upon successful reintegration into the *p*UAST-NDam plasmid, multiple lines for each isoform were isolated. DamID for the mutant Fru^{MB} isoform (Fru^{mutB}) at the adult time point was expected to produce one of two outcomes: a tiling array containing only background, indicative of no protein-protein interactions; or a tiling array showing peaks and thus evidence of protein-protein dimerization and/ or co-factor binding; the latter proved to be the case.

The peak finding process was perhaps the first point at which results could vary, depending on the stringency of the conditions chosen when processing the datasets. One could overly restrict the results, however this was avoided so as not to discard true genomic targets. This was perhaps the most important step, as a differing dataset at this point can dramatically alter future whole genome analysis, such as gene ontology.

Recently, many papers have suggested that transcription factors have the potential to bind their cognate binding site whenever and wherever it is available, thus specificity is brought about by the location and timing of transcription factor expression [170]. As a result, it is suggested that transcription factors in fact bind

up to thousands of genes strongly, and could bind tens of thousands more weakly but invariably do not due to the epigenomic state of the DNA; for example the presence or absence of other transcription factors having stronger affinities challenging the binding site region, or changes in the state of the chromatin that affect accessibility [169]. By sampling a series of time points, the intention is to show that Fru binds some genes transiently and perhaps others consistently throughout development. These results will help us piece together for the first time how Fru isoforms function at the level of DNA-binding to generate the diverse range of phenotypes associated with this gene.

5 DamID Analysis

5.1 Introduction

In the previous chapter, DamID experiments were carried out for three isoforms of Fru at three developmental time points, and for one DNA binding mutant at the adult time point. The intended outcome from these experiments was to identify specific downstream target genes of each Fru isoform in the NS throughout development. As a vast quantity of genome-wide data was generated using the DamID technique, this chapter describes the bioinformatic analyses used to process this data: analysing summary statistics; flymine analysis (flyatlas, chromosome distribution, protein domain analysis); gene ontology and gene list analysis. From these analyses and visualising the datasets, candidate genes were identified and further analysed in an attempt to define a direct genomic target (MOL analysis, co-staining using antibodies, motif analysis).

There are many approaches to analysing whole genome data. Fundamentally the difference is whether your analysis is bioinformatically driven (mining available data, using scripting programmes to analyse gene ontology, and patterns within the data) or functionally driven, identifying a series of candidate genes within the available datasets for analysis based on a predicted functional connection with Fru (whether that be their relationship to Fru identified by previous data, or the ability to be further analysed due to the existence of suitable tools) [2, 39, 79, 171]. To best analyse such a quantity of data, I have used a combination of these two approaches to identify candidate genes for further analysis. A number of bioinformatic approaches were used to categorise the data and identify a gene list that comprises genes unique to a specific time point or isoform, genes common across many time points or isoforms, or individual genes that stand out as being relevant in the downstream processing of Fru signals (Figure 5-1).

Recent studies have suggested that transcription factors are less specific than previously thought [169]. As sequence-specific transcription factors are known to bind relatively short and often highly degenerate DNA sequences [172], transcription factors are therefore capable of binding (and in fact do bind) thousands (or even tens-of-thousands) of sites within the genome [169]. Taking this into account, it is highly likely that the binding-sites identified in the previous chapter all represent true Fru-DNA binding. A recently proposed "Continuous Network" model of transcriptional networks is based on recent advancements in the studies of transcription factor binding on a genome-wide scale [173]: this model suggests that each transcription factor binds to almost all available sites *in vivo* "over a quantitative continuum of occupancy levels". Based on this model, the most highly occupied sites (represented by 'strongest binding') represent the biologically significant targets. The low occupancy sites (or 'weak binding') sites are therefore less likely to be biologically significant. In this chapter, candidate genes identified and further characterised are likely to represent the 'strongest binding' of Fru, and are therefore most likely to be biologically relevant in relation to Fru function [2, 34, 39, 79].

5.1.1 Aims and objectives

The aim of this chapter was to produce a reliable list of target genes for each dataset, and to identify a small cohort of genes for initial further analysis (in the region of ten genes). Further, to perform basic analysis of these identified genes to confirm their co-expression with Fru, and where possible attempt to elucidate their biological function as a downstream target of Fru.

These aims were completed through a series of processes to ensure each dataset was treated in the same way. Figure 5-1 outlines the protocol designed for dataset analysis.

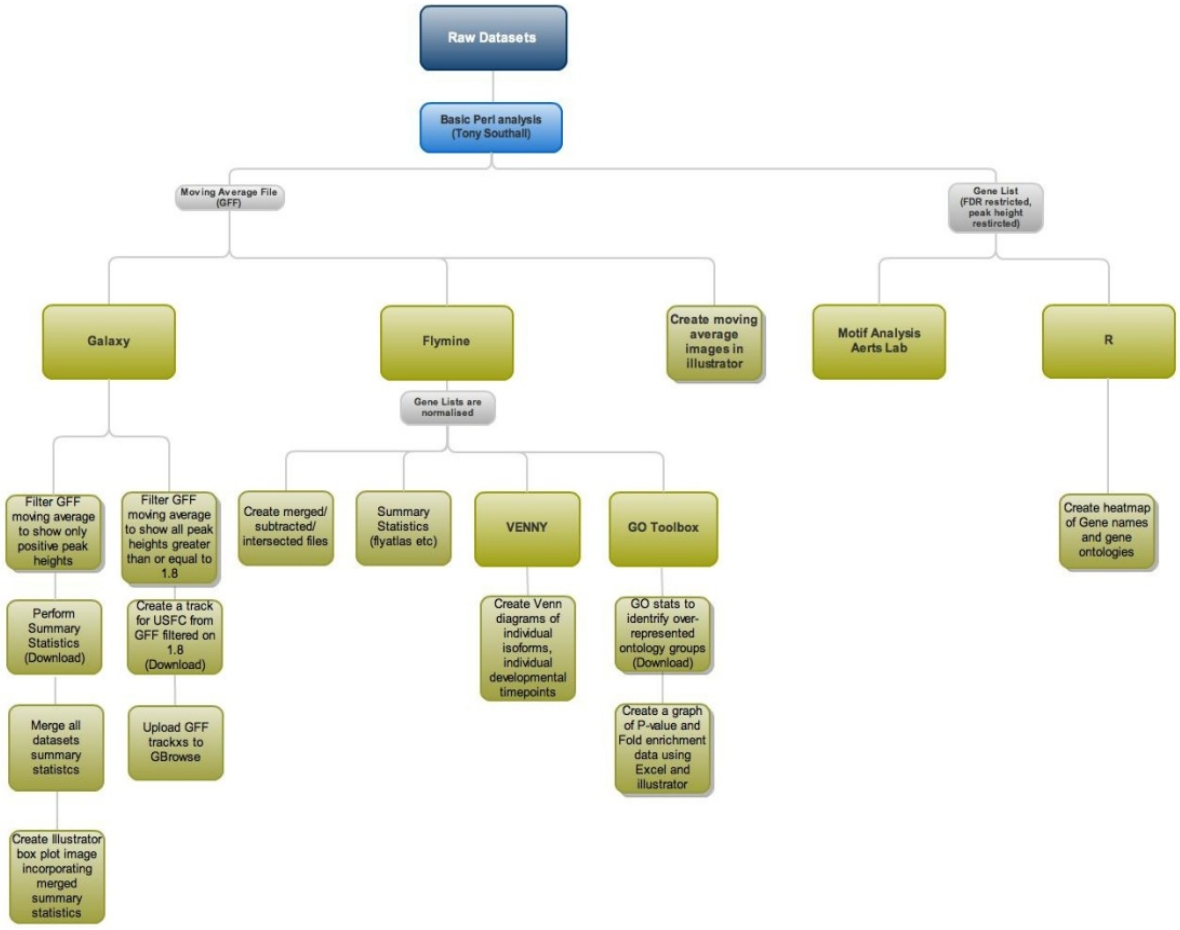


Figure 5-1 Flowchart depicting bioinformatic analysis of Fru DamID datasets. Processing of datasets in the correct order is very important to define true targets of Fru. The flowchart shows how information was processed, and from what original data subsequent files were derived.

5.2 Comparison of the datasets

To compare the spread of each dataset, box plots of the summary statistics were compiled using Excel (Microsoft), as seen in Figure 5-2. It was important to identify a number of targets from which further extrapolations could be made, without concern that too many 'true' targets had been discounted. Upon observation of the spread, it became clear that 5 %, a commonly used cut-off [174], was a suitable number to work with for the datasets. Therefore, a threshold

Fru-Dam/ Dam value of 1.8 (approximately 5 % for each dataset) was taken as an initial cut-off, in particular, when identifying a 'peak' from the raw data.

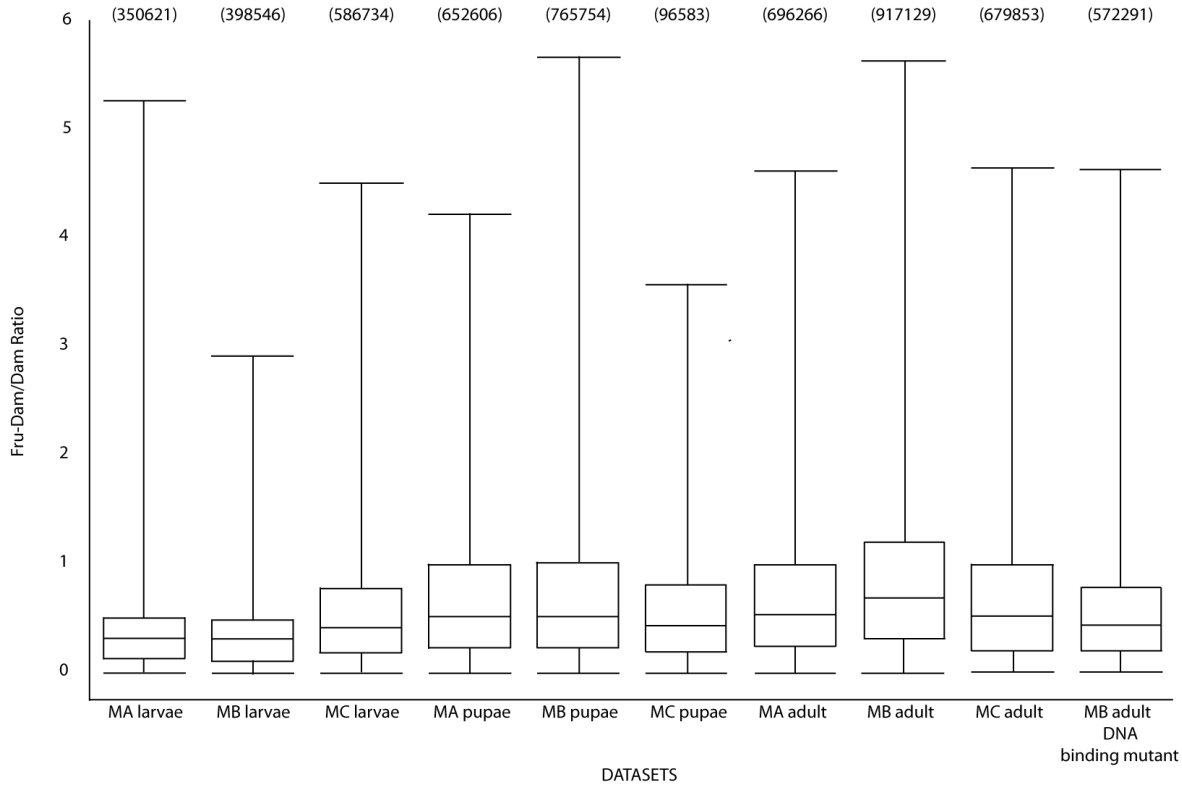


Figure 5-2 Boxplots representing the DamID datasets, and their spread. The DamID datasets are different in their spread of Intensities (Fru-Dam/ Dam ratio). Further, the size of each dataset differs greatly. By depicting the data in such a way, the similarities can be seen – and the size by which the sample should be restricted identifiable. The remaining numbers of genes given this cut off were listed in Table 5-1.

5.3 Flymine

To normalise all subsequent analysis and prevent discrepancies between gene nomenclatures, all isolated gene lists were analysed through Flymine (www.flymine.org) [175]. Flymine initially provides additional information for the gene lists including various identifiers: Flybase; CG number; gene symbol; full length gene name. Beyond standardising the datasets, Flymine provides a series of basic analyses, which interpret the spread of the data and provide a platform for future analysis. More generally, Flymine can be used to identify patterns in the targets, such as common functions, isoform specific functions and bias in the data.

5.3.1 Defining gene lists

The number of genes in each dataset after Flymine normalisation is shown in Table 5-1. The number of genes expressed in the larval time points were considerably fewer compared to the other developmental time points considered, which is consistent with our prior knowledge of Fru^{M} expression in the CNS: the onset of expression starts at the larval L3 phase and peaks at 48 hrs. into pupation [81]. The Fru^{MA} isoform had few genes targeted in the larvae, but by 46 hrs. pupae showed over a four-fold increase in targets, rising to over a five-fold increase by 48 hrs. into adulthood. Fru^{MB} was seen to bind more targets – albeit these datasets do not distinguish between strength of binding. It may be that Fru^{MB} has a more generic binding motif and is capable of binding more targets, but that many of the interactions represent low occupancy sites [173].

Dataset	Time point	Number of genes
Fru^{MA}	Larvae	409
	Pupae	1827
	Adult	2103
Fru^{MB}	Larvae	1756
	Pupae	2147
	Adult	2059
Fru^{mutMB}	Adult	1377
Fru^{MC}	Larvae	344
	Pupae	283
	Adult	1861

Table 5-1 Number of genes identified per dataset. Each dataset produces a different number of identified targets. In most instances the number is likely a reflection of the level of expression of each Fru isoform at each time point – the exception being Fru^{MC} pupae, which used a different tiling array with a sparser layout, which may account for a decrease in sensitivity. Original gene lists can be found in Appendix 1.

Interestingly, the Fru^{MB} isoform mutant (Fru^{mutMB}) dataset has a large number of genes in it indicating that many of the Fru^{MB} isoform interactions might occur through homodimerisation (Fru^{mutMB}/ Fru^{M}) or heterodimerisation with a co-factor or other Fru^{M} isoforms. The Fru^{MC} isoform appears to have relatively small numbers of targets at both the larval and pupal time points. This was to be expected at the larval time point, however was less expected for the pupal time point (the peak of Fru^{M}

expression). This might be explained by the different tiling array layout used for this experiment (which was completed prior to the release of the 2.1 M tiling arrays): the larger spacing between the oligos on these arrays resulted in decreased coverage of the genome, leading to less accurate capturing of phototopes and thus decreased numbers of genes recorded. To attempt to elucidate the reasoning for these differences between the transcripts targets, patterns between the datasets were investigated.

5.3.2 Chromosome distribution is even throughout development

Flymine allows a gene list to be compared to the expected spread of genes across the genome, highlighting any discrepancy or bias toward or away from a given chromosome. In all instances, the DamID datasets showed no chromosomal bias, being spread across the genome as predicted (see actual vs. expected in Figure 5-3).

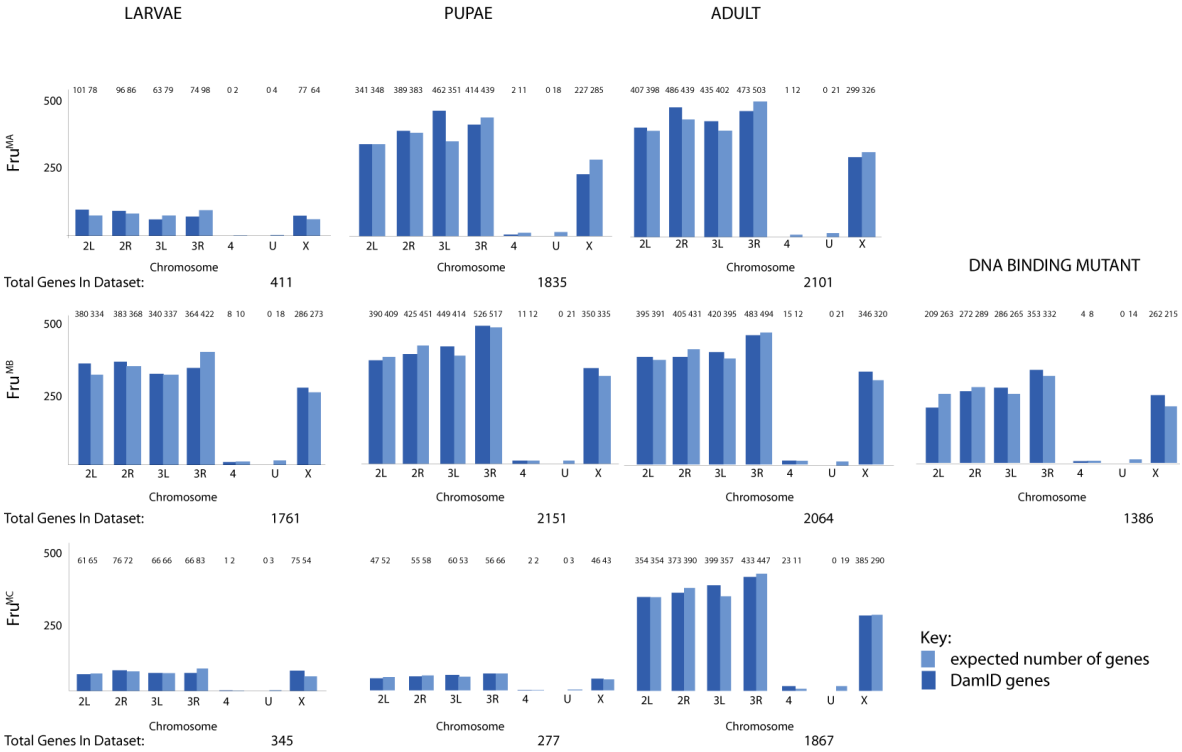


Figure 5-3 Chromosome distribution of DamID datasets. Distribution of the number of genes per dataset over the chromosomes in the genome. Potential bias towards any one chromosome is depicted. Datasets are scaled to be comparative, with each row representing a different isoform: row 1 shows Fru^{MA} data, row 2 shows Fru^{MB}, and row 3 shows Fru^{MC}. By column, the data is shown by developmental time point (left to right): larvae (L3 'wandering'), pupae (46 hr.), adult (48 hr.), DNA binding mutant (adult, 48 hr.). Dark Blue is the DamID data, Pale blue is the expected number of genes per chromosome.

5.3.3 Tissue specific expression of Fru *targets*

Flyatlas (www.flyatlas.org) is a tissue-specific affymetrix-derived map of gene expression in various sub-groups of tissues throughout the adult and larvae [176]. Flyatlas is used to analyse whole lists of genes and display the numbers of these genes up or down regulated in the organism). The Fru-target gene lists identified using DamID were analysed with this tool. Each gene was called using the following settings: present in 3 out of 4 arrays, enrichment greater than 2.0 and having an mRNA signal greater than 100 when compared to a matched whole-fly sample.

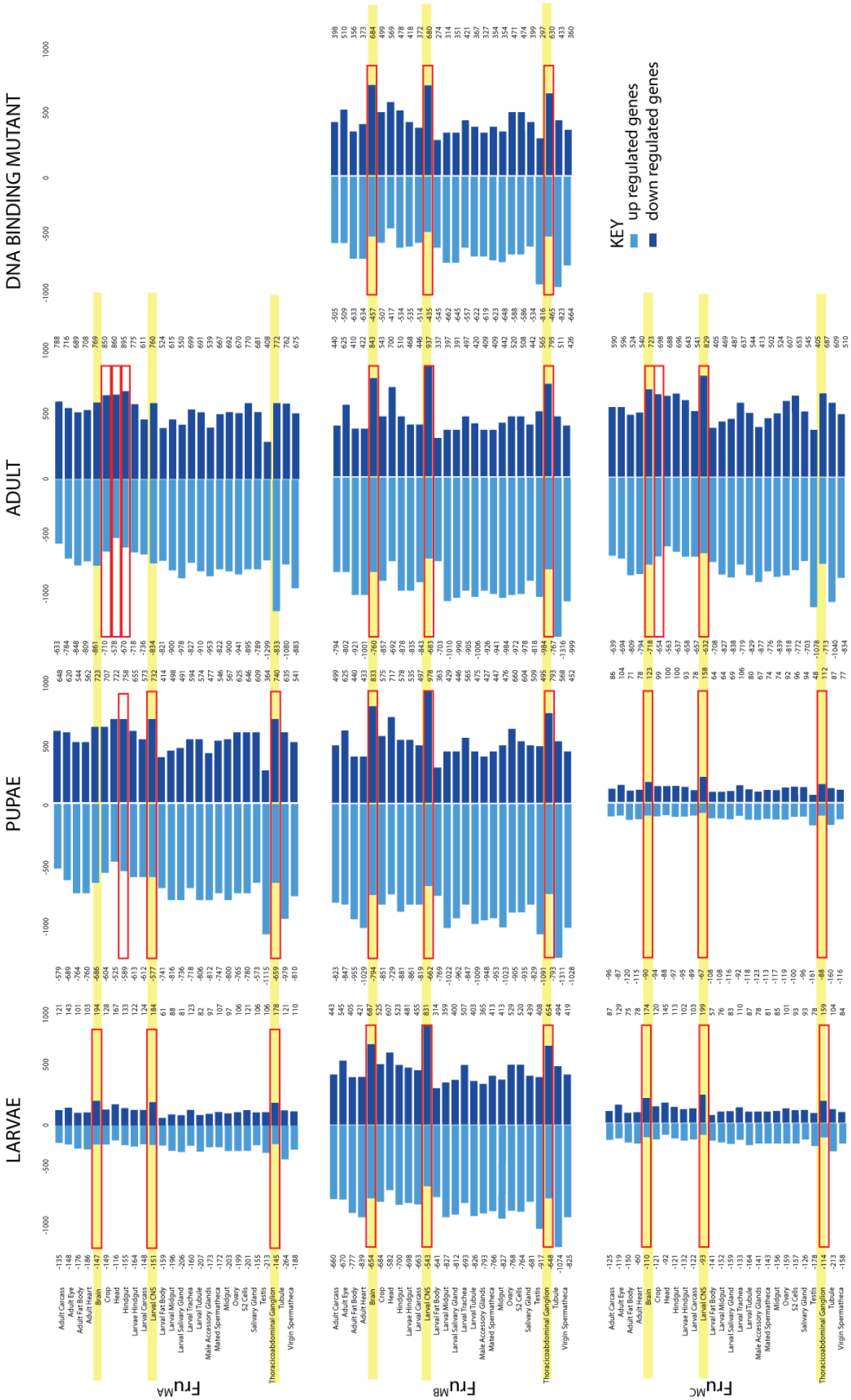

DNA BINDING MUTANT
ADULT
PUPAE
LARVAE
KEY
up regulated genes
down regulated genes
FruMA
FruMB
FruMC

Figure 5-4 Tissue specific expression of Fru target gene. Flyatlas widget representation of the tissue specificity of expression of Fru-target gene lists. In all instances the top three gene expression locations (indicated by a red box) are those associated with the nervous system: the head, the CNS and the VNC for larvae and pupae, however the crop is also represented in the adult time point. This is also the case for the DNA binding mutant, showing any indirect binding also follows this pattern. The nervous system (Brain, thoracicoabdominal ganglion, and larval CNS) is highlighted in yellow.

The larval time point was quite distinctive in that in each isoform the three same tissues represented the top three up-regulated tissues: brain; larval CNS; thoracicoabdominal ganglion (equivalent to VNC), highlighted in yellow in Figure 5-4. These top up-regulated tissue groups carried through to the pupae in Fru^{MB} and Fru^{MC}, and in fact throughout development in the case of Fru^{MB}, suggesting a strong role in NS development.

In the pupae the differences between the targets of the isoforms become apparent. Whereas both Fru^{MB} and Fru^{MC} were consistently associated with upregulated CNS groups, Fru^{MA} started to show differences, moving from highly upregulated brain genes, to highly upregulated genes associated with the hindgut. Further, by the adult time point Fru^{MA} contained genes highly upregulating in the head, crop and hindgut, a clear deviation from the strongly CNS related larval time point, suggesting that through metamorphosis and into the adult, the Fru^{MA} isoform has less association with CNS developmental genes, and perhaps moves towards a role driving physiology.

Throughout development, Fru^{MC} and Fru^{MB} showed similarity in their upregulated tissue groups, suggesting a potential overlap in their function. Fru^{MB} datasets are the largest, and are strongly associated with NS tissues throughout. The Fru^{MC} datasets are smaller and deviated slightly in the adult time point, showing a shift from thoracicoabdominal ganglion upregulated genes towards upregulated crop genes.

5.3.4 Protein domain analysis shows high homeodomain and immunoglobin enrichment

Each gene list was analysed for protein domain enrichment, as seen in Figure 5-5. This analysis looks for the enrichment of all InterPro protein domains (Data Source: UniProt) within the gene lists. The count is the number of times the domain appears for each gene in a given dataset. Data was correlated using the Bonferroni method, which takes into account multiple comparisons and maintains familywise error rates [174]. The p-value is the probability that a result occurs by chance, thus a lower p-value indicates greater enrichment.

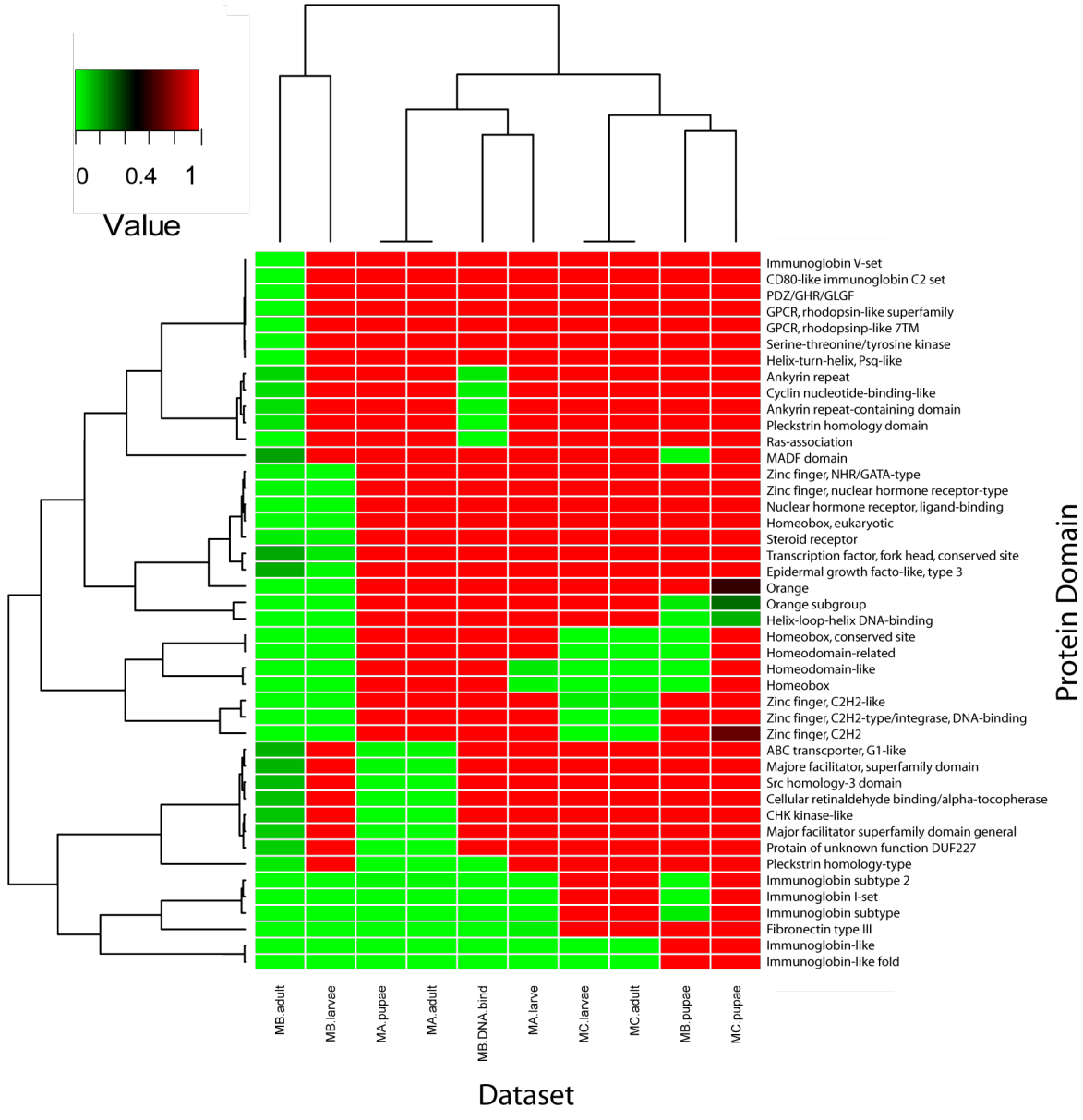


Figure 5-5 Protein domains enriched across the DamID datasets. A heatmap was used to show the enrichment of protein domains across the different DamID datasets for Fru. In the key, value is the p-value for each enriched domain, where a value closer to 0 is more significant. Dendrograms to the top and left of the heatmap show the relationships between the datasets (top) and protein domains (left). In all instances, *p-value*s shown are all less than 0.05, with the exception of Fru^{MC} pupae which showed no enrichment of domains at this stringency, and so to be able to compare this dataset with the others, the stringency had to be loosened.

All Fru^{MA} datasets showed immunoglobin domain enrichment: as the top five enrichments in the larval time point (peak p-value 5.16×10^{-5}, accounting for 88 gene targets); the top enrichment in pupae (3.00×10^{-9}, 49 gene targets) and four other positions in the top ten enrichments. In the adult time point, immunoglobin domain enrichments were seen as the second and third most enriched domains (the first being an unknown function) with p-values of 2.71×10^{-7} and 2.57×10^{-4} respectively, and having 49 and 28 genes in each group. Immunoglobins are known to have a regulatory roles in cell-cell communication and recognition [177]. Datasets for Isoforms Fru^{MB} and Fru^{MC} showed some enrichment for immunodomains (especially Fru^{MB} adult), however were more strikingly enriched for homeodomains.

Homeodomains are DNA binding domains common to transcription factors. For Fru^{MB}, homeodomains were the top four enriched domains for larvae, having a significant p-value of 5.38×10^{-25}, and encompassing 246 genes. Similarly, the top four domains remained the same at the pupae (albeit shuffled) with a significance of 2.11×10^{-13}, encompassing 253 genes. In the adult time point the third and fourth most enriched domains were homeodomains (p-values of 1.11×10^{-23} and 7.60×10^{-23}, 70 and 54 genes respectively) behind two immunoglobin like domains (3.8×10^{-28} and 5.07×10^{-27}, 67 and 79 genes respectively). Interestingly the 12^{th} enriched domain for the pupae is the C_2H_2 ZnF suggesting a potential for co-factor involvement. Comparatively, Fru^{MC} had four homeodomains in the top six enriched protein domains with a peak significance of 1.38×10^{-8} and

encompassing 71 genes. By the pupae time point, no homeodomains were in the top enrichments, and those that were enriched were a range of BTB and zinc finger domains. By the adult time point, homeodomains were highly enriched again, and covered the top four positions (p-values between 5.19×10^{-15} to 1.83×10^{-9}, encompassing 168 genes). In both the larvae and pupae for isoform Fru^{MC} the C_2H_2 ZnF domain was present at enrichment seven and four respectively. Often, transcription factors regulate other transcription factors, leading to cascades of genes regulating a function. Enrichment of similarly structured proteins could be indicative of this.

The DNA binding data for this analysis showed high enrichment for immunoglobin-like domains, which fill the top five enrichment positions. This suggests that the immunoglobin-like domains enriched in the adult Fru^{MB} data may well reflect the interaction with, and DNA-binding of, a co-factor. This further suggests that the homeodomain enrichment seen with the non-mutated Fru^{MB} data is due to direct Fru^{MB}-DNA binding.

This data reinforced similarities between the Fru^{MB} and Fru^{MC} isoforms, and similarly the apparent separation of these isoforms and the Fru^{MA} isoform.

5.3.5 Pathway enrichment

The KEGG database integrates genomic, chemical and systematic functional information through a series of categorisation, linking experimental data to higher level systemic functions to create pathway maps [178]. Reactome by comparison annotates pathways by cross-referencing proteins, genes, small molecules, primary research and GO to model pathways according to nucleic acids, small molecules, proteins and a range of other processes [179, 180]. The obvious limits

for both databases is that they, like many other programmes, define their results on human derived data and thus may be less relevant to fly data, however in lieu of a database designed specifically for the fly, they provide some relevant insights into pathway enrichments in the DamID datasets [179].

Using KEGG and Reactome data, each dataset was analysed for pathway enrichment (Bonferroni correction) [178-180]. In the larval time point for Fru^{MA}, axon guidance was the top pathway enriched (9.27×10^{-2} p-value). By the pupae time point the enrichment showed a shift to signalling and regulation which progressed to regulation of metabolism and signalling by the adult time point.

The Fru^{MB} isoform showed strong pathway enrichment for axon guidance, *robo* interactions (known to be involved in neuronal development) [181], wnt signalling and other developmental pathways, which remained throughout pupae development (with axon guidance becoming the most enriched, p-value 1.05×10^{-4}). By the adult time point, axon guidance was still high in the enrichment list, however it had been overtaken by signalling pathways, suggesting a shift from developmental functions to physiological functions used in the adult fly. Comparatively, axon guidance still featured in the DNA binding mutant dataset, along with signalling events, suggesting these interactions are potentially defined by co-factor binding or Fru- Fru heterodimer binding.

No significant pathways were enriched in the Fru^{MC} isoform dataset at the adult time point. At both the larvae and pupae time points, the enriched pathways were mainly metabolism/ signalling enrichments. This was unexpected in terms of the lack of NS related enrichments, however both of these datasets are very small compared to other datasets and so might not be very representative in generic

analyses. For all the Fru^{MC} isoform analysis, the p-values were barely significant (being in the realm of $p = 0.1$).

5.4 Comparing significant targets between the datasets

The number of putative downstream targets from each dataset makes it difficult to easily determine their similarities and differences. To aid in these further analysis, R was used to create heat maps comparing genes or additional properties of the datasets (Figures 5-5 and 5-6). The dendrograms on the top of each figure represents how similar each dataset is to each other.

A heat map was generated to compare the gene lists between the datasets, using peak heights of ≥ 1.8, the genes were then ranked based on FDR, by colour (Figure 5-6, original data: Appendix 2). Looking at the dendrograms at the top representing similarities between the dataset, it is clear that there is much more commonality between the Fru^{MB} and Fru^{MC} isoforms. In contrast, the pupae and adult Fru^{MA} isoform datasets derive from a completely different stem to the other datasets, showing their distinctive composition. This split from the Fru^{MA} larval dataset suggests a distinct change in the type of genes targeted from the pupae time point onwards, adding to evidence that from the pupae time point Fru^{MA} proteins have a different role, potentially physiological in nature. All the other datasets stem from the same origin, suggesting a strong overlap in the genes that they target. In particular the Fru^{MB} datasets (excluding the DNA binding mutant) are closely related, showing little variance in transcriptional targets for this isoform throughout development.

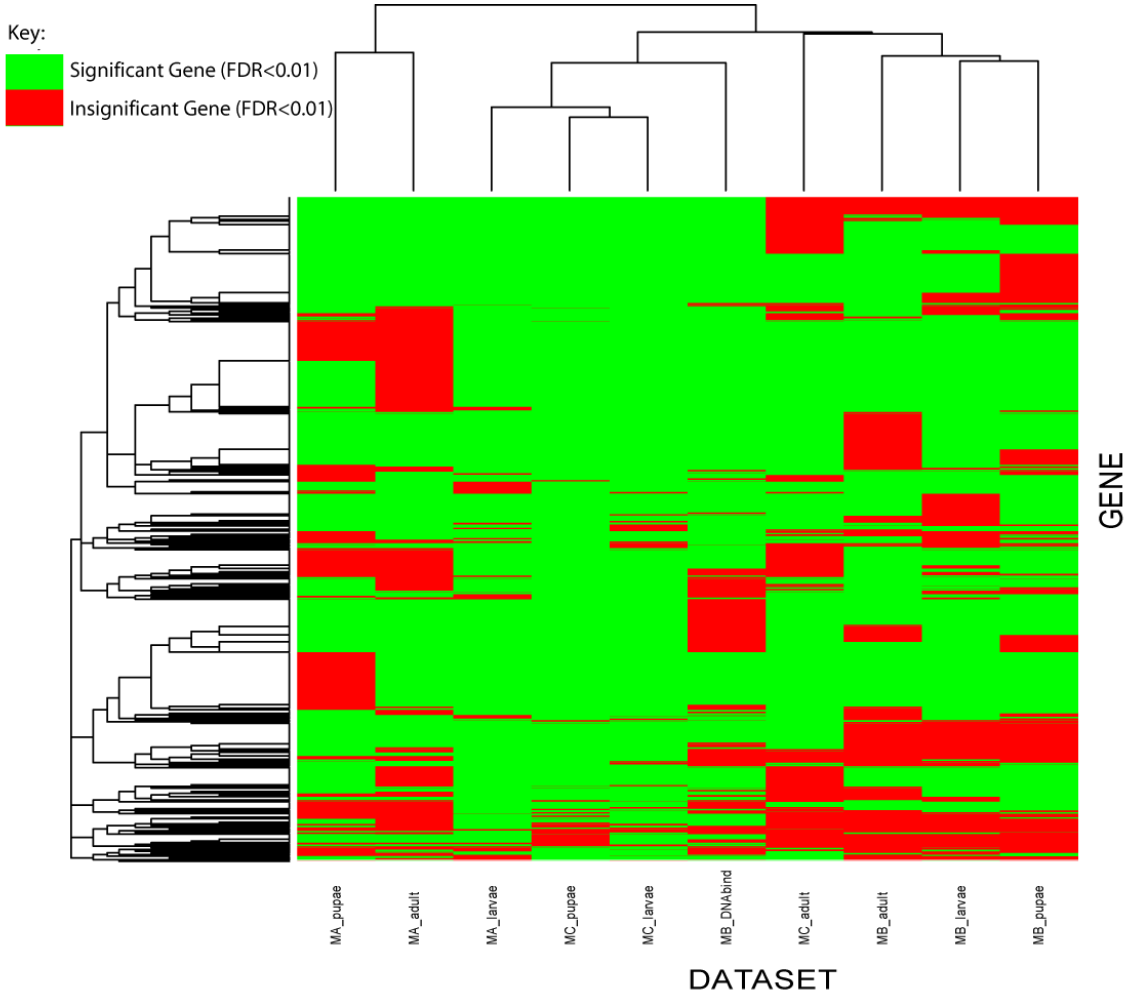


Figure 5-6 Significant genes in each dataset show a commonality between isoforms Fru^{MB} and Fru^{MC}. Significant genes (FDR < 0.01, peak height ≥ 1.8 are shown in green, those deemed insignificant by this cut-off point are coloured red. Due to the number of genes per dataset the individual genes cannot be read in this layout, however patterns between the datasets can be seen using the dendrograms on the left and top of the heat map.

5.5 Gene ontology and gene list analysis further distinguishes isoform Fru^{MA} from the Fru^{MB} and Fru^{MC} isoforms

To consider the types of genes represented in each gene list, gene ontology (GO) analysis was completed. GO analysis attributes standardised gene/ gene product information to sets of genes to allow characterisation for further analysis. Genes were described as a subset of one of three major categories: molecular function; biological function; cellular component. Each dataset was analysed in GOtoolbox, using the GOstats analysis with Bonferroni correction [182]. From this analysis the

p-value, along with calculated fold enrichment were graphed for the top 24 results (seen in Figures 5-8, 5-10, and 5-12).

The data showed little change in fold enrichment throughout the datasets, suggesting that high-ranking genes (by p-value) are of a similarly enriched state compared to the background. By combining GO data with overlap analysis of the original gene lists the differences between the datasets as a whole start to emerge.

Venn diagrams provide a useful means by which to look at the spread of the data across several datasets. Information can be gleaned from the commonality between the datasets, and those genes that are specific to datasets. To identify genes of interest for further characterisation, potential targets of interest might include those common to all isoforms within a time point, those unique to a specific time point or isoform, and those common to only one developmental time point implying these genes have temporally specific roles. For each Venn diagram the lists of genes in each are found in Appendix 3.

The whole GO data for each dataset was entered into a heat map. GO results negative in all datasets were removed for clarity and dendrograms added (ordering data based on its deviations from a common root), see Figure 42A (original data: Appendix 4). As previously seen, isoform Fru^{MA} pupae and adult were similar to each other but distinct from the other datasets. Figure 42B shows the most commonly represented categories, which show a strong involvement in the development of the brain and NS. The datasets with the least significant ontologies in the top 50 are the Fru^{MA} isoform sets, particularly the adult, suggesting distinct roles for this isoform.

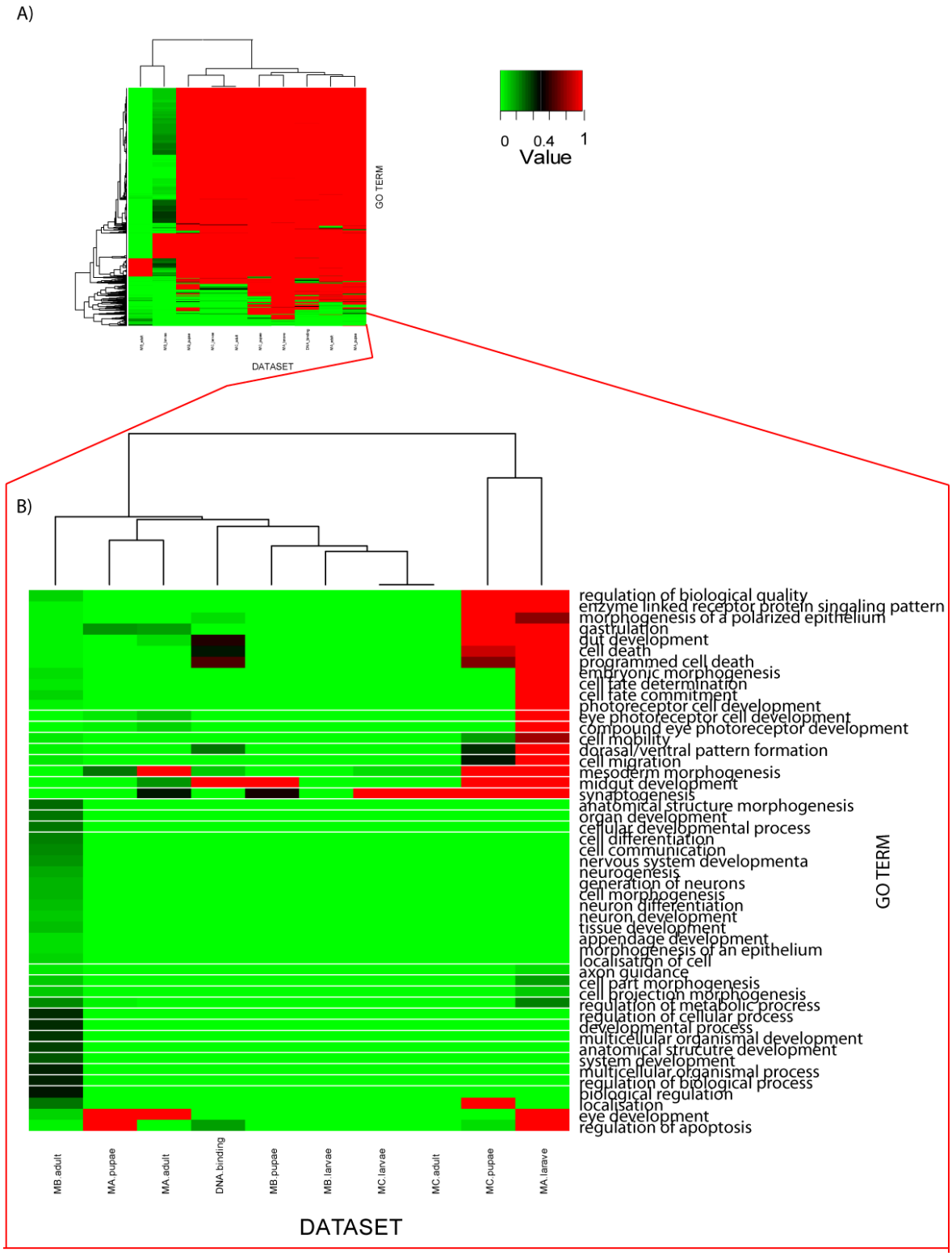


Figure 5-7 Overview of significant GO representation in the DamID datasets. A) All GO terms represented by the accumulation of DamID datasets. Datasets are in columns, GO terms are the rows, with both the datasets and GO terms being ranked in order of similarity by dendrograms. (B) Top 50 GO terms, as defined by those most frequently over-represented in all the datasets. The Red box represents these genes, which make-up the bottom portion of the heat map in (A). GO terms are defined to the right of the heat map, and again both GO terms and datasets are ranked by similarity. Heat map colours signify p-value, with smaller values (closer to 0) being more green, values around the middle value (0.5) being black, and those of less significance (closer to 1) being redder in colour. Data is derived from DamID experiments, run through the GO programme: GO Toolbox [182]. Dendrogram re-ordering occurs, due to the limited data shown in (B), therefore gene list ordering is altered.

To consider the datasets in more detail they were analysed by isoform, and then by time point.

5.5.1 The Fru^{MA} isoform: A distinct isoform with a physiological role

The Fru^{MA} isoform is quite distinct when compared to the other two Fru^{M} isoforms, Figure 5-8. It has been shown to be expressed in far fewer of the Fru-expressing neurons than either of the other isoforms [81]. These differences were further confirmed by the GO data analysis, which showed fewer significant ontology groups represented and different types of genes being targeted throughout development. p-values were also considerably lower, suggesting less specificity, and the range of ontology groups wider, which might represent fewer genes being common to individual ontology groups, or a less specific binding site [169]. Generally, the target gene still suggested NS involvement, however also showed high levels of regulation of physiology, particularly from the pupae time point onward.

The gene lists showed low levels of overlap between the larval time point (409 genes) and the pupae time point (1827 genes) and even fewer between the larval time point and the adult time point (2103 genes), 160 and 143 genes respectively, Figure 5-9. This suggested that the larval time point had quite distinct roles in comparison to the other two time points with over 50 % of the larval genes being specific to that time point. Comparatively, the overlap between pupae and adult time points, which represents 58 % of the pupae gene set, and 51 % of the adult dataset, showed similarity between these time points not seen between them and the larval time point. Only 106 genes were seen in all three time points. This data adds weight to the hypothesis that at the later time points, Fru^{MA} is following a

distinctly different role to its larval role: potentially a role in the regulation of neural physiology in the fly.



Figure 5-8 GO by isoform: Isoform Fru^{MA}. The top 24 GO categories for each developmental time point of Isoform A. Notably, the p-values of Fru^{MA} are only in the range of 0 to 10^{-25}. In each

instance, several categories associated with the nervous system and development are present, suggesting that the genes over-enriched in the DamID datasets are derived from these areas of the fly. Fold enrichment are in the range of 1.5 – 3.

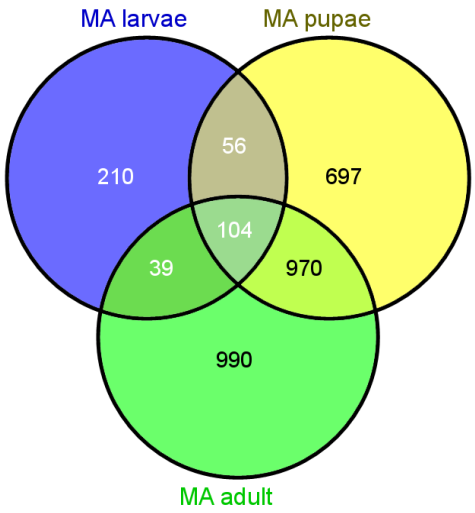


Figure 5-9 Venn diagrams depicting overlap of Fru^{A}. This data alludes to changes in function through development, and change in numbers of genes targeted. Larval time point is shaded blue, pupae time point is shaded yellow, and the adult time point is shaded green. The Fru^{MA} isoform showed low levels of overlap between the larval and pupae time points, and also the larval and adult time points, suggesting the larval time point is distinct to the other time points, which share some commonality.

5.5.2 The Fru^{MB} isoform: A functionally broad isoform with a clear role in the nervous system

The Fru^{MB} isoform showed a strong association with genes involved in NS development throughout, as seen in Figure 5-10. The most significant ontology groups, based on p-value, were connected to the NS and its development. As development progressed, it appeared that the Fru^{MB} isoform began to incorporate physiologically relevant targets. The most striking effect of the Fru^{MB} isoform was the high level of overlap between all three time points, suggesting little temporal change in genes targeted though development; 925 genes identified were present in all three datasets, Figure 5-11, which was reflected in GO analysis. The larval time point was again the smallest dataset (1756 genes), again this is unsurprising as Fru^{MB} is only just beginning to be expressed at this late larval stage. However, all the datasets were considerably larger for this isoform than either of the others, with adult containing 2059 genes and pupae 2147 genes.


MB LARVAE
TRANSCRIPTION
REGULATION OF RNA METABOLIC PROCESS
REGULATION OF CELLULAR METABOLIC PROCESS
REGULATION OF MACROMOLECULE METABOLIC PROCESS
REGULATION OF TRANSCRIPTION, DNA-DEPENDENT
REGULATION OF PRIMARY METABOLIC PROCESS
NEURON DIFFERENTIATION
CELL DEVELOPMENT
CELL FATE COMMITMENT
ORGAN MORPHOGENESIS
REGULATION OF NUCLEOBASE, NUCLEOSIDE, NUCLEOTIDE AND NUCLEIC ACID METABOLIC PROCESS
REGULATION OF TRANSCRIPTION
REGULATION OF GENE EXPRESSION
REGULATION OF BIOSYNTHETIC PROCESS
REGULATION OF CELLULAR BIOSYNTHETIC PROCESS
REGULATION OF MACROMOLECULE BIOSYNTHETIC PROCESS
ANATOMICAL STRUCTURE MORPHOGENESIS
GENERATION OF NEURONS
NEUROGENESIS
REGULATION OF CELLULAR PROCESS
CELLULAR DEVELOPMENTAL PROCESS
NERVOUS SYSTEM DEVELOPMENT
ORGAN DEVELOPMENT
SYSTEM DEVELOPMENT
MB PUPAE
CELL DEVELOPMENT
ORGAN MORPHOGENESIS
NEURON DEVELOPMENT
REGULATION OF NUCLEOBASE, NUCLEOSIDE, NUCLEOTIDE AND NUCLEIC ACID METABOLIC PROCESS
REGULATION OF METABOLIC PROCESS
REGULATION OF MACROMOLECULE METABOLIC PROCESS
IMAGINAL DISC DEVELOPMENT
REGULATION OF PRIMARY METABOLIC PROCESS
REGULATION OF CELLULAR METABOLIC PROCESS
REGULATION OF GENE EXPRESSION
REGULATION OF TRANSCRIPTION
NEURON DIFFERENTIATION
REGULATION OF CELLULAR BIOSYNTHETIC PROCESS
REGULATION OF BIOSYNTHETIC PROCESS
REGULATION OF MACROMOLECULE BIOSYNTHETIC PROCESS
CELL FATE COMMITMENT
ANATOMICAL STRUCTURE MORPHOGENESIS
REGULATION OF CELLULAR PROCESS
CELL DIFFERENTIATION
GENERATION OF NEURONS
NEUROGENESIS
ORGAN DEVELOPMENT
NERVOUS SYSTEM DEVELOPMENT
SYSTEM DEVELOPMENT
MB ADULT
TRANSCRIPTION
REGULATION OF TRANSCRIPTION, DNA-DEPENDENT
REGULATION OF METABOLIC PROCESS
NEURON DEVELOPMENT
ANATOMICAL STRUCTURE MORPHOGENESIS
REGULATION OF CELLULAR METABOLIC PROCESS
REGULATION OF MACROMOLECULE METABOLIC PROCESS
REGULATION OF GENE EXPRESSION
REGULATION OF TRANSCRIPTION
REGULATION OF NUCLEOBASE, NUCLEOSIDE, NUCLEOTIDE AND NUCLEIC ACID METABOLIC PROCESS
CELL DEVELOPMENT
REGULATION OF PRIMARY METABOLIC PROCESS
CELL FATE COMMITMENT
REGULATION OF MACROMOLECULE BIOSYNTHETIC PROCESS
NEURON DIFFERENTIATION
REGULATION OF CELLULAR BIOSYNTHETIC PROCESS
REGULATION OF BIOSYNTHETIC PROCESS
ORGAN DEVELOPMENT
GENERATION OF NEURONS
NEUROGENESIS
CELL DIFFERENTIATION
SYSTEM DEVELOPMENT
REGULATION OF CELLULAR PROCESS
NERVOUS SYSTEM DEVELOPMENT
MB DNA BINDING
REGULATION OF MULTICELLULAR ORGANISMAL PROCESS
CELLULAR COMPONENT MORPHOGENESIS
CELL PART MORPHOGENESIS
TISSUE DEVELOPMENT
AXONOGENESIS
NEURON PROJECTION MORPHOGENESIS
CELL MORPHOGENESIS INVOLVED IN NEURON DIFFERENTIATION
CELL PROJECTION MORPHOGENESIS
NEURON PROJECTION DEVELOPMENT
CELL MORPHOGENESIS
IMAGINAL DISC DEVELOPMENT
CELL MORPHOGENESIS INVOLVED IN DIFFERENTIATION
ANATOMICAL STRUCTURE MORPHOGENESIS
CELL DEVELOPMENT
NEURON DEVELOPMENT
SIGNAL TRANSDUCTION
ORGAN DEVELOPMENT
CELL DIFFERENTIATION
NEURON DIFFERENTIATION
GENERATION OF NEURONS
NEUROGENESIS
SYSTEM DEVELOPMENT
NERVOUS SYSTEM DEVELOPMENT
REGULATION OF CELLULAR PROCESS
4
2
0
1x10-35
1 x10-70
FOLD ENRICHMENT
P-VALUE

Figure 5-10 GO by isoform: Isoform Fru^{MB}. The top 24 GO categories for each developmental time point of Isoform Fru^{MB}. Notably, the p-values of Fru^{MB} are much smaller than those seen for isoform Fru^{MA}, with some as small as 1 x 10^{-80}. In each instance, several categories associated with the nervous system and development are present, suggesting that the genes over-enriched in the DamID datasets are involved in the development of this tissue. Fold enrichment are in the range of 1.5 – 5.

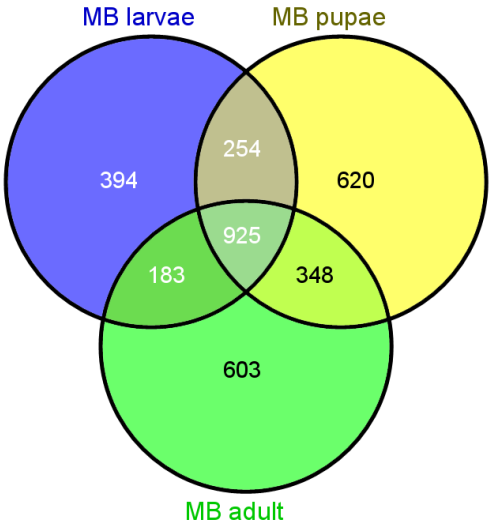


Figure 5-11 Venn diagrams depicting overlap of Fru^{B}. This data alludes to changes in function through development, and change in numbers of genes targeted. Larval time point is shaded blue, pupae time point is shaded yellow, and the adult time point is shaded green. The Fru^{MB} isoform showed high overlap throughout development, suggesting a common function through development.

5.5.3 Isoform Fru^{MC}: An isoform with narrow specificity and a clear role in the nervous system

Like the Fru^{MB} isoform, the Fru^{MC} isoform also showed a strong association with genes involved in NS development throughout, Figure 5-12. Similarly, as development progresses, more ontologies were suggestive of being involved in aspects of physiological regulation, mirroring the progress of Fru^{MB}'s involvement in the development of the fly.

The Fru^{MC} isoform datasets were generally smaller, and unfortunately harder to compare between datasets as the pupae dataset was created using an older, less accurate tiling array, potentially leading to a smaller number of genes being identified (283 genes; see Figure 5-13). Noticeably, the adult time point was considerably larger (1861 genes) than either of the larval or pupal datasets,

implying a huge increase in activity at this time point, or that this is an artefact of the methylation building up in the adult fly (which is unlikely as it had not been seen in any other datasets), but potentially due to the low number of targets identified at the pupae time point. As the peak of Fru expression is known to be at the pupae time point, it seems most likely that the decrease in expected genes for the pupae time point was related to the different tiling array used. Although only 2 % of genes overlapped between all three time points, strong overlap was noted between the larvae and adult (69 %) and the pupae and the adult (89 %), potentially suggesting that the early time points are only targeting the strongest interactions, and that targeting is broader by the adult, alluding to a shift to wider regulation of physiology.

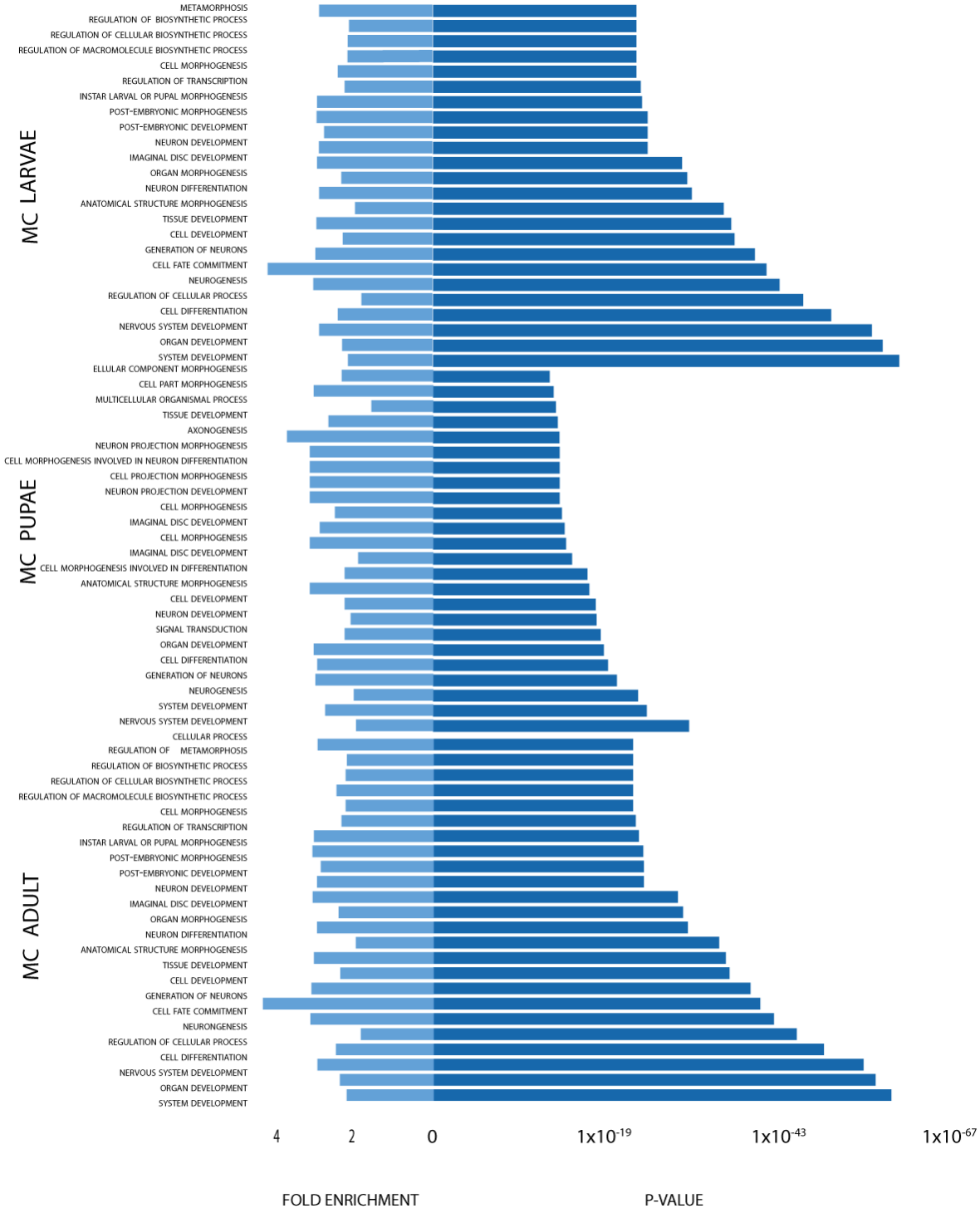


Figure 5-12 GO by isoform: Isoform Fru^{MC}. The top 24 GO categories for each developmental time point of Isoform Fru^{MC}. The *p-value*s of isoform Fru^{C} are in the range of 0 to 10^{-70}. In each instance, several categories associated with the nervous system and development are present, suggesting that the genes over-enriched in the DamID datasets are involved in the development of this tissue. Fold enrichment's are in the range of 1-5.

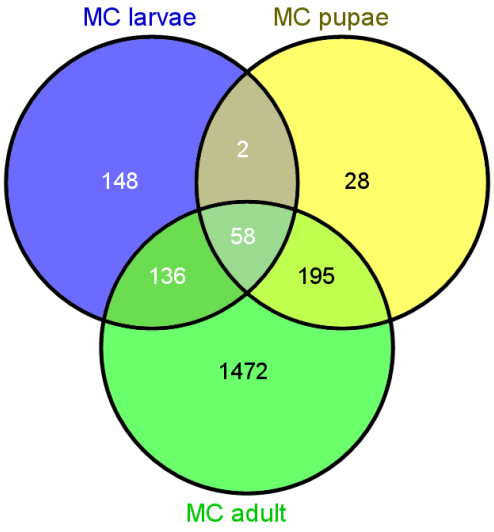


Figure 5-13 Venn diagrams depicting overlap of Fru^{C}. This data alludes to changes in function through development, and change in numbers of genes targeted. Larval time point is shaded blue, pupae time point is shaded yellow, and the adult time point is shaded green. The Fru^{MC} isoform showed high overlap throughout development, suggesting a common function through development.

5.5.4 GO by time point: Common elements link the isoforms

The larval time point showed the most similarity between isoforms of all the time points (see Figures 5-9, 5-11 and 5-13). At this stage, gene ontologies would be expected to be mainly developmental. This was seen in the 'imaginal disc' related terms, differentiation and morphogenesis ontologies as well as basic development ontology groups such as 'system development' seen for isoform Fru^{MA}. For isoforms Fru^{MB} and Fru^{MC} noted ontologies were 'regulation of cellular metabolic process' and neuron development ontologies including 'generation of neurons'.

The pupae time point would still be expected to be largely developmental. As these genes are downstream targets of Fru, it would be wise to assume that these developmental ontologies would start to have an increased involvement in the development of the CNS and PNS. With the Fru^{MA} isoform, this was not obvious, with only one term related directly to the development of the NS,

however in isoforms Fru^{MB} and Fru^{MC}, five of the top 24 ontologies were directly linked to the development of the NS: Neuron development; neuron differentiation; generation of neurons; neurogenesis; nervous system development. With reference to the mounting information suggesting a strong link between the downstream targets of isoforms Fru^{MB} and Fru^{MC}, the fact that these five ontology groups were identical adds more weight to this argument. Further, the most significant six ontologies (by p-value) for isoform Fru^{MC}, and the most significant eight ontologies (by p-value) for isoform Fru^{MB} were found in the reciprocal ontology analysis showing strong overlap between the types of genes targeted by each isoform. The data from isoform Fru^{MC} shows links to regulatory roles with localisation and signalling ontologies.

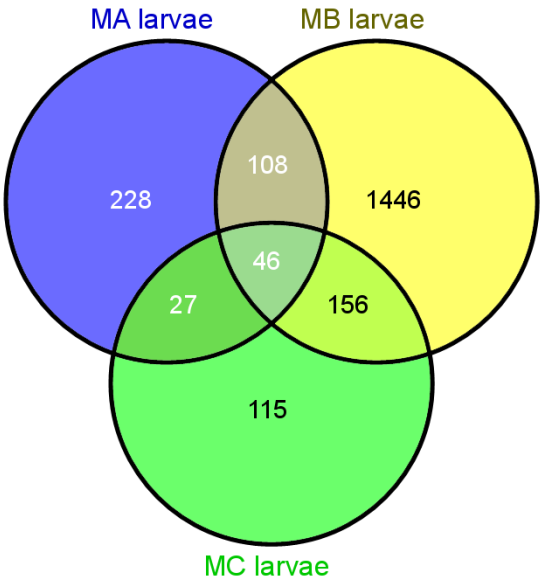
The adult time point depicts the effect of Fru once development had been achieved and thus the effects of Fru in a more regulatory role – the maintenance of the adult fly. In isoform Fru^{MA}, the ontology groups suggested a shift towards more of a role in general regulation of the cell, with ontologies described with broad descriptors suggesting regulation of whole systems. Comparatively isoforms Fru^{MB} and Fru^{MC} both still showed a strong connection to the NS, and otherwise showed an involvement of the targets in the regulation of cellular processes.

5.5.5 Venny by time point: Gene targets change throughout development

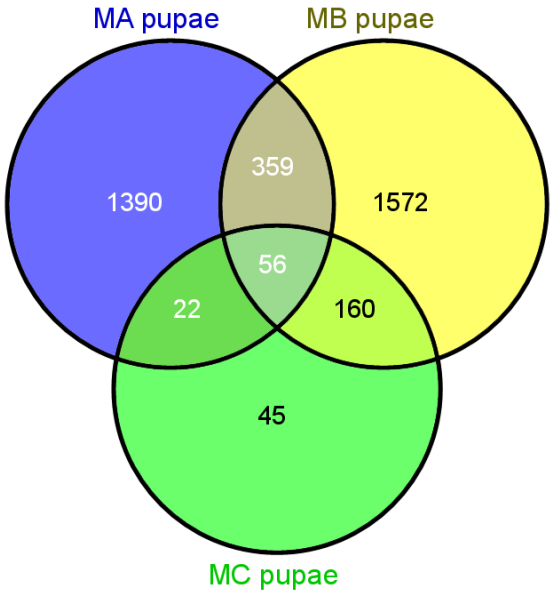
When comparing each isoform at the same time point it was immediately noticeable that the isoforms had significant overlap with each other, suggesting a commonality between their targets at a given developmental stage (Figure 5-14).

This overlap, however, only accounts for a small percentage of the whole datasets, therefore each isoform shows clear specificity.

A)



B)



C)

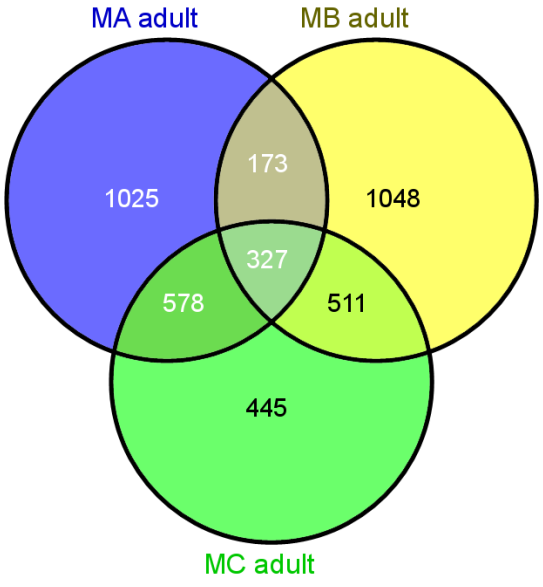


Figure 5-14 Venn diagrams show the overlap between individual time points. The analysis allows the similarities and differences between the datasets to be analysed. For these images Isoform Fru^{MA} data is in blue, Isoform Fru^{MB} is shown in yellow and Isoform Fru^{MC} is in green. (A) The larval time point shows relatively little overlap between the datasets. (B) The pupae time point, whose datasets are large with the exception of Fru^{MC} (likely to be an artefact of the tiling array used). (C) The adult time point are large had have a high degree of overlap (328 genes found in all three isoforms, compared to fewer than 50 for the other two time points).

The overlaps seen between the Fru isoforms at the larval time point are seen in Figure 5-14A. Most of the genes were unique to the individual isoforms suggesting a diverse, isoform specific role at the larval time point, perhaps indicative of the huge scope of changes required in early development. Similarly, although there was some overlap between the samples, this was small: 2 % overlap between all three isoforms, and 7 %, 10 % and 10 % overlap between isoform Fru^{MA} and Fru^{MB}, Isoforms Fru^{MA} and Fru^{MC} and Isoforms Fru^{MB} and Fru^{MC}, respectively. With the exception of Fru^{MB}, that showed a large number of targets, the datasets were small, potentially reflecting the decreased levels of expression at this developmental stage. For Fru^{MC} this was an unusual deviation from its often-broad similarity to Fru^{MB}. These observations suggest diverse roles for the different isoforms at this developmental stage.

The datasets for the pupal time point are generally much larger, in line with the increase in Fru expression at this developmental stage, Figure 5-14B. However, there is still relatively little overlap between the isoforms, especially between all three datasets. Those that do overlap with all three isoforms include *svp, fru,* and *flw,* genes that potentially play a key role during this developmental period. There was a large amount of overlap between isoform Fru^{MC} and Fru^{MB}, with 76 % of all genes from the Fru^{MC} gene list also being found within the Fru^{MB} gene list. Generally, isoform Fru^{MC} appears to be underrepresented at the pupae time point, likely due to the differences in the tiling array used to collect this data.

The adult time point showed the greatest amount of overlap between the datasets, Figure 5-14C. Approximately 5 % of all datasets overlapped in all three isoforms. Further, when considering the overlap between pairs of datasets: 12 % between isoforms Fru^{MA} and Fru^{MB}; approximately 23 % between isoforms Fru^{MA} and Fru^{MC}; and approximately 21 % between isoforms Fru^{MB} and Fru^{MC}. This suggested that by the adult time point, there is much more functional overlap between isoforms.

5.5.6 DNA binding mutants: identifying the effects of co-factors

Only one DamID for a Fru-DNA binding mutant datasets had been processed. The mutant chosen was Fru^{mutMB} at the adult time point, which was chosen, as the Fru^{MB} adult data appeared to be the most robust and compelling. As the Fru^{mutMB} mutant in theory should not be able to bind DNA directly itself, we were unsure what results to expect when using it for DamID, for example it was possible that the dataset would contain no targets. This would suggest that either Fru^{MB} does not bind co-factors, or that even if co-factor binding does occur, it is the Fru^{MB}-DNA binding ability that is solely responsible for targeting the DNA. This however, was not the case. The Fru^{mutMB} DamID results produced many targets, in fact more than were potentially expected, suggesting that co-factor binding (be that to other Fru proteins through BTB-dimerisation, or other proteins altogether) occurs regularly and accounts for many of Fru's interactions. Unfortunately, the data cannot be separated into binding events formed through interactions with Fru, and those formed through interactions with other proteins.

Figure 5-15 shows the overlap between the Fru^{MB} dataset and the Fru^{mutMB} dataset. The blue circle represents targets that were targeted by Fru^{MB} directly interacting with DNA, The yellow circle represents those sites Fru^{mutMB} where any

direct DNA interaction is due to binding with a co-factor or dimer. The overlap would represent targets in the Fru^{MB} dataset that are not due to direct DNA-binding.

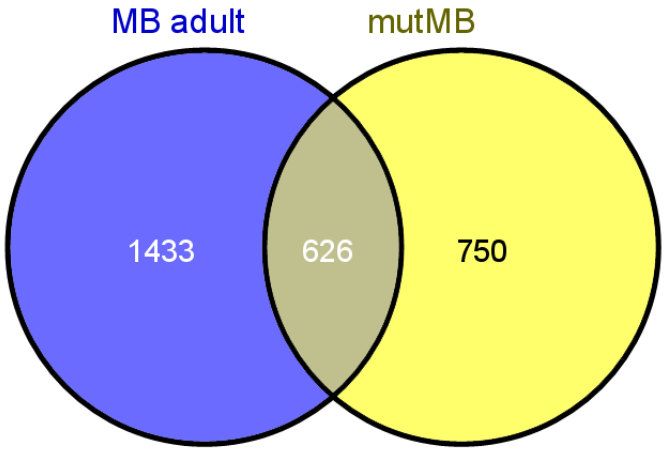


Figure 5-15 Fru^{MB} adult and Fru^{mutMB}. The DNA binding mutant cannot bind DNA directly, therefore the targets identified using this mutant are those bound by co-factors, or Fru dimers.

Perhaps some level of understanding as to whether or not Fru^{MB} is binding to other Fru proteins could be derived from looking at the overlap between all the adult datasets, Figure 5-16. Initial interpretation suggests that the DNA binding domain of Fru^{MB} alone is required to correctly bind approximately 30 % of its targets. Further, in the overlapping region of all three isoforms, there is a small decrease from 173 genes to 127 genes – suggesting that in some instances, the genes work co-operatively, and that in some situations Fru^{MB} DNA-binding is required. Although the mutant data was informative, given the ability of Fru isoforms to homo- and hetero-dimerise the interpretation of the data was less straightforward then we had anticipated. For these reasons, no further DamID experiments were completed at this time. However, the data was very useful for motif analysis and therefore the other DNA-binding mutants may be used in the future for additional DamID experiments.

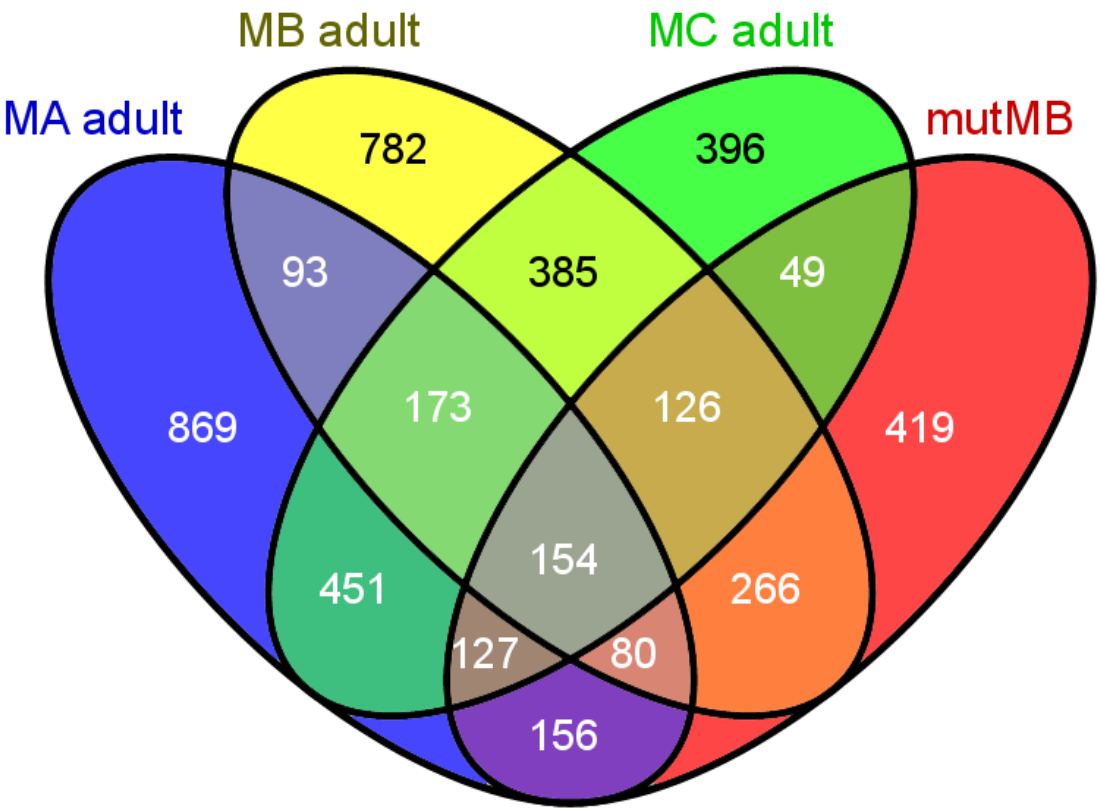


Figure 5-16 Comparison of the DNA binding mutant with other isoforms in adults. This analysis compares all Fru DamID dataset targets at the adult developmental stage with each other and the Fru^{mutMB} mutant.

5.6 The representation of the nervous system gene ontology group

Due to Fru's obvious connection with genes associated with the development of the NS along with its NS specific expression, the gene ontology group 'Nervous System Development' (NSD) was further analysed. This analysis compared the number of targets found within this group and its downstream directly-linked ontologies [35, 78, 79, 107]. The Gene Ontology Consortium define the category as "Nervous System Development (GO:0007399): the process whose specific outcome is the progression of nervous tissue over time from its formation to its mature state" [183]. In *Drosophila melanogaster*, 1261 genes were identified in the NSD ontology and its downstream ontologies. Of these genes, 519 (59 %) were found in one or more of the Fru DamID datasets, showing a strong relationship between Fru and the NSD ontology group, Figure 5-17. The dendrograms at the

top of the heat maps continue to show a significant difference between the pupae and adult datasets for Isoform Fru^{MA} and the rest of the Fru data, suggesting that this isoform has a different, non-developmental role at these time points. Only 32 genes were expressed in Fru^{MA} isoform throughout development (all of which are expressed in both the larval and pupal time points) including *dsx* and *chinmo*, and Isoform Fru^{MC} had 31 genes in the NSD expressed development, all of which were expressed during the larval time point (including *CycE, EcR* and *pros*).

Only two genes were found throughout all the datasets, *fru* and *apt*. When the Fru^{mutMB} mutant dataset was not considered, this increased to include *svp*. There are clear differences when comparing the Fru^{mutMB} and Fru^{MB} datasets. Including the DNA binding mutant set, 127 genes from the NSD were expressed at some point by the Fru^{MB} isoform; these include *pros*, *CycE*, *EcR*, *ttk* and *robo3*, of which *EcR* had previously been suggested as downstream targets of Fru [119]. When the DNA binding mutant dataset was discounted, this number rose to 263 genes, further including *lola*, *pdm2, dsx,* and *chinmo.* Fru^{MB} produced larger datasets throughout the time points, and the NSD ontology group was more significantly enriched in the Fru^{MB} datasets when compared to the other Fru isoforms. Fru^{MB} appears to be the predominant isoform affecting the NS.

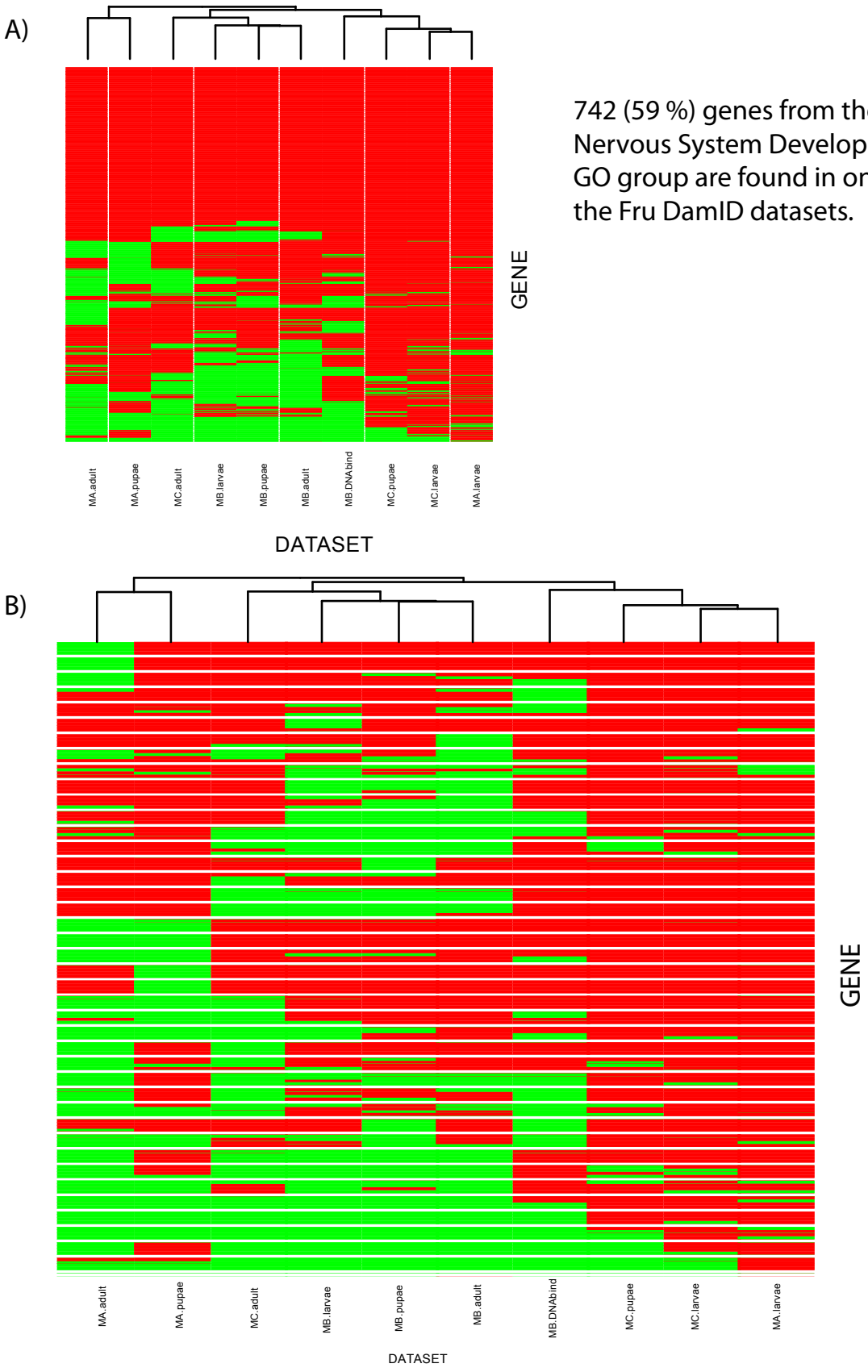


Figure 5-17 Representation of downstream Fru targets in the nervous system development ontology. Of the 1261 genes in the Nervous System Development Ontology, 519 are found in one or more Fru DamID. The highlighted box shows a heat map for the overlapping genes showing a greater difference between Isoform Fru^{MA} and Isoforms Fru^{MB} and Fru^{MC}. Fru^{MB} has the most significant enrichment of genes within this ontology group (however this may be due to the Fru^{MB} datasets being considerably larger rather than a significantly increased overlap). Red signifies a gene that does not appear in the dataset, green represents a Nervous System Development ontology gene also found in a DamID Fru dataset.

The data confirms the similarities between Isoforms Fru^{MB} and Fru^{MC}, having 64 % overlap, compared with only 13 % overlap with Fru^{MA}. These isoforms stem from a common branch of the dendrogram, suggesting conservation between the exons.

5.7 Motif analysis: Identifying Fru DNA-binding sequences

Motif analysis was undertaken by collaborators in the Aerts Lab, Katholieke Universiteit, Leuven, Belgium. Using data collected through DamID, The Aerts Lab used CisTargetX to look for over represented known sequences in the ten DamID datasets (see Figure 5-18) [184]. Again, they showed that Fru^{MA} is significantly different from the other two isoforms, and that the Fru^{MA} larval time point differs significantly from the other Fru^{MA} developmental time points. A Fru^{MA} DNA binding site has been identified from the on going genome-wide project using the bacterial one-hybrid method to systematically describe the binding site preferences of transcription factors in Drosophila melanogaster [185]. This sequence was only enriched in the Fru^{MA} larval dataset. Isoform's Fru^{MB} and Fru^{MC} showed a great deal of similarity in their DNA-binding site specificity, showing enriched sequences resembling Ttk and Kruppel (Kr). Significantly, these sequences were not enriched in the targets of the Fru^{mutMB}, confirming the specificity of the observed enrichment. The two candidate sequences the Aerts Lab identified as being most likely to be motifs for Fru were related to known binding sites for the transcription factors Ttk and Kr, and it is these sequences upon which they based further analysis.

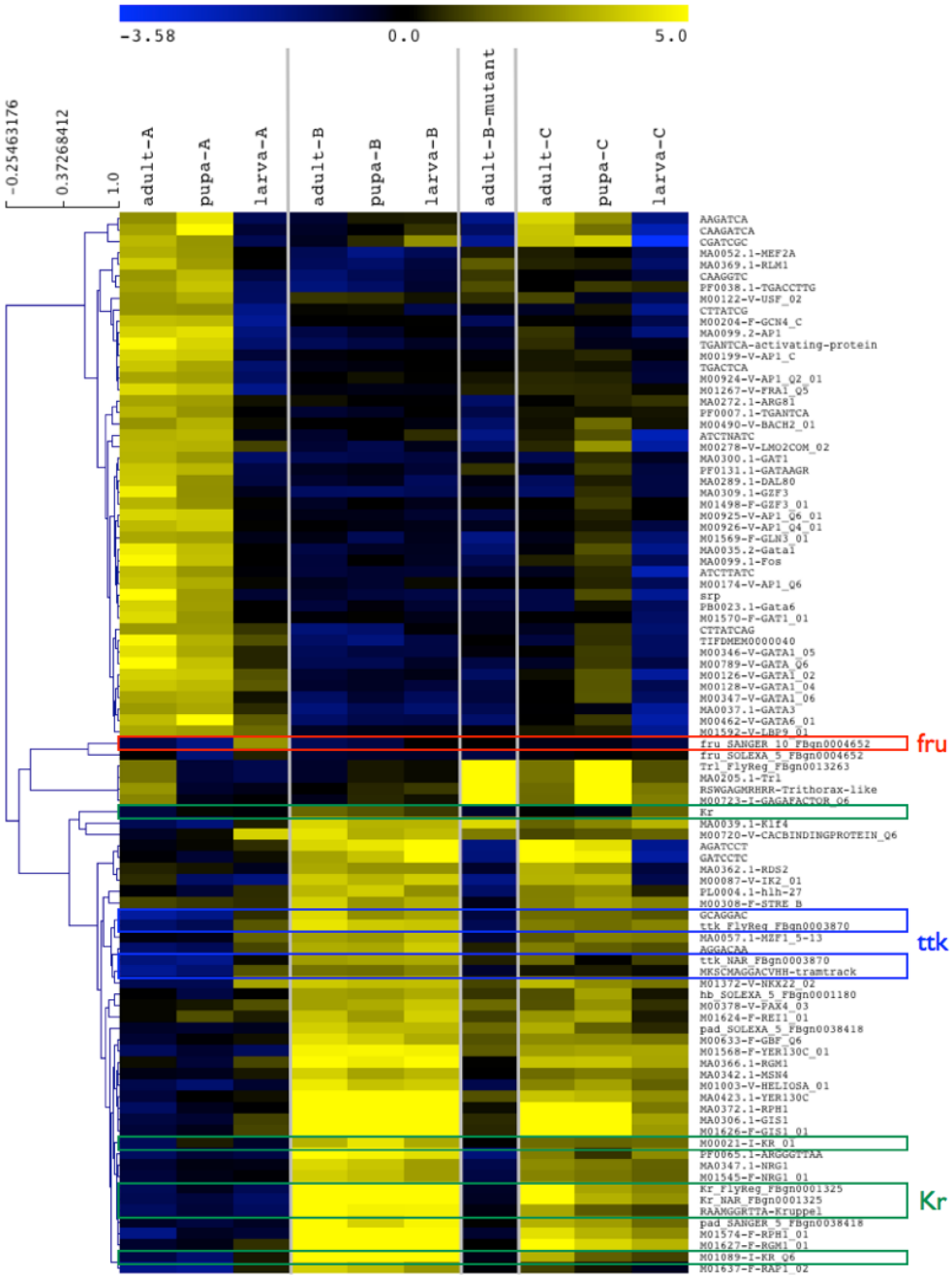


Figure 5-18 Identifying putative MOTIFs in DamID targets. Using known and non-specific MOTIFs, all ten DamID datasets were compared to identify re-occurring sequences within the top 1000 putative targets (B. Van der Sande). Yellow represents strong motif enrichment and blue represents weak motif enrichment. Sequences related to Ttk and Kr are indicated by blue and green boxes, respectively. The Fru^{A} binding site identified using the one-hybrid system is shown in red.

As Ttk is the most closely related gene to Fru in *Drosophila melanogaster*, with Ttk's protein structure showing striking similarities to Fru (see Figure 5-19C), the motifs identified as being Ttk-like are the most likely to be direct DNA-binding targets of the Fru^{MB} and Fru^{MC} isoforms (Figure 5-19A).

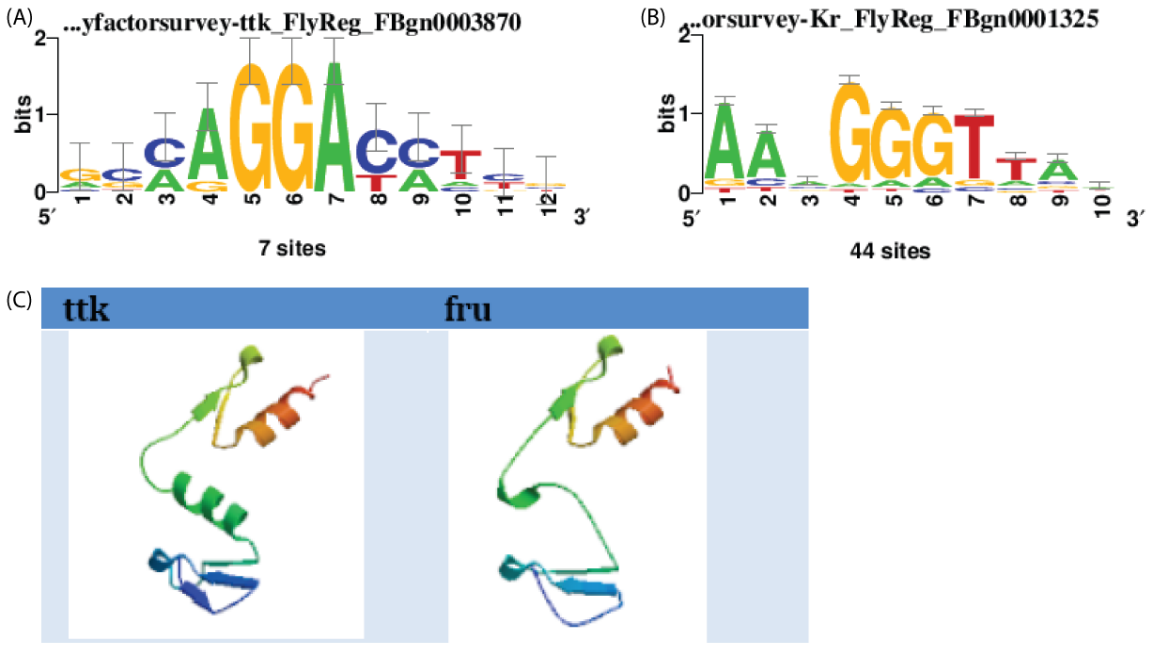


Figure 5-19 Ttk and Kr as potential MOTIF sequences for Fru. MOTIF analysis has suggested Ttk and Kr-like DNA sequences represent Fru's binding sites (A) Ttk motif identified from Fru targets. (B) Kr motif identified from Fru targets. (C) Predicted 3D protein structure comparing Ttk and Fru C_2H_2 DNA binding domains.

From these findings, gene lists were created from the original DamID gene lists identifying those targets that contain the Ttk, Kr and Fru motifs. 180 Fru containing targets were identified for the Fru^{MA} isoform, 71 Ttk and 91 Kr containing targets were found in all Fru^{MB} datasets. Thus far the Fru^{MC} isoform has not been analysed.

5.8 Visualisation of the peaks

5.8.1 Signalmap™: Fruitless self-regulates promoters three and four

Initially, visualisation was via the Nimblegen visualisation software, Signalmap™, which allows visualisation of the various datasets compared to a genome layout – to which gene schematics can be added through conversion to illustrator. These visualisation methods are helpful for when comparing various datasets. This

approach was used to generate a visual display of Fru isoforms apparent self-regulation.

Fru was found to self-regulate two promoters, P3 and P4, Figure 5-20. All isoforms regulated at least one promoter at one point in the developmental process. For Fru^{MA}, there was no binding in the larval time point, but by the pupal stage both P3 and P4 were targeted, which was maintained through the adult time point. By comparison, Fru^{MB} binds P3 at the larval stage, and possibly P4. This pattern is maintained through the pupal and adult stages, albeit the signal is slightly downstream of the P4 promoter. This however, does not necessarily represent the location of the binding site, as DamID uses local GATC sequences for identifying targets, and thus the local GATC sites may be further downstream at this point. Interestingly, the DNA binding mutant for Fru^{mutMB} shows no signal at P3, suggesting that this targeting requires DNA binding of Fru^{MB} directly. Isoform Fru^{MC} also binds at P3 but not P4 at the larval time point and this pattern is maintained through the remainder of the developmental time points.



Figure 5-20. Fruitless self-regulates at promoters P3 and P4. *fru* was found self-regulated throughout development. Fru^{MA} is shown in purple, Fru^{MB} is shown in blue and Fru^{MC} is shown in green. The yellow shading represents Fru binding. The Y scale is in base pairs, and represents the genome. At the top of the image, a schematic of the *fru* locus is depicted, where boxes represent exons, and black lines alternative splice variants. The X-axis is Fru-Dam/ Dam ratio. The red line represents a threshold of 1.8 (approximately 5 % of each dataset) signifying a relevant peak.

5.8.2 Improving visualisation of the datasets

Although Signalmap™ was a convenient way to scan through this data, it was not very compatible with other systems. To be able to visually compare and contrast our datasets with other genome data that is available, a web based genome browser was required. To modify this data to be compatible with a genome browser, an R script was designed to extract and re-orientate the information into a .wig file. Wig files were chosen as the file for conversion as these files are able to show the data as an intensity file, whereby the stronger the intensity recorded from a given probe, the darker the shade of grey is visualised, allowing a simple way to successfully analyse multiple data streams simultaneously – providing very clear imagery of the most specific targets. The probes used for the DamID protocol were overlapping, and so to accommodate the requirements of the .wig format, the files had to be re-ordered, and altered so that the probes ran sequentially, rather than overlapping. For a full description of the R scripts used see 2.5.2. This data was uploaded to the UCSC genome browser. The full datasets can be visualised at the following website: www.genome.ucsc.edu

Session information: User Name: Elizabeth Ashley
Password: Fruitless

The resultant data is seen in Figure 5-21, which as an example shows the same Fru data as seen in Signalmap™ in Figure 5-20.

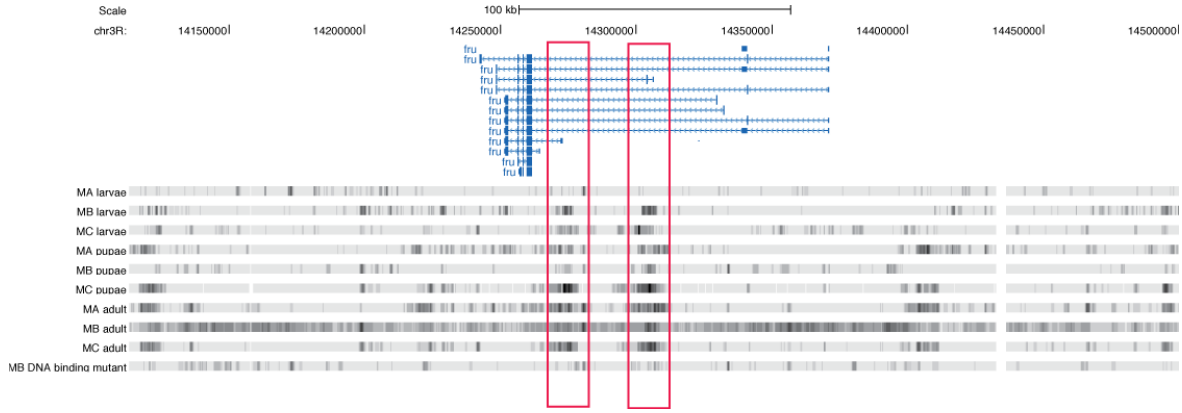


Figure 5-21 DamID datasets shown as a .wig file on UCSC genome browser. Intensity mapped DamID data is shown as a single line for each dataset. Above in blue, a *fru* gene schematic is shown, indicating all isoforms. The red boxes identify strong signals at promoters P3 and P4 in *fru*. The scale, in base pairs, runs along the Y-axis. Intensity stream is on a scale of Fru-Dam/Dam 1-6.

The data for Fru is visually much cleaner using this method of visualisation and shows neat intensity peaks under *fru* promoters P3 and P4 in nearly all datasets.

5.9 Genes of interest

5.9.1 Merging, intersecting and subtracting datasets can identify potential priority genes of interest

Flymine allows datasets to be merged, intersected and subtracted, allowing easy identification of genes representative of all categories (intersect), those which are not common to multiple datasets (subtract) and the ability to merge multiple datasets into one file (merge). Certain genes can be identified as being of potential interest due to their localisation to a specific isoform, developmental time point or their commonality between all of the datasets. Some potential candidate genes identified using this approach are seen in Table 5-2.

Candidate Gene	Isoform/time point										Interest
	A			B				C			
	l	p	a	l	p	a	mut	l	p	a	
Flapwing (flw)		1	6			12		11	1	6	A/B/C
Chinmo	12			8							larvae
VGlut											A_{larvae}
Cyclin E (CycE)				12	15						A/B/C
POU domain protein 2 (pdm2)					4	1					B
Fruitless (fru)							7				A/B/C
Longitudes lacking (lola)									23		pupae
ovo								2		3	C

Key:

0	
>100	
>200	
>500	
>1000	
Not present	

Table 5-2 Candidate genes of immediate interest. A/ B/ C represent the Fru isoforms, l/ p/ a/ mut represent larvae/ pupae/ adult/ DNA binding mutant respectively. The values in the key show the colour representation of peak position based on statistical significance and peak-height. Where a value is given in the table, this represents the position of the peak within the top 25 peaks

5.10 Genes of interest: Further analysis

From the above results, a selection of potential Fru-targets genes were identified for further analysis. They were chosen either for their high ranked position within a single dataset, their occurrence in multiple datasets, or having functions potentially related to Fru. The first step in the analysis of these candidate genes was to research any known information on the candidates found in the literature. The following genes were identified for further analysis:

5.10.1 Chronologically inappropriate morphogenesis (Chinmo)

Chinmo is a BTB-ZnF protein (like Fru) involved in neuronal temporal identity in postembryonic development of the *Drosophila* brain [186, 187]. It aides in successful fate determination in the mushroom body and projection neuron

neuroblasts, and if absent causes both types of neuroblasts to assume late-born neuronal identity at the early-born stage. This is achieved by a gradient of Chinmo expression throughout development, which controls birth order-dependent fates in such neuroblast lineages, Figure 5-22.

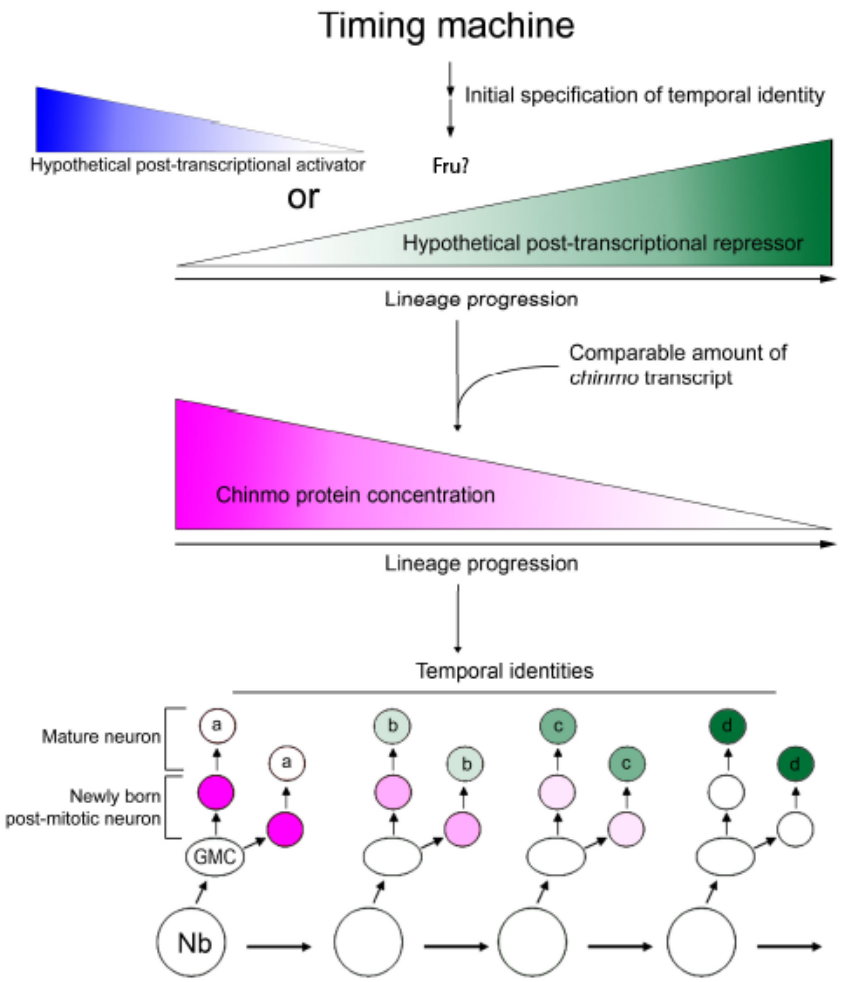


Figure 5-22 The role of *chinmo* in establishing neuronal cell identity. *chinmo* forms a gradient, which results in specifying cell fates of mushroom body and projection neuron neuroblasts throughout post-embryo development. This gradient is thought to be controlled either by an activator, which over time decreases, or a repressor, which over time increases. Either of these could likely be a transcription factor, and therefore have the potential to be Fru. Modified from Zhu et al., [186].

Chinmo also potentially has a role in the established neuroblast lineage fate cascade: Hunchback (Hb) → Kr → POU domain transcription factors (Pdm) →Castor (Cas). All of which are transcription factors, several of which also rank highly as targets of Fru [186]. *chinmo* has been shown to be required during neurogenesis, but not subsequently in active neuroblasts – suggesting a

developmental role only, which is reflected in the DamID data, where *chinmo* is a clear target in the larval time point of isoforms Fru^{MA} and Fru^{MB} (see Table 5-2 and Figure 5-23).

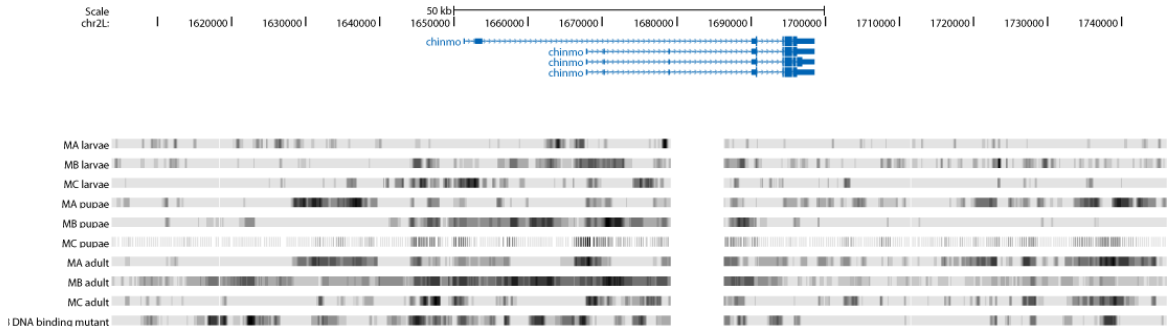


Figure 5-23 DamID binding of *chinmo*. Intensity mapping of Fru targeting the *chinmo* genomic region. Where Intensity mapping is on a scale of 1-6 (Fru-Dam/Dam ratio). The X-axis is the genome in base pairs, a schematic of the *chinmo* gene is shown in blue.

5.10.2 POU protein domain 2 (pdm2)

The Pou protein domain 2 (pdm2) gene is a sequence-specific DNA binding transcription factor with a role in neuronal cell fate determination. It provides cell identity function in the developing nervous system via regulation by Cas and Hb [188, 189]. More recently it has been shown that pdm2 is expressed specifically in proprioceptive neurons in the pNS, which send projections to the CNS, where it appears to have a role in the development of dendritic arborisations and axonal targeting [190].

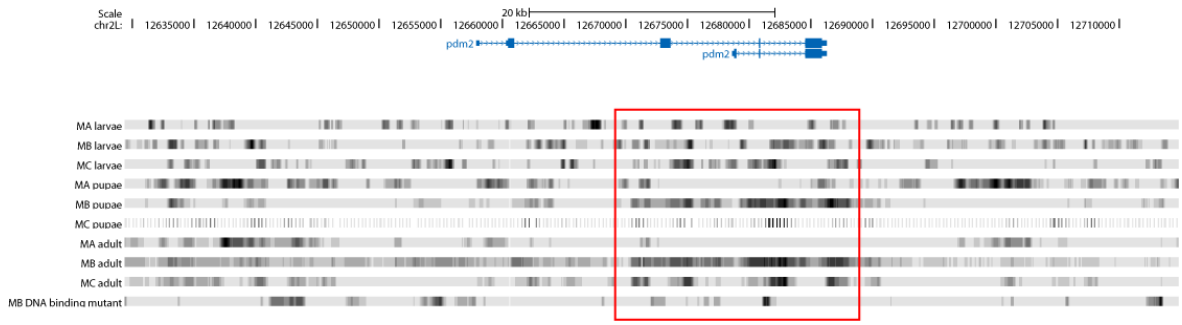


Figure 5-24 DamID binding of *pdm2*. Intensity mapping of Fru targeting the *pdm2* genomic region. Intensity mapping is on a scale of 1-6 (Fru-Dam/ Dam ratio). The X-axis is the genome in base pairs with the *pdm2* gene schematic in blue. The red box depicts the area of binding associated with the smaller *pdm2* isoform.

Within the DamID datasets, only isoform Fru^{MB} showed a strong interaction, which appeared to be largely associated with the smaller of the two identified *pdm2* transcripts, Figure 5-24. This association appears to be DNA-binding dependent as it was abolished in the DNA-binding mutant Fru^{mutMB}.

5.10.3 Kruppel (Kr)

Kr is a C_2H_2 zinc finger transcription factor, which is an essential gap gene that has been shown to have a role in neuroblast determination and axon guidance [187, 191]. In a recent screen in the compound eye for modifiers of roughness induced by fru^+ overexpression, 4 enhancers were identified as *Kr* alleles and 1 suppressor as *hb*, a known transcriptional regulator of *Kr* [31]. This screen identified both *Kr* and *hb* as potential mediators of Fru actions. Further analysis of Hb suggested it is essential for male-typical shaping of the contralateral neurites by Fru, however any direct connection between Fru and Kr remains unclear. *Kr* only produces one transcript, which is heavily targeted by all Fru isoforms, at all time points, bar isoform Fru^{MA} in the larvae (Figure 5-25). Unfortunately, due to the different tiling array employed for the Fru^{MC} pupae dataset, no data is available for Fru^{MC} pupae in this region of the genome. There is also clear binding in the Fru^{mutMB} mutant dataset, therefore this binding may be due to Fru-Fru dimerization or reflect the association of Fru with a co-factor.

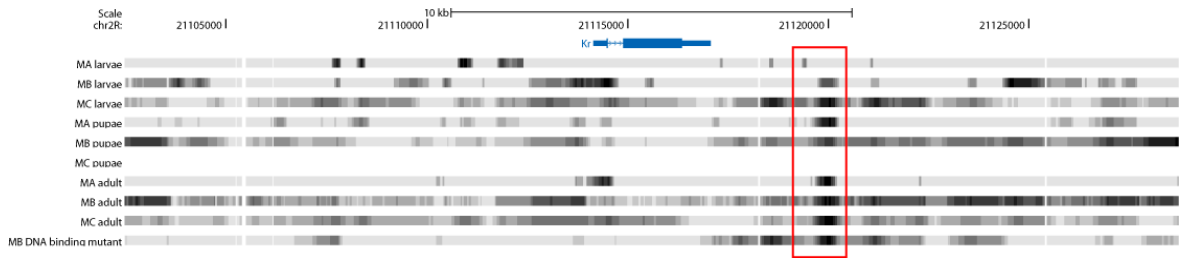


Figure 5-25 DamID binding of *Kr*. Intensity mapping of Fru targeting the *Kr* genomic region. Intensity mapping is on a scale of 1-6 (Fru-Dam/ Dam ratio). The X-axis is the genome in base pairs with the *Kr* gene schematic in blue. The red box depicts the peak of Fru binding, which is

seen in all isoforms, bar Fru^{MA} larvae. There is no data for Fru^{MC} pupae at the *Kr* locus due to a lack of coverage in the differing tiling arrays used.

5.10.4 *flapwing (flw)*

flapwing (*flw*) codes for serine/threonine protein phosphatase type 1β (PP1β) and was initially identified in a mutant screen as an effector of indirect flight muscle (IDF) formation [192]. Regulation of non-muscle myosin activity is the single essential *flw* function that is non-redundant with the three closely related PP1α genes (85 % sequence identity) [193]. In addition, *flw* has been shown to control oocyte polarization by regulating myosin activity [194]. Myosin II has also been shown to play an important role in asymmetric cell division in dividing neuroblasts [195]. A potential role for *flw* as a target of Fru would be to regulate myosin II during the critical stage of asymmetric neuroblast cell division. *flw* is seen to be targeted by Fru^{MB} throughout development, and also by Fru^{MC} in pupal and adult stages (Figure 5-26). This target potentially highlights the similarities between Fru^{MB} and Fru^{MC} isoforms, perhaps reflecting cooperation or redundancy between the isoforms.

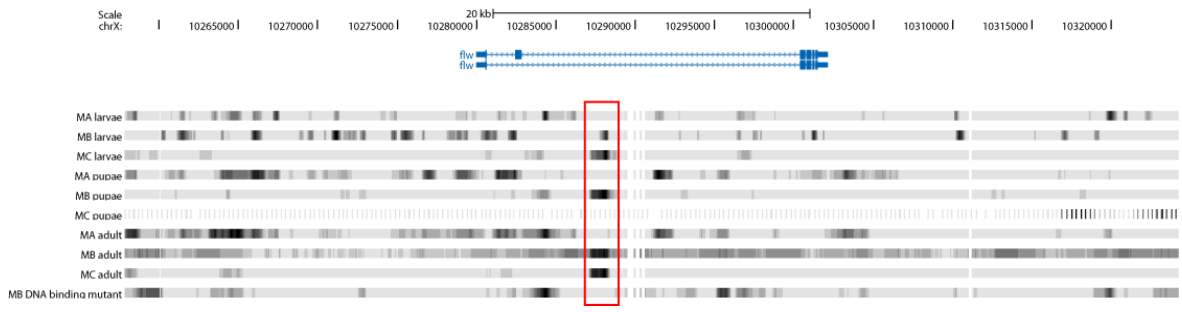


Figure 5-26 DamID binding of *flw*. Intensity mapping of Fru targeting the *flw* genomic region. Where Intensity mapping is on a scale of 1-6 (Fru-Dam/ Dam ratio). The X-axis is the genome in base pairs with the *flw* gene schematic in blue. The red box defines a region of high intensity binding in both Isoforms Fru^{MB} and Fru^{MC}.

5.10.5 *Longitudinals lacking (Lola)*

lola, like *fru*, encodes a range of C_2H_2 BTB ZnF transcription factors, having 26 described transcripts, with suggested roles in the male adult reproductive system [91, 94, 121]. However, due to the complexities of the gene, it is difficult to define which transcripts Fru directly targets. In the DamID experiments, *lola* appears to be targeted at the 5' end by all Fru isoforms (apart from the DNA-binding mutant; Figure 5-27). *lola* has clearly defined functions during early embryonic axonogenesis, however *lola* is also involved in construction or function of neural circuits throughout the lifespan of the fly, including roles in dendritic morphogenesis and the dendritic and axonal development of adult olfactory projection neurons [91]. Moreover, *lola* has been identified in a number of genetic and genomic screens for genes involved in complex behaviours, including locomotion, alcohol sensitivity, mating and aggression [196-199]. The varied roles of *lola* along with its relatedness Fru make it an attractive target for future analysis.

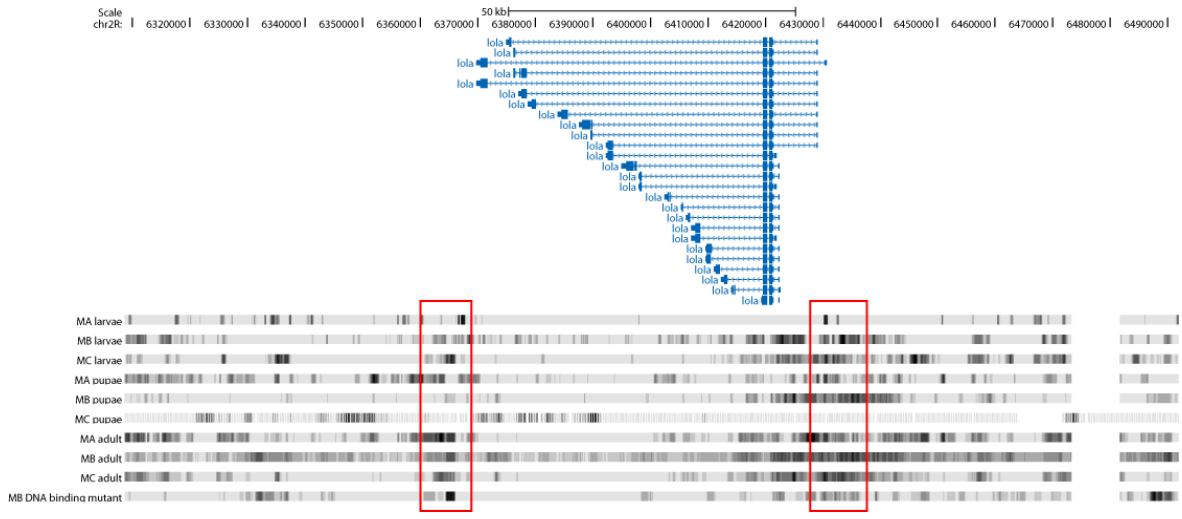


Figure 5-27 DamID Binding of *lola*. Intensity mapping of Fru targeting the *lola* genomic region. Where Intensity mapping is on a scale of 1-6 (Fru-Dam/ Dam ratio). The X-axis is the genome in base pairs with the *lola* gene schematic in blue.. The red boxes define regions of high intensity binding: the left hand box showing a region where all isoforms bar Fru^{MB} and Fru^{MC} pupae are binding (including the DNA binding mutant, suggesting potential for co-factors); and the right hand box shows binding from all isoform bar Fru^{MC} in pupae, in this region the Fru^{mutMB} DNA binding mutant also shows only weak binding.

5.10.6 Tramtrack (Ttk)

Ttk is a key *Drosophila* transcription factor, which is involved in various aspects of development and cell differentiation, including male adult reproductive system development [95, 200]. Ttk proteins have been shown to be transcriptional repressors and inhibitors of neuronal cell fate. *ttk* encodes a BTB C_2H_2 ZnF protein and is the most closely related gene to *fru* in the drosophila genome. The motif analysis of the Fru DamID data has shown a strong enrichment for Ttk-like DNA-binding motifs associated with Fru^{MB} and Fru^{MC} isoforms. As Fru C_2H_2 ZnF motifs are closely related to Ttk, this could reflect the Fru DNA-binding motif, or this could represent Ttk binding as a co-factor associated with Fru. The *ttk* locus is strongly targeted by all Fru isoforms throughout development, apart from Fru^{MA} larvae, Figure 5-28.

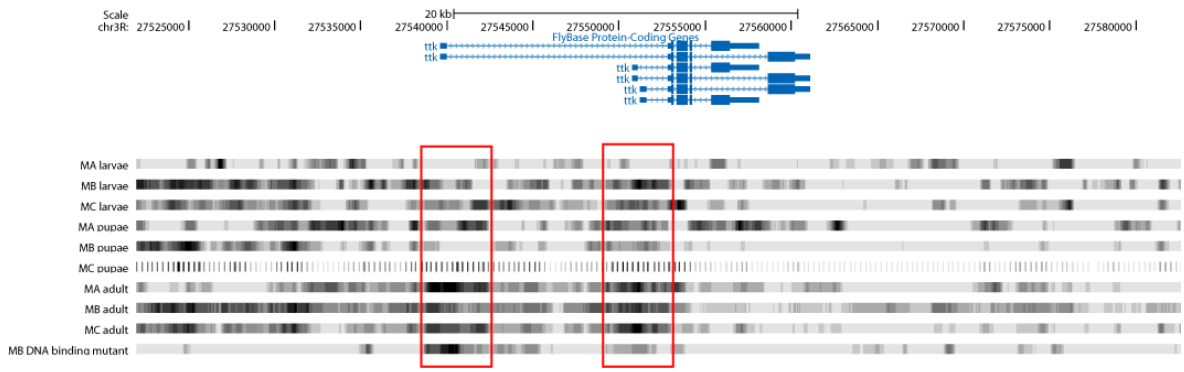


Figure 5-28 DamID binding to *ttk*. Intensity mapping of Fru targeting the *ttk* genomic region. Where Intensity mapping is on a scale of 1-6 (Fru-Dam/ Dam ratio). The X-axis is the genome in base pairs with the *ttk* gene schematic in blue. The red boxes define regions of high intensity binding.

5.10.7 RNAi analysis of fertility

As described above, some initial genes of interest were identified for further analysis. For these genes RNAi lines were identified from one of two sources: the Vienna Drosophila Research Centre, or the Bloomington Stock Centre, Indiana

University. Each of these lines was crossed to the *fru*^{M}-specific GAL4 expressing driver, *fru*^{GAL4}. In addition, UAS-*Dicer-2* was included in these crosses, as it has been shown that co-expressing *Dicer-2* enhances transgenic RNAi effects (although it can also increase off-target effects [201]. To investigate the effects of decreased expression of the gene of interest specifically in *fru*^{M} cells, the RNAi lines were crossed to *fru*^{GAL4} (+*Dicer-2*), CS (+) or themselves (RNAi line), as seen in Table 5-3. As disruption of *fru*^{M} function has been shown to specifically effect male fertility, disrupting the expression of a gene directly regulated by Fru could potentially 'phenocopy' the *fru* mutant phenotype, resulting in decreased fertility. The results of these experiments are seen in Table 5-3.

RNAi line	/*fru*^{GAL4} (+*Dicer-2*)		/+		/RNAi line	
	M	F	M	F	M	F
Flw^{V}	91	95	100	100	100	100
Flw^{B}	95	100	95	86	100	100
Flw^{B1}	97	90	95	100	100	100
Ttk	100	95	100	96	100	100
CG9650^{V}	100	100	96	96	100	100
CG9650^{B}	100	100	95	97	100	100
Chinmo	100	91	100	100	100	100
Kr	100	100	100	100	100	100
Lola^{V}	100	91	95	96	100	100
Lola^{B}	100	86	80^{1}	78^{1}	100^{1}	100^{1}
Pdm2	90	95	90	91	100	100

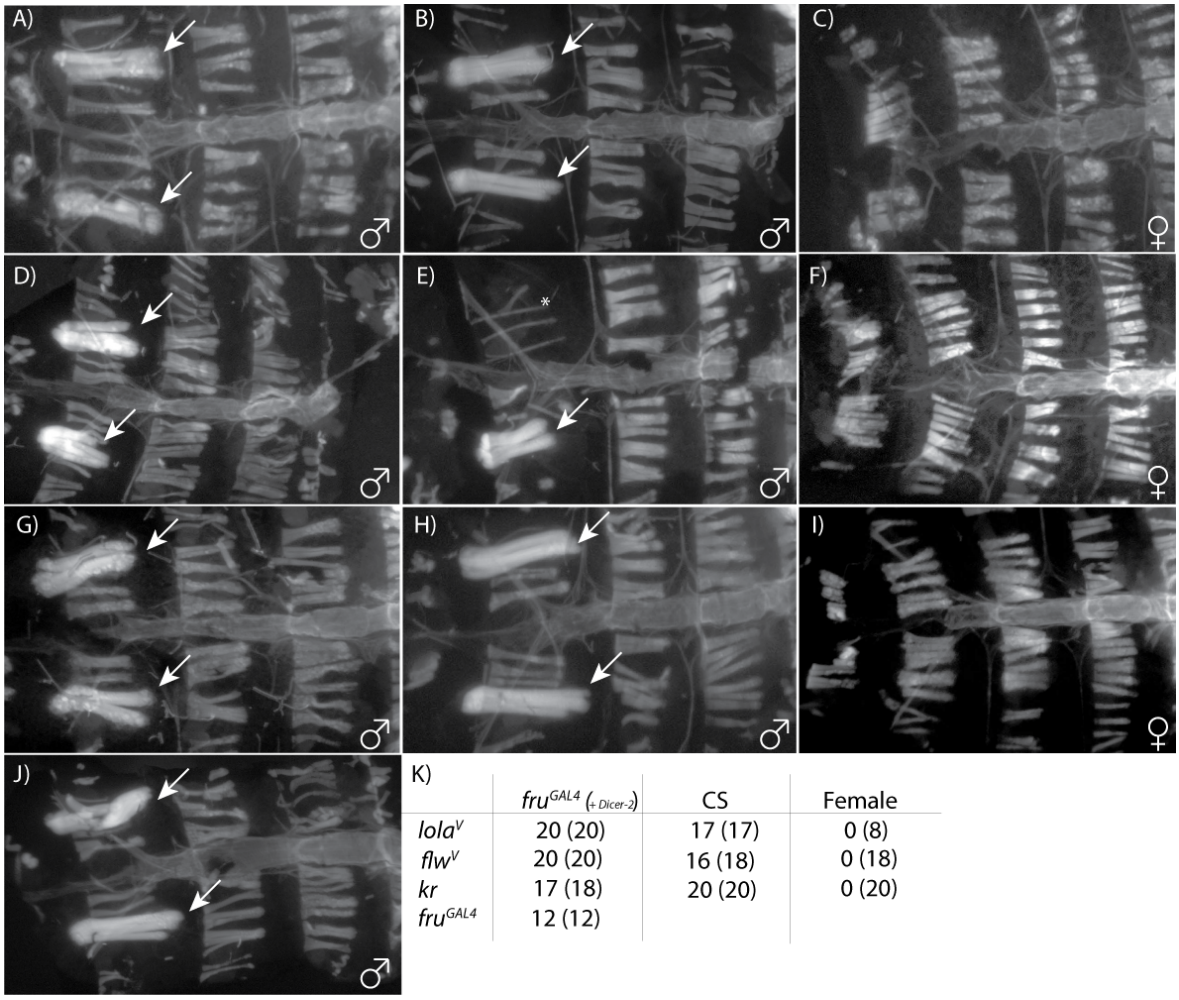
Table 5-3 Fertility data for candidate genes using RNAi. Candidate gene RNAi lines were crossed to *fru*^{GAL4} (*Dicer-2*+) flies and their progeny crossed to WT (CS) and themselves and incubated at 25°C for 10 days. If at this point, progeny were seen in the vial, the cross is classed as fertile. The table shows percentage fertility for an n>20. Where more than one line exists, ^{V} indicates a VDRC originating line, and ^{B} indicates Bloomington origin. ^{1} represents n<20, due to the loss of the Lola^{B} line during the experiment. All results are not significant p<0.05 (Fisher's exact test).

The results of the fertility tests using candidate gene RNAi lines showed few fertility abnormalities, all of which were found to be not statistically significant.

5.10.8 Analysis of the MOL

The candidate genes shown to be targeted by Fru^{C} (*flw, lola, Kr*) were used to observe their effect on the formation of the MOL, a structure known to be formed through Fru^{C} and its downstream targets [2]. RNAi lines for the candidate genes were crossed to fru^{GAL4} (+*Dicer-2*) and dissected at 5 days post eclosion. Their MOL's were stained with Phalloidin, as described previously (2.1.5.4).

Of the RNAi lines tested, only one line showed any potential interesting change in phenotype, Figure 5-29. *flw*-RNAi/ + males in two instances (out of 18) were lacking a MOL (Figure 5-29E). Further, the lack of MOL did not default to the female state, rather, three very thin muscles were formed. This phenotype was not observed when *flw*-RNAi was crossed to fru^{GAL4}, therefore the phenotype seen was most likely associated with a 'leaky' UAS RNAi line. Given a potential connection with Flw and the formation of the MOL, it would be worth considering trying the dissections using fru^{GAL4} (-*Dicer-2*) at 29 °C as well as trying alternative available *flw*-RNAi lines.



	fru^{GAL4} (+ Dicer-2)	CS	Female
$lola^{V}$	20 (20)	17 (17)	0 (8)
flw^{V}	20 (20)	16 (18)	0 (18)
kr	17 (18)	20 (20)	0 (20)
fru^{GAL4}	12 (12)		

Figure 5-29 Formation of a MOL in candidate gene down regulated flies. The formation of the MOL is dependent on the Fru^{MC} isoform alone. Candidate genes associated with the Fru^{MC} isoform were identified and RNAi lines available for each crossed to fru^{GAL4} (+*Dicer-2*). The progeny of these crosses were assayed for MOL formation by staining dissections with Phalloidin, which stains muscle. (A) *lola*/ fru^{GAL4} male with two fully formed MOLs. (B) *lola*/ + male with two fully formed MOLs. (C) *lola*/ fru^{GAL4} female with no MOL formation. (D) *flw*/ fru^{GAL4} male with two fully formed MOLs. (E) *flw*/ + male with one MOL and one structure which is neither a MOL nor the female alternative. (F) *flw*/ fru^{GAL4} female with no MOL formation. (G) *Kr*/ fru^{GAL4} male with two fully formed MOL. (H) *Kr*/ + male with two fully formed MOL. (I) *Kr*/ fru^{GAL4} female with no MOL formation. (J) fru^{GAL}/ fru^{GAL} control showing a fully formed MOL in the absence of RNAi. (K) Numbers of MOL recorded for each phenotype (n in brackets). White arrow indicates a fully formed MOL, * indicates a malformed MOL, which is different to the female state.

5.11 Discussion

The hardest part of genome screening is confirming a direct downstream target. Through DamID, putative downstream targets of Fru have been identified for each isoform, at three developmental time points: L3 ‘wandering’ larvae, 46 hr. pupae and 48 hr. adult. However, each individual experiment has generated hundreds, and in some cases thousands of genes. To identify true targets in these lists a specific, *in vivo* connection with *fru* function must be made.

As has been shown in this chapter, normalised lists of these datasets can be narrowed down using various bioinformatic techniques. Whole dataset analysis such as the tools available through flymine looked for large trends in the data, whereas RNAi and MOL analysis used candidates identified from the trend analyses to find direct targets. Overall, bioinformatic methods identified common and unique elements in each dataset – alluding to the potential functions of each isoform, and motif analysis attempted to identify the sequence to which each isoform specifically binds. By combining these methods, a number of 'highly significant' targets were identified, and taken on for further analysis to try to identify that elusive confirmed target.

Initially, some genes were of obvious interest due to their being unique to a time point or isoform. Identifying the function of these targets might highlight functions specific to an individual isoform. Other genes were identified for spanning many datasets which might identify common functions of Fru.

Flymine was used to look at the datasets and identify functionality. The targets of Fru were found to be equally spread over the chromosomes, but it was via flyatlas that the differences between the isoforms started to emerge. At the larval time point, up-regulated tissues were distinctively development: brain, larval NS and thoracicoabdominal ganglion which continued through pupae for Fru^{MB} and Fru^{MC}, and to the adult in Fru^{MB}. Fru^{MA} showed control of physiological function beyond the larval time point. Fru^{MC} showed a great deal of similarity to Fru^{MB}, but had fewer targets.

These findings were further highlighted in protein domain analysis where Fru^{MA} was enriched for the regulatory immunoglobin domain, and Fru^{MB} and Fru^{MC} were enriched for homeodomains (which bind DNA). When pathway enrichment was

considered, Fru^{MA} initially showed axonal development enrichment, which moved to signalling and regulation from the pupae time point onwards. Comparatively, Fru^{MB} mainly enriched NS related pathways. Fru^{MC} had no enrichments for the adult time point, and for the other two time points, showed enrichment for metabolism and signalling. However, the lack of results may be due to very small datasets (e.g. if only a couple of genes per pathway were present in the dataset, it would not register as enriched, which may regularly occur in a small dataset). To further investigate these patterns, GO was analysed for all datasets. All datasets had similar ontology enrichments, suggesting a commonality in key targets.

The larval time point was highly enriched for NS development ontology groups in all datasets – this was where the isoforms were most similar to each other (this similarity was further enforced in the Venn diagram analysis). Beyond this point the Fru^{MA} isoform became more distinct. It appeared to become more involved in control of physiology. It generally appeared to be more specific, and produced a narrower list of significant genes. Isoforms Fru^{MB} and Fru^{MC} remained NS orientated with several overlapping NS ontologies showing similarity between the isoforms. Although these two isoforms become more regulatory in nature in the mature fly, they remained strongly NS biased.

The pupae time point showed an explosion in the number of gene targets identified, which is in line with pupae being the time point of peak *fru* expression. The exception to this was isoform Fru^{MC} – however the lack of targets for this time point were thought to be due to the older style tiling array used, as this dataset was completed prior to the release of the 2.1 M array and so used a custom design array. The adult time point had the largest overlap between all isoforms,

showing a move towards commonality again, as the fly matures and all isoforms move to a role in controlling physiology of the fly.

The DNA binding mutant DamID analysis allowed the contribution of DNA-binding to Fru-DNA associations to be investigated. Only one DNA binding mutant was analysed, Fru^{mutMB}, which showed that it is still able to associate with DNA, perhaps through co-factor binding or through dimerization with endogenous Fru isoforms through the BTB domain.

Finally, the NSD ontology was analysed as many of Fru's targets appear to be linked to this classification. 59 % of NSD genes were found to be targets of Fru. Interestingly, co-factor binding may be an important contribution to this binding as 127 genes (42 %) of Fru^{MB} targets in the NSD group were bound in the absence of DNA-binding (with Fru^{mutMB}), these included *EcR* and *Robo3*, previously suggested targets of Fru [119].

A key finding of the analysis of Fru-genomic targets was that Fru self-regulates promoters three and four (common promoters) and that it does so throughout development, with each isoform binding at some stage. Although this is currently more difficult to substantiate further, it is a crucial finding.

Motif identification would make a huge difference for Fru understanding, and so was a key aim for these experiments. Collaborating with the Aerts Laboratory, they identified potential Fru DNA-recognition sequences: for Fru^{MA} a motif was enriched that has been shown in the bacterial one-hybrid system to be Fru^{A}-specific, and for Fru^{MB} and Fru^{MC} *ttk* and *Kr*-like motifs were identified. Ttk was identified as most likely representative of a $Fru^{B/C}$ motif. The fact that the same

motif has been identified for both isoforms reinforces the similarity between these two isoforms.

All this data lead to several candidates being taken on for verification as Fru targets. *chinmo, Kr, ttk* and *lola* are all C_2H_2 BTB ZnF proteins like Fru, with functions related to NSD. *chinmo* is a strong target for the larvae time point, and *lola* for the pupae time point. *Kr* and *ttk* were identified by motif analysis. Finally, *pdm2* and *flw* were chosen for their ubiquitous identification; for *pdm2* throughout development with the Fru^{MB} isoform and for *flw* through development with all isoforms. Analysis by RNAi showed no significant fertility results. There are additional RNAi lines available for these candidates and in the future these should be tested for their effects on fertility. Furthermore, fertility is a very general measure of courtship behaviour, in the future these RNAi lines should be tested for there effects on more specific aspects of male courtship behaviour, such as measurements of male courtship indices.

We know more about isoform Fru^{MC} than the other two isoforms due to the existence of an EMS mutant, $Fru^{\Delta C}$ [2]. This study has shown it is the sole isoform involved in the formation of the MOL, it regulates the neuron that innervates the MOL, leading to its formation. Therefore, using RNAi lines for the candidate genes associated with isoform Fru^{C} (*lola, kr, flw*) were expressed specifically in fru^{+} cells. MOL dissections and analysis showed no clear affect on MOL formation due to RNAi knock-down. A possible explanation for this result is the delay in RNAi knockdown using the GAL4/UAS system in *fru* neurons, resulting in a knockdown only after the 'developmental window' for MOL formation has passed. Other GAL4 lines can be used in the future that have been previously shown to effect MOL formation, for example OK371, which is expressed in a subset of

motorneurons in the VNC [202, 203]. As with the fertility analysis, additional RNAi lines targeting these genes can also be tried.

This data is only the beginning of our understanding of the downstream targets of Fru. There are huge steps to be taken to confirm the targets, and fully understand their involvement in the sex-determining hierarchy. This is a long-term project due to the size of the target datasets produced using the DamID approach.

6 Discussion

6.1 Introduction

Prior to this project, many aspects of *fru* function were well defined: it's role in the sex determining hierarchy; its essential role in most aspects of male courtship behaviour; its interation with *dsx* to complete all aspects of male courtship behaviour properly [2, 34, 39, 78, 79, 81, 110]. Within the SDH we knew the genes that regulate *fru*, but not those with which *fru* targets itself, downstream.

This project aimed to build on existing knowledge of *fru*, both by creating mutants of isoforms fru^{A} and fru^{B} to understand the specific functions associated with the individual isoforms, and by completing DamID experiments throughout development to identify putative downstream targets. Through the duration of this study several HR mutants were created: five $fru^{\Delta A}$ mutants and four $fru^{\Delta B}$ mutants. Ten DamID experiments were completed and analysed covering the larvae, pupae and adult time points for each of the three sex-specific isoforms: Fru^{MA}, Fru^{MB}, and Fru^{MC}, as well as the adult time point for Fru^{mutMB}.

The data collected covers the developmental periods of Fru^{M} expression in the CNS, indicating both the individuality and cooperativity of the different isoforms throughout development. The role of Fru in determining sexuality, and defining courtship behaviour in *Drosophila melanogaster* is by no means fully understood. However, the work completed in this study has strengthened our understanding of its isoforms, its function and its interactions with the genes that act downstream of it to coordinate observed phenotypes.

6.2 Homologous recombination mutants define isoform-specific functions of Fru

Creating mutants through homologous recombination fills a gap in the existing knowledge of *fru*. The Goodwin group successfully identified and characterised a fru^{C} isoform mutant in 2006: $fru^{\Delta C}$, which revolutionised our knowledge and understanding of *fru* by being able to associate specific functions with an individual isoform [2]. fru^{C} was identified as being involved in the formation of the MOL, which is found to be missing in $fru^{\Delta C}$ flies. In this work, I further isolated mutants using HR for $fru^{\Delta A}$ and $fru^{\Delta B}$ to fully characterise individual contributions of each isoform to viability and male courtship behaviour.

6.2.1 fru^{A} function is necessary for normal courtship behaviour and wing extension

The HR mutants produced for fru^{A} bore solidly reproducible characteristics. No evidence for a role in viability or fertility was observed.

Lack of the fru^{A} isoform did not result in severe defects in male courtship initiation, time to copulation and copulation duration. However, when CI was considered, a highly significant change in time spent courting was found. Interestingly, $fru^{\Delta A}$ mutants showed abnormal levels of wing extension during courtship, spending approximately 50 % less time extending their wing during courtship. Might this impact on the ability to perform successful song? Although abnormal wing extension does not directly affect some courtship parameters, females potentially respond differently to these males. Prior to copulation, *Drosophila* females must assess a potential mate for species type and fitness

before they sanction mating. This process involves evaluating the courting male through a variety of cues, including courtship song. If the female is receptive she will facilitate copulation by slowing down, ceasing rejection behaviours and opening her vaginal plate. Through collaborators we are investigating any defects associated with courtship song in these mutants. These results might further allude to the small, yet significant contribution of *fru*^{A} transcripts to courtship behaviour.

6.2.2 *fru*^{B} mutants are homozygous lethal and show dramatic courtship defects

The initial data derived from *fru*$^{\Delta B}$ HR mutants was inconsistent. Although initial mutants were partially homozygous viable, post-cantonisation the mutants were found to be completely inviable in the homozygous state. This result suggests the presence of a second-site suppressor either in the original strains used for HR, or the incorporation of a suppressor 'mutation' as result of the HR process itself. Post-cantonisation, viability and fertility assays showed all *fru*$^{\Delta B}$ mutants were homozygous inviable, yet fertile when over a deficiency that removes only the sex-specific isoforms.

The *fru*$^{\Delta B}$ mutant behaviours were seen to be much more severe than *fru*$^{\Delta A}$, and that it showed some similarity with phenotypes seen in the previously isolated *fru*$^{\Delta C}$ mutant (semi-lethality and leg defects). Although missing the *fru*^{MB} isoform did not affect fertility over ten days, *fru*$^{\Delta B}$ mutant males showed severe defects in courtship; they took significantly longer to initiate courtship, and showed a highly significant decrease in time spent performing courtship behaviours (CI), with only 16 % of them successfully copulating within one hour.

Whereas *fru*$^{\Delta A}$ also displayed a drop in general courtship behaviors (decreased CI), *fru*$^{\Delta B}$ was considerably stronger in its effect: a 50 % drop, compared to a 33 % drop, respectively. Due to time constraints, the individual components of male courtship behavior in *fru*$^{\Delta B}$ mutants have yet to be characterised. Analysis of the minutiae of courtship behaviour associated with the *fru*$^{\Delta B}$ mutants is an immediate aim beyond the scope of this project, in fact, work is already underway to this end.

Beyond the sex-specific contribution of the *fru*^{MB} isoform, the effects of missing the common *fru*^{B} isoform was seen in the inability of adults to emerge from their pupal cases. Even when pharate adults were dissected from their pupal casings they survived less than 24 hours. Given the broad expression of the common *fru* isoforms it is unclear what specific defects underly the lethal phenotype [81, 86]. Additionally, dissected pharate adults showed leg defects similar to those seen in the homozygote *fru*$^{\Delta C}$ mutant, along with an inability to inflate their wings. It is interesting to note that a very similar range of phenotypes are seen in mutants that disrupt the role of the neuropeptide bursicon, which plays a role in both cuticle hardening and wing expansion post-eclosion, however was more recently found to play an essential developmental role (pre-eclosion) [204]. *fru*$^{\Delta B}$ mutants do appear to have similar (but often stronger) defects to those observed in *fru*$^{\Delta C}$ mutants, in respect to the roles of the common isoforms (e.g. viability), however, *fru*$^{\Delta C}$ mutants appear to have more pronounced and defined sex-specific defects: MOL formation, fertility defects, and genital formation defects [2].

6.2.3 fru^{MC} is Solely Responsible for the Formation of the MOL

Previously, the fru^{MC} isoform was shown to be involved in the formation of the MOL. When $fru^{\Delta C}$ is not present, no MOL is formed [2]. Until now, there were no null mutants for the remaining isoforms to confirm whether or not they also played a role in MOL formation. Using the HR mutants created in this project, formation of the MOL was assayed, and showed no deviation in the formation of the MOL. Therefore, isoform $fru^{\Delta C}$ is solely responsible for the formation of the male-specific MOL. This result confirms that at least some functions of *fru* are isoform-specific.

6.2.4 Summary of HR mutations

The HR mutants derived from this study fill the gap in our ability to investigate the roles of the individual isoforms of *fru*. Although in this study only certain aspects of their character have been analysed, it is already much clearer which transcripts control which functions of *fru*: isoform fru^{A} is involved in courtship behaviour, particularly wing extension, and may subsequently be shown to be involved in *fru*'s control of song; isoform fru^{B} is likely the most dominant transcript in terms of viability (function derived from the common transcripts) however the sex-specific courtship behaviour; and isoform fru^{C} is solely responsible for MOL formation, successful fertilisation of the female, and also plays a role in viability (which could potentially be a redundancy of the function of the fru^{B} isoform, being less strong in its inhibition). The overlapping functions associated with fru^{B} and fru^{C} suggest a hypothesis that both isoforms stem from a common ancestor. It is possible that the essential functions were retained by the fru^{B} isoform, while the roles played by the fru^{C} isoform developed after its duplication enabling it to take on more sex-specific functions.

The characterisation of these mutants is by no means complete. Further to completing analysis of courtship behaviour and song analysis, there is the potential to try and acquire the elusive $fru^{\Delta A/ B}$ double mutant, which could be a very interesting mutant given the new information attributed to these individual mutants in this body of work. I would hypothesise that this mutant would be homozygous inviable, and given the combined effects of the mutant phenotypes observed in this work, may well be infertile, as the combinatory effects of $fru^{\Delta B}$'s poor courtship abilities, teamed with wing extension issues (that might further effect song) of $fru^{\Delta A}$.

This data exemplifies how complex behaviours can be controlled by a single locus through multiple isoforms regulating both developmental and physiological pathways in different neuronal substrates. The DamID data acquired further highlights the differences and similarities of the transcrips derived from *fru*, and their control over the phenotypes displayed.

6.3 DamID data identifies large datasets of putative targets of Fru, which change through development

The mutants created in this study have the potential to provide insight into the genes that Fru targets, potentially being a useful tool for identifying genes that only interact with specific Fru isoforms. Now having the means to tease apart the roles of each isoform, the downstream effects of Fru were targeted in this study in an attempt to understand the wider reaching effects of Fru, and how its binding of certain genes leads to the displaying of certain phenotypes with which it is associated.

To identify putative downstream targets of Fru, DamID was used in the absence of a suitable antibody. These created a huge amount of data suggesting that, depending on the time point and isoform, Fru binds from 283 downstream targets per isoform per time point to 2147 downstream targets. Recent opinions on transcription factor targeting accept that some complex gene transcripts target hundreds of genes strongly, and thousands in weaker binding complexes, and that successful binding is competitive [169, 173]. However, from an analytical perspective, this created a problem. How to identify a manageable number of candidates from these lists to attempt to prove a direct interaction between them and Fru.

To address these large datasets, a number of bioinformatic techniques were employed which highlighted functional aspects of Fru, as well as prioritising candidate genes.

6.3.1 Downstream targets of Fru share common functions

Gene Ontology showed genes targeted by Fru have a strong connection to development of the NS, ontologies of which are heavily over-represented in all isoforms through the larval time point (such as nervous system development, axon guidance and neurogenesis). Beyond the larval time point, these ontologies are largely restricted to isoforms Fru^{MB} and Fru^{MC}. In fact, isoforms Fru^{MB} and Fru^{MC} show considerable overlap throughout bioinformatics analysis suggesting a strong similarity and, potentially, that they are derived from an ancient common ancestor. This is substantiated in dendrograms of the various analyses: gene list analysis and GO, as well as in the analysis of $fru^{\Delta B}$ and $fru^{\Delta C}$ mutants [2]. Dendrogram analysis shows isoforms Fru^{MB} and Fru^{MC} to be derived from a

common stem, compared to the pupae and adult time points for Fru^{MA}, which stem from a separate lineage.

The commonality between isoforms Fru^{MB} and Fru^{MC} was also observed in protein domain enrichment analysis- with both isoforms showing over representation of homeodomains domains (common to transcription factors), compared to immunoglobin enriched in isoform Fru^{MA} data. Homeodomains are common to transcription factors, whereas immunoglobin domains are more associated with signalling and regulation in a cell, suggesting a key difference between the role of Fru^{MA}, and the other two sex-specific isoforms, Fru^{MB} and Fru^{MC}.

Generally, the data shows Fru^{MA} to be distinct from the other two isoforms studied, and that it has a strong role regulating physiology beyond the larval time point, compared to isoforms Fru^{MB} and Fru^{MC}, which remain strongly associated with genomic targets invovled in the development of the nervous system.

6.3.2 The Isoforms of Fru have Individual Functions

The alternatively spliced isoforms of Fru are one way in which Fru might manage to have such a wide-ranging number of functions throughout development. The individual zinc-finger exons give the potential for them to bind different sequences. Further, via their common BTB domain, they have the potential to bind either to themselves, each other or other proteins (co-factors).

6.3.2.1 Isoform Fru^{MA}

Isoform Fru^{MA} has been suggested to have a more restricted role in the nervous system. This was first suggested by the more restricted patter of expression, as seen by antibody staining, when compared to the broad expression patterns see

for Fru^{MB} and Fru^{MC}; this decresed expression pattern is particularly pronounced in the VNC [2]. Of the assays performed in this body of work, the phenotypes associated with the fru^{A} mutant, although subtler compared to those displayed by other isoforms, still have specific effect on courtship behaviour. Decreased wing extension (and potentially disrupted song production) may make flies deficient in this isoform less competitive in a wild—type environment. This is generally backed up by the DamID data, which shows Fru^{MA} to be quite distinct in comparison to the other isoforms– beyond the larval time point. It shows less links to the NS, and instead appears to have a role in the physiology of the fly: targeting genes involved in regulation – signalling and cell interactions. Furthermore, motif analysis identifying a distinctly different motif association for this isoform. Future studies using the Fru^{MA} DNA-binding mutant for DamID analysis may establish a strong role in co-factor binding. Investigating this will further shed light on the differences between Fru^{MA} and the remaining sex-specific isoforms, Fru^{MB} and Fru^{MC}.

6.3.2.2 Isoform Fru^{MB}

Isoform Fru^{B} appears to show at least some functional overlap with isoform Fru^{C}, furthermore they show a similar expression patterns throughout development. The Fru^{MB} isoform has by far the most genes per dataset, and has been shown to be strongly associated with targets known to play a role in the development of the NS. The DNA-binding mutant analysis at the adult time point showed a 49 % overlap in targets when compared to the wild-type analysis, suggesting that half the genes targeted by the Fru^{MB} adult isoform are through a co-factor. It is possible that these interactions are associated with one role over another (viability or sex behaviour), which could be potentially evaluated as direct interactors

become identified. Whether this action is homodimeric or heterodimeric is unclear. As Fru^{MB} is the most strongly associated with the NS and appears to be involved in the development of the CNS in general with a more specific role in the formation of axons. Its motif is also found to be most similar to the closely related Ttk binding motif.

6.3.2.3 Isoform Fru^{MC}

The Fru^{MC} isoform, as previously discussed, was most clearly understood prior to this body of work, in large part due to the work completed by Billeter et al., who identified and characterised the $fru^{\Delta C}$ mutant [2]. Knowing more about its role in *fru's* function, it was easier to interpret potential candidate targets for this isoform. Assays aimed at indentifying a role for Fru^{MC} targets in MOL formation potentially identifying Kr, which alongside Ttk was identified as having its binding motif highly enriched in Fru^{MB} and Fru^{MC} bound regions of the genome.

6.3.3 Fru binds cooperatively with unspecified proteins to further expand its scope of regulation

Restrictions on time prevented a full set of DNA binding mutant experiments from being completed for this study; only a single time point for isoform Fru^{MB} was completed. This partial data however has already answered an important question: does Fru bind other proteins to target downstream genes? The answer is a definitive yes. However, due to the uncertain nature of the means by which Fru binds other proteins (Figure 6-1), little more could be gleaned as to the type of interactions formed. Venn diagram analysis of gene lists added weight to yeast-2-hybrid data showing the Fru is able to homo-dimerise throught its BTB domain,

therefore various homo/heterodimerisation and co-factor binding scenarios are possible (M. Neville, unpub.).

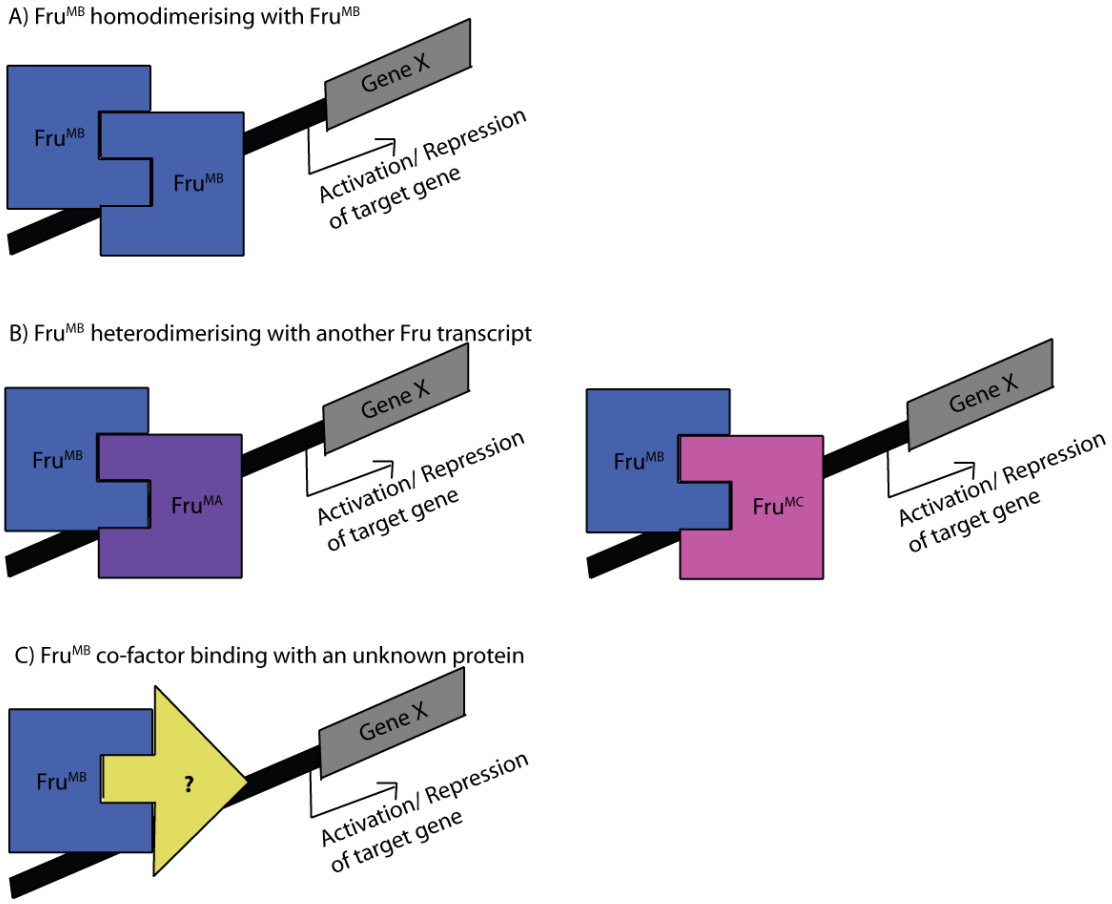


Figure 6-1 Potential Cooperative Binding of Fru^{MB}. Fru^{MB} has been shown to co-factor in the DamID data for Fru^{mutMB}. The method by which it binds is unknown. The above schematics diagram potential methods for binding between Fru^{MB} and (A) itself, (B) other transcripts of *fru* (denoted as the same shape, but different colours which represent different downstream 3´ exons), (C) other unknown proteins (yellow shape, denoted '?'). The arrow represents an initiation of downstream activation or repression of gene transcription.

What was clear from this data was that Fru^{MB}'s indirect interactions with genomic targets (in the absence of binding DNA itself) appear to be highly biased toward genes that play a developmental role. This concept of highly selective binding of individual targets, combined with highly selective binding with co-factoring proteins is in line with the 'widespread DNA binding' theory of binding, suggested by Biggin in 2011 [173]. Biggin suggests that this method works, not by competitive binding of available transcription factors so much, as by nucleosomes and proteins blocking binding sites not used in a given temporal or

spatial time point. This is also consistent with information available on the number of transcription factor molecules per animal cells, which appear to fall within the range of 10,000 to 300,000 molecules per nucleus, suggesting high concentration in all cells [173, 205].

6.3.4 Summary of DamID data

In summary, DamID data has provided a wealth of new information with respect to the binding of Fru with its downstream genomic targets, however the data itself is only the beginning of a very large, long-term project. For this data to be truly meaningful, and beneficial to the field, direct links between Fru and its putative targets must be defined. Potentially, using knowledge gained from the isoforms in this study will allude to which phenotypes is associated with each isoform.

As an initial attempt to achieve these goals, six putative targets were selected for further study. Using RNAi, the down regulation of these genes on the effect of fertility was measured, however no confirmed effect was seen. Given the subsequent ruling out of Isoforms Fru^{MA} and Fru^{MB}'s involvement in fertility, in hindsight these fertility results were unlikely to yield many results. These tests remain a potential way of identifying candidate genes involvement in the downstream processes of Fru^{MC}. Identifying further candidates derived from Fru^{C} datasets and assessing them with the same assay may yet identify downstream targets specifically involved in this aspect of Fru^{MC} function. In terms of the other two isoforms, more detailed behavioural analysis are currently underway that will help define isoform-specific assays that can them be used to test RNAi lines to identify candidate genes: for example, wing extension for Fru^{MA}, and both viability and subsequently identified courtship minutiae for Fru^{B}.

Having shown that isoform Fru^{MC} produces the only transcripts involved in MOL development, RNAi lines derived from isoform Fru^{MC} candidates were tested for MOL formation. Only one candidate showed a potential result: *flw*. Here no MOL formed in two of 18 MOL dissected, however these were only in control animals. Longer-term, a larger sample size is required to show a definitive result. Again, further Fru^{MC} candidates can be assayed in this way to identify genes involved in the downstream specification of the MOL. Although in the time-scale of this study we have been unable to prove a definitive link between Fru and a specific candidate target gene, on going experiments aimed at reaching this goal are currently underway.

6.4 Moving Forward: Using this study as a basis for future investigations

6.4.1 Unanswered Questions: Immediate steps

Behaviour

To further character the isoforms created in this study, there are several obvious steps to take which unfortunately couldn't be completed due to constraints of time. Song analysis should be completed on all the *fru* isoform mutant males, given the observed disruptions to wing extension with the $fru^{\Delta A}$ data and decreased courtship levels seen with $fru^{\Delta B}$ males, it seems likely that a song-related phenotype will be observed. This is currently being analysed by a collaborator, and will hopefully be completed soon. If the sine and pulse waves of song are not correct the female may not recognise the male as being of the same

species, which would go some way to explaining the delayed time to copulation seen in these mutants [39].

The steps to be taken to build on data for $fru^{\Delta B}$ are as yet unclear, however a more detailed behavioural analysis, involving breaking down the individual behaviours that make up the CI are the next step: checking for standard levels of motion, wing extension, licking and tapping. These mutants will also be tested for an increase in male-male courtship, a common feature of *fru* mutants, to define the isoform-specific roles of male-female preference. Further understanding of the isoform-specific aspects of Fru will allude to the direction with which to attempt to characterise the downstream targets identified with Fru.

DamID

The most obvious benefit in terms of the DamID datasets would be to produce the remaining DNA-binding mutant datasets, it would be most prudent to complete adult datasets for the remaining two isoforms, but long-term also taking into account the other two developmental time points for each isoform to identify any changes in co-factor binding levels through development, or whether co-factor binding is a unique feature of the Fru^{MB} isoform.

Strong binding events vs. weak binding events

The data acquired through DamID should be interpreted carefully to most efficiently identify genes with direct links to Fru. This will be a more long-term goal, with a hope that as some links are made, the pattern of progression becomes more apparent, alluding to further significant interactions. Further to this, Biggin (2011) has some interesting thoughts on the significance of target

data identified by DamID style whole-genome screens (in fact, his concepts are largely defined considering ChIP-chip data). He accepts that a small FDR must be used to decrease background noise (albeit that specific FDR changes depend on the quality of the data in question) [173]. Further, he notes that as optimised versions of the assays become available the number of targets identified increase. This was based on papers by Cao [206, 207], and these points were further enforced by studies completed by MacArthur in 2009, which showed that several factors all bound the same regions, and that these numbers increased when stringencies were relaxed [208]. This highlights several important points: primarily that the Fru^{MC} pupae data may well be identifying fewer targets due to it being assayed on an older model chip which was less accurate. In terms of accurately interpreting the Fru^{C} datasets, it would be worth considering re-running the Fru^{C} pupae dataset, to find out whether or not the pupae time point is unusually small in target numbers, or whether this was an artefact of the chip used. Further, it is very important to use a high stringency FDR to increase likelihood of targeting the highest stringency targets, as theoretically all the targets identified might contain the binding sequence, however initially it is more likely there is a higher chance of showing a direct link between Fru and a higher stringency target.

6.4.2 Direction: Understanding the underlying network governing Fru function

Through analysis of this data we have a much better understanding of Fru and the methods by which it produces phenotypes, however several important questions remain unanswered:

- What are the behavioural idiosyncrasies specific to each Fru isoform?

- How can a direct target be shown, given the gene lists derived from DamID?

The idiosyncrasies of individual Fru isoforms shall hopefully be alluded to, following experimental plans discussed in the previous section (6.4.1). In terms of defining direct downstream targets of Fru derived from DamID data, in the short term, showing co-localisation with Fru (using available antibodies), although not showing a direct link, confirms the presence of the two genes at the same time and space. RNAi and over-expressing lines seem the most likely way of showing a direct link: combining such tools with behavioural assays and fertility and viability assays, taking into account information as to which isoforms are putatively linked to the candidates and which assays detect a functional deficit associated with that/ those isoforms.

Long-term, the central aim is to create a network map of the downstream targeting of Fru, showing the overlap leading to common functions, and the cascades of genes linked directly to a specific isoform. There are many ways by which such a network can be defined, and given just the information already collected a network could be created. However, the weight of such interactions would be limited without direct links being identified to add strength to the network formed. Currently, the best means to create a network lies with Cytoscape: an open sourced network analysing tool, which integrates with many programmes including R to amass as much data concerning the genes of the network as possible [209]. Cytoscape has been designed specifically with the intention of be used in the mapping of protein-protein interactions (PPI) and protein-DNA interactions, and allows for manual editing of the weighting applied to each node and its edges.

6.5 Concluding Remarks

We have the tools to understand the isoform-specific aspects of Fru: individual mutants for each isoform, and the ability to screen for the common transcripts (Fru^{com} antibody), and sex-specific transcripts, both as a whole (Fru^{M}) and as individual isoforms ($Fru^{A/\ B/\ C}$). Through analyses with these tools, our understanding of Fru is broad and detailed. We now have the datasets from which to tackle the next aspect of understanding Fru: the elusive direct link between Fru and downstream target genes Although this may take time, there are several strong candidates we are actively pursuing. Hopefully as more targets are defined, an understanding of the way in which Fru functions will start to form with a long-term goal of creating a network of Fru binding to explain the complex idiosyncrasies of Fru.

References

1. Choksi, S.P., et al., *Prospero acts as a binary switch between self-renewal and differentiation in Drosophila neural stem cells.* Developmental Cell, 2006. **11**(6): p. 775-789.

2. Billeter, J.C., et al., *Isoform-specific control of male neuronal differentiation and behavior in Drosophila by the fruitless gene.* Current Biology, 2006. **16**(11): p. 1063-1076.

3. Sokolowski, M.B., *Drosophila: Genetics meets behaviour.* Nature Reviews Genetics, 2001. **2**(11): p. 879-890.

4. Bate, M., *Making sense of behavior.* International Journal of Developmental Biology, 1998. **42**(3): p. 507-509.

5. Billeter, J.C., et al., *Control of male sexual behavior in Drosophila by the sex determination pathway.* Current Biology, 2006. **16**(17): p. R766-R776.

6. Dukas, R., *Evolutionary biology of insect learning*, in *Annual Review of Entomology*. 2008, Annual Reviews: Palo Alto. p. 145-160.

7. Dudai, Y., et al., *Dunce, A Mutant Of Drosophila Deficient In Learning.* Proceedings of the National Academy of Sciences of the United States of America, 1976. **73**(5): p. 1684-1688.

8. Bellen, H.J., C. Tong, and H. Tsuda, *Timeline 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future.* Nature Reviews Neuroscience, 2010. **11**(7): p. 514-+.

9. Vosshall, L.B., *Into the mind of a fly.* Nature, 2007. **450**(7167): p. 193-197.

10. Konopka, R.J. and S. Benzer, *Clock Mutants Of Drosophila-Melanogaster.* Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(9): p. 2112-&.

11. Duffy, J.B., *GAL4 system in Drosophila: A fly geneticist's Swiss army knife.* Genesis, 2002. **34**(1-2): p. 1-15.

12. Ryder, E. and S. Russell, *Transposable elements as tools for genomics and genetics in Drosophila.* Briefings in Functional Genomics & Proteomics, 2003. **2**(1): p. 57-71.

13. Mohr, S. and P. N, *RNAi screening: new approaches, understandings, and organisms*, 2011.

14. van Steensel, B. and S. Henikoff, *Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase.* Nature Biotechnology, 2000. **18**(4): p. 424-428.

15. Ho, J.W.K., et al., *ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis.* Bmc Genomics, 2011. **12**.

16. Holland, S., et al., *Comparison of Dam tagging and chromatin immunoprecipitation as tools for the identification of the binding sites for S-pombe CENP-C.* Chromosome Research, 2005. **13**(1): p. 73-83.

17. Adams, M.D., et al., *The genome sequence of Drosophila melanogaster.* Science, 2000. **287**(5461): p. 2185-2195.

18. Hummel, T. and C. Klaembt, *P-element mutagenesis.* Methods in Molecular Biology, 2008. **420**: p. 97-117.

19. Spradling, A.C. and G.M. Rubin, *Transposition Of Cloned P Elements Into Drosophila Germ Line Chromosomes.* Science, 1982. **218**(4570): p. 341-347.

20. Rubin, G.M. and A.C. Spradling, *Vectors For P-Element-Mediated Gene-Transfer In Drosophila.* Nucleic Acids Research, 1983. **11**(18): p. 6341-6351.

21. Cooley, L., R. Kelley, and A. Spradling, *Insertional Mutagenesis Of The Drosophila Genome With Single P-Elements.* Science, 1988. **239**(4844): p. 1121-1128.

22. Brand, A.H. and N. Perrimon, *Targeted Gene-Expression As A Means Of Altering Cell Fates And Generating Dominant Phenotypes.* Development, 1993. **118**(2): p. 401-415.

23. Jones, W.D., *The expanding reach of the GAL4/UAS system into the behavioral neurobiology of Drosophila.* Bmb Reports, 2009. **42**(11): p. 705-712.

24. Ito, K., J. Urban, and G.M. Technau, *Distribution, Classification, And Development Of Drosophila Glial-Cells In The Late Embryonic And Early Larval Ventral Nerve Cord.* Rouxs Archives of Developmental Biology, 1995. **204**(5): p. 284-307.

25. Yang, J.J., J.S. Kang, and R.S. Krauss, *Transformation-Restoring Factor - A Low-Molecular-Weight Secreted Factor Required For Anchorage-Independent Growth Of Oncogene-Resistant Mutant-Cell Lines.* Oncogene, 1995. **10**(7): p. 1291-1299.

26. Pfeiffer, B.D., et al., *Refinement of Tools for Targeted Gene Expression in Drosophila.* Genetics, 2010. **186**(2): p. 735-U488.

27. Clynen, E., et al., *Conservation of fruitless' role as master regulator of male courtship behaviour from cockroaches to flies.* Development Genes and Evolution, 2011. **221**(1): p. 43-48.

28. Kitamoto, T., *Conditional modification of behavior in drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons.* Journal of Neurobiology, 2001. **47**(2): p. 81-92.

29. Koenig, J.H., K. Saito, and K. Ikeda, *Reversible Control Of Synaptic Transmission In A Single Gene Mutant Of Drosophila-Melanogaster.* Journal of Cell Biology, 1983. **96**(6): p. 1517-1522.

30. Lee, T.M. and L.Q. Luo, *Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development.* Trends in Neurosciences, 2001. **24**(5): p. 251-254.

31. Goto, J., et al., *Sexually dismorphic shaping of interneuron dendrites involves the hunchback transcription factor.* J Neurosci, 2011. **31**(14): p. 5.

32. Claridge-Chang, A., et al., *Writing Memories with Light-Addressable Reinforcement Circuitry.* Cell, 2009. **139**(2): p. 405-415.

33. Villella, A. and J.C. Hall, *Neurogenetics of Courtship and Mating in Drosophila.* Advances in Genetics, Vol 62, 2008. **62**: p. 67-184.

34. Anand, A., et al., *Molecular genetic dissection of the sex-specific and vital functions of the Drosophila melanogaster sex determination gene fruitless.* Genetics, 2001. **158**(4): p. 1569-1595.

35. Hall, J.C., *The Mating Of A Fly.* Science, 1994. **264**(5166): p. 1702-1714.

36. Ferveur, J.F., *Cuticular hydrocarbons: Their evolution and roles in Drosophila pheromonal communication.* Behavior Genetics, 2005. **35**(3): p. 279-295.

37. Gailey, D.A., et al., *Functional conservation of the fruitless male sex-determination gene across 250 Myr of insect evolution.* Molecular Biology and Evolution, 2006. **23**(3): p. 633-643.

38. Von Schilcher, F., *The Role Of Auditory Stimuli In The Courtship Of Drosophila-Melanogaster.* Animal Behaviour, 1976. **24**(1): p. 18-26.

39. Rideout, E.J., J.C. Billeter, and S.F. Goodwin, *The sex-determination genes fruitless and doublesex specify a neural substrate required for courtship song.* Current Biology, 2007. **17**(17): p. 1473-1478.

40. Greenspan, R. and J.-F. Ferveur, *Courtship in Drosophila.* Genetics, 2000. **34**: p. 27.

41. Spieth, H.T. and J.M. Ringo, *Mating and sexual isolation in Drosophila*. The Genetics and Biology of Drosophila, ed. M. Ashburner, H.L. Carson, and J.N. Thompson. Vol. 3c. 1983: Academic Press.

42. Fowler, G.L., *Some Aspects Of The Reproductive Biology Of Drosophila Sperm Transfer Sperm Storage And Sperm Utilization.* Caspari, E.W. (Ed.). Advances in Genetics, Vol. 17. Vii+516p. Illus. Academic Press: New York, N.Y., U.S.a.; London, England, 1973: p. 293-360.

43. Burnet, B. and K.J. Connolly, *Activity and Sexual Behaviour in Drosophila melanogaster*. The Genetics of behaviour, ed. J.H.F. van Abeelen. 1974, Amsterdam.

44. Connolly, K., B. Burnet, and D. Sewell, *Selective Mating And Eye Pigmentation - An Analysis Of Visual Component In Courtship Behavior Of Drosophila-Melanogaster.* Evolution, 1969. **23**(4): p. 548-&.

45. Tompkins, L., et al., *The Role Of Female Movement In The Sexual-Behavior Of Drosophila-Melanogaster.* Behavior Genetics, 1982. **12**(3): p. 295-307.

46. Stocker, R.F., *The Organization Of The Chemosensory System In Drosophila-Melanogaster - A Review.* Cell and Tissue Research, 1994. **275**(1): p. 3-26.

47. Billeter, J.C., S.F. Goodwin, and K.M.C. O'Dell, *Genes mediating sex-specific behaviors in Drosophila*, in *Advances in Genetics, Vol 47*. 2002. p. 87-116.

48. Bennetcl.Hc and A.W. Ewing, *Pulse Interval As A Critical Parameter In Courtship Song Of Drosophila-Melanogaster.* Animal Behaviour, 1969. **17**: p. 755-&.

49. Kyriacou, C.P. and J.C. Hall, *Circadian-Rhythm Mutations In Drosophila-Melanogaster Affect Short-Term Fluctuations In The Males Courtship Song.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1980. **77**(11): p. 6729-6733.

50. Kyriacou, C.P. and J.C. Hall, *The function of courtship song rhythms in drosophila.* Animal Behaviour, 1982. **30**(AUG): p. 794-801.

51. Kyriacou, C.P. and J.C. Hall, *Learning And Memory Mutations Impair Acoustic Priming Of Mating-Behavior In Drosophila.* Nature, 1984. **308**(5954): p. 62-65.

52. Kyriacou, C.P. and J.C. Hall, *Interspecific Genetic-Control Of Courtship Song Production And Reception In Drosophila.* Science, 1986. **232**(4749): p. 494-497.

53. Ewing, A.W. and Bennetcl.Hc, *Courtship Songs Of Drosophila.* Behaviour, 1968. **31**: p. 288-+.

54. Bennetcl.Hc and A.W. Ewing, *Stimuli Provided By Courtship Of Male Drosophila Melanogaster.* Nature, 1967. **215**(5101): p. 669-&.

55. Liu, H.F. and E. Kubli, *Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(17): p. 9929-9933.

56. Avila, F.W., et al., *Insect Seminal Fluid Proteins: Identification and Function.* Annual Review of Entomology, Vol 56, 2011. **56**: p. 21-40.

57. Garcia-Bellido, A., *The Secretion Of Paragonia As A Stimulus For The Fecundity Of Drosophilamelanogaster Females.* Z Naturforsch B. , 1964. **19**: p. 4.

58. Merle, J., *Ovarian function and sexual receptivity of Drosophila melanogaster after implantation of fragments of the male genital tract.* J Insect Physiol., 1968. **14**(8): p. 9.

59. Chen, P.S., *The Functional-Morphology And Biochemistry Of Insect Male Accessory-Glands And Their Secretions.* Annual Review of Entomology, 1984. **29**: p. 233-255.

60. Isaac, R.E., et al., *Angiotensin-converting enzyme as a target for the development of novel insect growth regulators.* Peptides, 2007. **28**(1): p. 153-162.

61. Domanitskaya, E.V., et al., *The hydroxyproline motif of male sex peptide elicits the innate immune response in Drosophila females.* Febs Journal, 2007. **274**(21): p. 5659-5668.

62. Manning, A., *Sperm Factor Affecting Receptivity Of Drosophila Melanogaster Females.* Nature, 1962. **194**(4825): p. 252-&.

63. Tompkins, L., *Evidence That Association Of Excitatory And Inhibitory Sex-Pheromones Modifies The Courtship Of Drosophila-Melanogaster Males.* Behavior Genetics, 1982. **12**(6): p. 599-599.

64. Cook, R., *Courtship Tracking Of Drosophila-Melanogaster.* Biological Cybernetics, 1979. **34**(2): p. 91-106.

65. Cook, R., *The Extent Of Visual Control In The Courtship Tracking Of D Melanogaster.* Biological Cybernetics, 1980. **37**(1): p. 41-51.

66. Joiner, M.L.A. and L.C. Griffith, *CaM kinase II and visual input modulate memory formation in the neuronal circuit controlling courtship conditioning.* Journal of Neuroscience, 1997. **17**(23): p. 9384-9391.

67. Geer, B.W. and M.M. Green, *Genotype, Phenotype And Mating Behavior Of Drosophila-Melanogaster.* American Naturalist, 1962. **96**(888): p. 175-&.

68. Stocker, R.F. and N. Gendre, *Courtship Behavior Of Drosophila Genetically Or Surgically Deprived Of Basiconic Sensilla.* Behavior Genetics, 1989. **19**(3): p. 371-385.

69. Vosshall, L.B. and R.E. Stocker, *Molecular architecture of smell and taste in Drosophila.* Annual Review of Neuroscience, 2007. **30**: p. 505-533.

70. Gailey, D.A., R.C. Lacaillade, and J.C. Hall, *Chemosensory Elements Of Courtship In Normal And Mutant, Olfaction-Deficient Drosophila-Melanogaster.* Behavior Genetics, 1986. **16**(3): p. 375-405.

71. Scott, D., *Inhibition Of Female Drosophila-Melanogaster Remating By A Seminal Fluid Protein (Esterase-6).* Evolution, 1986. **40**(5): p. 1084-1091.

72. Ferveur, J.F. and G. Sureau, *Simultaneous influence on male courtship of stimulatory and inhibitory pheromones produced by live sex-mosaic Drosophila melanogaster.* Proceedings of the Royal Society of London Series B-Biological Sciences, 1996. **263**(1373): p. 967-973.

73. Billeter, J.-C., et al., *Specialized cells tag sexual and species identity in Drosophila melanogaster.* Nature, 2009. **461**(7266): p. 987-U250.

74. Nothiger, R., et al., *Sex Determination In The Germ Line Of Drosophila Depends On Genetic Signals And Inductive Somatic Factors.* Development, 1989. **107**(3): p. 505-518.

75. Goldman, T.D. and M.N. Arbeitman, *Genomic and functional studies of Drosophila sex hierarchy regulated gene expression in adult head and nervous system tissues.* Plos Genetics, 2007. **3**(11): p. 2278-2295.

76. Billeter, J.C. and S.F. Goodwin, *Characterization of Drosophila fruitless-gal4 transgenes reveals expression in male-specific fruitless neurons and innervation of male reproductive structures.* Journal of Comparative Neurology, 2004. **475**(2): p. 270-287.

77. Usui-Aoki, K., et al., *Formation of the male-specific muscle in female Drosophila by ectopic fruitless expression.* Nature Cell Biology, 2000. **2**(8): p. 500-506.

78. Ryner, L.C., et al., *Control of male sexual behavior and sexual orientation in Drosophila by the fruitless gene.* Cell, 1996. **87**(6): p. 1079-1089.

79. Goodwin, S.F., et al., *Aberrant splicing and altered spatial expression patterns in fruitless mutants of Drosophila melanogaster.* Genetics, 2000. **154**(2): p. 725-745.

80. Song, H.J., et al., *The fruitless gene is required for the proper formation of axonal tracts in the embryonic central nervous system of drosophila.* Genetics, 2002. **162**(4): p. 1703-1724.

81. Lee, G., et al., *Spatial, temporal, and sexually dimorphic expression patterns of the fruitless gene in the Drosophila central nervous system.* Journal of Neurobiology, 2000. **43**(4): p. 404-426.

82. Demir, E. and B.J. Dickson, *fruitless splicing specifies male courtship behavior in Drosophila.* Cell, 2005. **121**(5): p. 785-794.

83. Manoli, D.S., et al., *Male-specific fruitless specifies the neural substrates of Drosophila courtship behaviour.* Nature, 2005. **436**(7049): p. 395-400.

84. Rong, Y.S. and K.G. Golic, *Gene targeting by homologous recombination in Drosophila.* Science, 2000. **288**(5473): p. 2013-2018.

85. Lee, G. and J.C. Hall, *Abnormalities of male-specific FRU protein and serotonin expression in the CNS of fruitless mutants in Drosophila.* Journal of Neuroscience, 2001. **21**(2): p. 513-526.

86. Dornan, A.J., D.A. Gailey, and S.F. Goodwin, *GAL4 enhancer trap targeting of the Drosophila sex determination gene fruitless.* Genesis, 2005. **42**(4): p. 236-246.

87. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(16): p. 9440-9445.

88. Bonchuk, A., et al., *Drosophila BTB/POZ Domains of "ttk Group" Can Form Multimers and Selectively Interact with Each Other.* Journal of Molecular Biology, 2011. **412**(3): p. 423-436.

89. Zollman, S., et al., *The Btb Domain, Found Primarily In Zinc-Finger Proteins, Defines An Evolutionarily Conserved Family That Includes Several Developmentally-Regulated Genes In Drosophila.* Proceedings of

the National Academy of Sciences of the United States of America, 1994. **91**(22): p. 10717-10721.

90. Stogios, P.J., et al., *Sequence and structural analysis of BTB domain proteins.* Genome Biology, 2005. **6**(10).

91. Spletter, M.L., et al., *Lola regulates Drosophila olfactory projection neuron identity and targeting specificity.* Neural Development, 2007. **2**.

92. Albagli, O., et al., *The Btb/Poz Domain - A New Protein-Protein Interaction Motif Common To Dna-Binding And Actin-Binding Proteins.* Cell Growth & Differentiation, 1995. **6**(9): p. 1193-1198.

93. Hu, S., et al., *The Drosophila Abrupt Gene Encodes A Btb-Zinc Finger Regulatory Protein That Controls The Specificity Of Neuromuscular Connections.* Genes & Development, 1995. **9**(23): p. 2936-2948.

94. Giniger, E., et al., *Lola Encodes A Putative Transcription Factor Required For Axon Growth And Guidance In Drosophila.* Development, 1994. **120**(6): p. 1385-1398.

95. Bonchuk, A., et al., *Drosophila BTB/POZ Domains of "ttk group" Can Form Multimers and Selectively Interact with Each Other.* Journal of Molecular Biology, 2011: p. 14.

96. Bardwell, V.J. and R. Treisman, *The Poz Domain - A Conserved Protein-Protein Interaction Motif.* Genes & Development, 1994. **8**(14): p. 1664-1677.

97. Dhordain, P., et al., *Corepressor SMRT binds the BTB/POZ repressing domain of the LAZ3/BCL6 oncoprotein.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(20): p. 10762-10767.

98. Dhordain, P. and O. Albagli, *SMRT, a corepressor for both nuclear receptors and the POZ domain-containing oncoproteins, LAZ3 and PLZF.* M S-Medecine Sciences, 1998. **14**(2): p. 219-222.

99. Brayer, K.J. and D.J. Segal, *Keep your fingers off my DNA: Protein-protein interactions mediated by C2H2 zinc finger domains.* Cell Biochemistry and Biophysics, 2008. **50**(3): p. 111-131.

100. Wolfe, S.A., E.I. Ramm, and C.O. Pabo, *Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers.* Structure with Folding & Design, 2000. **8**(7): p. 739-750.

101. Krishna, S.S., I. Majumdar, and N.V. Grishin, *Structural classification of zinc fingers.* Nucleic Acids Research, 2003. **31**(2): p. 532-550.

102. Matthews, J.M. and M. Sunde, *Zinc fingers - Folds for many occasions.* Iubmb Life, 2002. **54**(6): p. 351-355.

103. Wolfe, S.A., L. Nekludova, and C.O. Pabo, *DNA recognition by Cys(2)His(2) zinc finger proteins.* Annual Review of Biophysics and Biomolecular Structure, 2000. **29**: p. 183-212.

104. Dennig, J., et al., *Members Of The Sp Transcription Factor Family Control Transcription From The Uteroglobin Promoter.* Journal of Biological Chemistry, 1995. **270**(21): p. 12737-12744.

105. Lee, J.A., et al., *Transcriptional activity of Sp1 is regulated by molecular interactions between the zinc finger DNA binding domain and the inhibitory domain with corepressors, and this interaction is modulated by MEK.* Journal of Biological Chemistry, 2005. **280**(30): p. 28061-28071.

106. Suzuki, T., et al., *Regulation of interaction of the acetyltransferase region of p300 and the DNA-binding domain of Sp1 on and through DNA binding.* Genes to Cells, 2000. **5**(1): p. 29-41.

107. Villella, A., et al., *Extended reproductive roles of the fruitless gene in Drosophila melanogaster revealed by behavioral analysis of new fru mutants.* Genetics, 1997. **147**(3): p. 1107-1130.

108. Ito, H., et al., *Sexual orientation in Drosophila is altered by the satori mutation in the sex-determination gene fruitless that encodes a zinc finger protein with a BTB domain.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(18): p. 9687-9692.

109. Baker, B.S., B.J. Taylor, and J.C. Hall, *Are complex behaviors specified by dedicated regulatory genes? Reasoning from Drosophila.* Cell, 2001. **105**(1): p. 13-24.

110. Rideout, E.J., et al., *Control of sexual differentiation and behavior by the doublesex gene in Drosophila melanogaster.* Nature Neuroscience, 2010. **13**(4): p. 458-U79.

111. Datta, S.R., et al., *The Drosophila pheromone cVA activates a sexually dimorphic neural circuit.* Nature, 2008. **452**(7186): p. 473-477.

112. Gaunt, M.W. and M.A. Miles, *An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks.* Molecular Biology and Evolution, 2002. **19**(5): p. 748-761.

113. Gailey, D.A., et al., *The muscle of lawrence in Drosophila: A case of repeated evolutionary loss.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(9): p. 4543-4547.

114. Bi, X. and Y.S. Rong, *Genome manipulation by homologous recombination in Drosophila.* Briefings in Functional Genomics & Proteomics, 2003. **2**(2): p. 142-146.

115. Rong, Y.S. and K.G. Golic, *A targeted gene knockout in drosophila.* Genetics, 2001. **157**(3): p. 1307-1312.

116. Hinnen, A., J.B. Hicks, and G.R. Fink, *Transformation of yeast.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(4): p. 1929-1933.

117. Bellaiche, Y., V. Mogila, and N. Perrimon, *I-SceI endonuclease, a new tool for studying DNA double-strand break repair mechanisms in Drosophila.* Genetics, 1999. **152**(3): p. 1037-1044.

118. Thompson, A.J., et al., *Cleavage And Recognition Pattern Of A Double-Strand-Specific Endonuclease (I-Crei) Encoded By The Chloroplast 23s-Ribosomal-Rna Intron Of Chlamydomonas-Reinhardtii.* Gene, 1992. **119**(2): p. 247-251.

119. Dalton, J.E., et al., *Ecdysone Receptor Acts in fruitless-Expressing Neurons to Mediate Drosophila Courtship Behaviors.* Current Biology, 2009. **19**(17): p. 1447-1452.

120. Lebo, M.S., et al., *Somatic, germline and sex hierarchy regulated gene expression during Drosophila metamorphosis.* Bmc Genomics, 2009. **10**.

121. Goeke, S., et al., *Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila.* Nature Neuroscience, 2003. **6**(9): p. 917-924.

122. Pappu, K.S. and S.L. Zipursky, *Axon Guidance: Repulsion and Attraction in Roundabout Ways.* Current Biology, 2010. **20**(9): p. R400-R402.

123. Badenhorst, P., J.T. Finch, and A.A. Travers, *Tramtrack co-operates to prevent inappropriate neural development in Drosophila.* Mechanisms of Development, 2002. **117**(1-2): p. 87-101.

124. van Steensel, B., J. Delrow, and H.J. Bussemaker, *Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(5): p. 2580-2585.

125. Perez-Torrado, R., D. Yamada, and P.A. Defossez, *Born to bind: the BTB protein-protein interaction domain.* Bioessays, 2006. **28**(12): p. 1194-1202.

126. Southall, T.D. and A.H. Brand, *Chromatin profiling in model organisms.* Briefings in Functional Genomics & Proteomics, 2007. **6**(2): p. 133-140.

127. Rong, Y.S., *Gene targeting by homologous recombination: a powerful addition to the genetic arsenal for Drosophila geneticists.* Biochemical and Biophysical Research Communications, 2002. **297**(1): p. 1-5.

128. Bainbridge, S.P. and M. Bownes, *Staging The Metamorphosis Of Drosophila-Melanogaster.* Journal of Embryology and Experimental Morphology, 1981. **66**(DEC): p. 57-80.

129. Hanahan, D., *Studies On Transformation Of Escherichia-Coli With Plasmids.* Journal of Molecular Biology, 1983. **166**(4): p. 557-580.

130. Sambrook, J. and D.W. Russell, *Molecular cloning: A laboratory manual.* Molecular cloning: A laboratory manual, 2001.

131. Kessler, C., *Non radioactive analysis of Biomolecules*. 2 ed. 2000: Springer Laboratory. 78.

132. Hamilton, B.A. and K. Zinn, *From clone to mutant gene.* Methods in Cell Biology; Drosophila melanogaster: Practical uses in cell and molecular biology, 1994. **44**: p. 81-94.

133. Toedling, J., et al., *Ringo - an R/Bioconductor package for analyzing ChIP-chip readouts (vol 8, pg 221, 2007).* Bmc Bioinformatics, 2007. **8**.

134. Castrillon, D.H., et al., *Toward a Molecular Genetic Analysis of Spermatogenesis in.* Genetics, 1993. **155**: p. 16.

135. Gailey, D.A. and J.C. Hall, *Behavior And Cytogenetics Of Fruitless In Drosophila-Melanogaster - Different Courtship Defects Caused By Separate, Closely Linked Lesions.* Genetics, 1989. **121**(4): p. 773-785.

136. Villella, A. and J.C. Hall, *Courtship anomalies caused by doublesex mutations in Drosophila melanogaster.* Genetics, 1996. **143**(1): p. 331-344.

137. Gill, K.S., *A mutation causing abnormal courtship and mating behaviour is males of Drosophila melanogaster.* Am Zool, 1963. **3**(507).

138. Miguel-Aliaga, I., D.W. Allan, and S. Thor, *Independent roles of the Dachshund and Eyes Absent genes in BMP signaling, axon pathfinding and neuronal specification.* Development, 2004. **131**(23): p. 5837-5848.

139. Lawrence, P.A. and P. Johnston, *The Genetic Specification Of Pattern In A Drosophila Muscle.* Cell, 1984. **36**(3): p. 775-782.

140. Dolezal, T., et al., *Genetic analysis of the ADGF multigene family by homologous recombination and gene conversion in Drosophila.* Genetics, 2003. **165**(2): p. 653-666.

141. Deng, C.X. and M.R. Capecchi, *Reexamination Of Gene Targeting Frequency As A Function Of The Extent Of Homology Between The Targeting Vector And The Target Locus.* Molecular and Cellular Biology, 1992. **12**(8): p. 3365-3371.

142. Papadopoulou, B. and C. Dumas, *Parameters controlling the rate of gene targeting frequency in the protozoan parasite Leishmania.* Nucleic Acids Research, 1997. **25**(21): p. 4278-4286.

143. Gloor, G.B., et al., *Targeted Gene Replacement In Drosophila Via P-Element-Induced Gap Repair.* Science, 1991. **253**(5024): p. 1110-1117.

144. Hasty, P., J. Riveraperez, and A. Bradley, *The Length Of Homology Required For Gene Targeting In Embryonic Stem-Cells.* Molecular and Cellular Biology, 1991. **11**(11): p. 5586-5591.

145. Hastings, P.J., et al., *Ends-In Vs Ends-Out Recombination In Yeast.* Genetics, 1993. **135**(4): p. 973-980.

146. Kurada, P. and K. White, *Ras promotes cell survival in Drosophila by downregulating hid expression.* Cell, 1998. **95**(3): p. 319-329.

147. Mansfield, E.S., et al., *Nucleic-Acid Detection Using Nonradioactive Labeling Methods.* Molecular and Cellular Probes, 1995. **9**(3): p. 145-156.

148. Sarup, P., et al., *Candidate Genes Detected in Transcriptome Studies Are Strongly Dependent on Genetic Background.* Plos One, 2011. **6**(1).

149. Roy, S. and C.M. Hart, *Targeted gene replacement by homologous recombination in Drosophila stimulates production of second-site mutations.* Fly, 2010. **4**(1): p. 12-17.

150. Mukai, T., et al., *Mutation Rate And Dominance Of Genes Affecting Viability In Drosophila-Melanogaster.* Genetics, 1972. **72**(2): p. 335-&.

151. Kaufmann, B.P., *Spontaneous Mutation Rate In Drosophila.* American Naturalist, 1947. **81**(796): p. 77-80.

152. *Dripps, I.*, *F. Diao*, and *B. White*, *Neurogenetics and neural development* in *Drosophila Research Council*2011: San Diego, CA.

153. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology.* Nature Reviews Genetics, 2009. **10**(10): p. 669-680.

154. Sandmann, T., J.S. Jakobsen, and E.E.M. Furlong, *ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos.* Nature Protocols, 2006. **1**(6): p. 2839-2855.

155. Negre, N., et al., *A Comprehensive Map of Insulator Elements for the Drosophila Genome.* Plos Genetics, 2010. **6**(1).

156. Slattery, M., et al., *Genome-Wide Tissue-Specific Occupancy of the Hox Protein Ultrabithorax and Hox Cofactor Homothorax in Drosophila.* Plos One, 2011. **6**(4).

157. Alekseyenko, A.A., et al., *High Resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the Male X chromosome.* Genes and Development, 2006. **20**: p. 9.

158. Zeitlinger, J., et al., *Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo.* Genes & Development, 2007. **21**(4): p. 385-390.

159. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nature Biotechnology, 1999. **17**(10): p. 1030-1032.

160. Bianchi-Frias, D., et al., *Hairy transcriptional repression targets and cofactor recruitment in Drosophila.* Plos Biology, 2004. **2**(7): p. 975-990.

161. Solomon, M.J., P.L. Larsen, and A. Varshavsky, *Mapping Protein Dna Interactions Invivo With Formaldehyde - Evidence That Histone-H4 Is Retained On A Highly Transcribed Gene.* Cell, 1988. **53**(6): p. 937-947.

162. Ren, B., et al., *Genome-wide location and function of DNA binding proteins.* Science, 2000. **290**(5500): p. 2306-+.

163. Barras, F. and M.G. Marinus, *The Great Gatc - Dna Methylation In Escherichia-Coli.* Trends in Genetics, 1989. **5**(5): p. 139-143.

165. Greil, F., C. Moorman, and B. van Steensel, *DamID: Mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase.* DNA Microarrays Part a: Array Platforms and Wet-Bench Protocols, 2006. **410**: p. 342-359.

166. Pym, E.G.C., et al., *The homeobox transcription factor Even-skipped regulates acquisition of electrical properties in Drosophila neurons.* Neural Development, 2006. **1**(3).

167. de Wit, E., F. Greil, and B. van Steensel, *Genome-wide HP1 binding in Drosophila: Developmental plasticity and genomic targeting signals.* Genome Research, 2005. **15**(9): p. 1265-1273.

168. van Steensel, B., *Mapping of genetic and epigenetic regulatory networks using microarrays.* Nature Genetics, 2005. **37**: p. S18-S24.

169. Nuzhdin, S.V., A. Rychkova, and M.W. Hahn, *The strength of transcription-factor binding modulates co-variation in transcriptional networks.* Trends Genet, 2011. **26**(2): p. 3.

170. Beyer, A., et al., *Integrated assessment and prediction of transcription factor binding.* Plos Computational Biology, 2006. **2**(6): p. 615-626.

171. Watanabe, K., et al., *Gr39a, a Highly Diversified Gustatory Receptor in Drosophila, has a Role in Sexual Behavior.* Behavior Genetics, 2011. **41**(5): p. 746-753.

172. Bryne, J.C., et al., *JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.* Nucleic Acids Research, 2008. **36**: p. D102-D106.

173. Biggin, M.D., *Animal Transcription Networks as Highly Connected, Quantitative Continua.* Developmental Cell, 2011. **21**(4): p. 611-626.

174. Crawley, M.J., *The R Book*. 2007: Wiley.

175. Lyne, R., et al., *FlyMine: an integrated database for Drosophila and Anopheles genomics.* Genome Biology, 2007. **8**(7).

176. Chintapalli, V.R., J. Wang, and J.A.T. Dow, *Using FlyAtlas to identify better Drosophila melanogaster models of human disease.* Nature Genetics, 2007. **39**(6): p. 715-720.

177. Vogel, C., S.A. Teichmann, and C. Chothia, *The immunoglobulin superfamily in Drosophila melanogaster and Caenorhabditis elegans and the evolution of complexity.* Development, 2003. **130**(25): p. 6317-6328.

178. Kanehisa, M., et al., *KEGG for integration and interpretation of large-scale molecular data sets*, 2011.

179. Matthews, L., et al., *Reactome knowledgebase of human biological pathways and processes.* Nucleic Acids Research, 2009. **37**: p. D619-D622.

180. Croft, D., et al., *Reactome: a database of reactions, pathways and biological processes.* Nucleic Acids Research, 2011. **39**: p. D691-D697.

181. Pappu, K.S., et al., *Robo-3-mediated repulsive interactions guide R8 axons during Drosophila visual system development.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(18): p. 7571-7576.

182. Martin, D., et al., *GOToolBox: functional analysis of gene datasets based on Gene Ontology.* Genome Biology, 2004. **5**(12).

183. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* 2000. **25**: p. 5.

184. Aerts, S., et al., *Robust Target Gene Discovery through Transcriptome Perturbations and Genome-Wide Enhancer Predictions in Drosophila Uncovers a Regulatory Basis for Sensory Specification.* Plos Biology, 2010. **8**(7).

185. Noyes, M.B., et al., *Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites.* Cell, 2008. **133**(7): p. 1277-1289.

186. Zhu, S., et al., *Gradients of the Drosophila Chinmo BTB-zinc finger protein govern neuronal temporal identity.* Cell, 2006. **127**(2): p. 409-422.

187. Isshiki, T., et al., *Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny.* Cell, 2001. **106**(4): p. 511-521.

188. Kambadur, R., et al., *Regulation of POU genes by castor and hunchback establishes layered compartments in the Drosophila CNS.* Genes & Development, 1998. **12**(2): p. 246-260.

189. Tran, K.D., M.R. Miller, and C.Q. Doe, *Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity.* Development, 2010. **137**(9): p. 1421-1430.

190. Corty, M. and W. Grueber, *Co-ordinated control of sensory dendrite morphology and axon targeting by CUT and PDM transcription factors.*, in *Neurobiology of Drosophila Conference*2011, CSH: Cold Spring Harbour, New York. p. 62.

191. Reichert, H., *Drosophila neural stem cells: cell cycle control of self-renewal, differentiation, and termination in brain development.* Results and problems in cell differentiation, 2011. **53**: p. 529-46.

192. Deak, II, *Mutations Of Drosophila-Melanogaster That Affect Muscles.* Journal of Embryology and Experimental Morphology, 1977. **40**(AUG): p. 35-63.

193. Raghavan, S., et al., *Protein phosphatase 1 beta is required far the maintenance of muscle attachments.* Current Biology, 2000. **10**(5): p. 269-272.

194. Sun, Y., et al., *Regulation of somatic myosin activity by Protein Phosphatase 1 beta controls Drosophila oocyte polarization.* Development, 2011. **138**(10): p. 1991-2001.

195. Smith, N.R. and K.E. Prehoda, *Robust Spindle Alignment in Drosophila Neuroblasts by Ultrasensitive Activation of Pins.* Molecular Cell, 2011. **43**(4): p. 540-549.

196. Yamamoto, A., et al., *Neurogenetic networks for startle-induced locomotion in Drosophila melanogaster.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(34): p. 12393-12398.

197. Morozova, T.V., R.R.H. Anholt, and T.F.C. Mackay, *Transcriptional response to alcohol exposure in Drosophila melanogaster.* Genome Biology, 2006. **7**(10).

198. Mackay, T.F.C., et al., *Genetics and genomics of Drosophila mating behavior.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**: p. 6622-6629.

199. Edwards, A.C., et al., *Mutations in many genes affect aggressive behavior in Drosophila melanogaster*, 2009.

200. Boyle, M.J. and C.A. Berg, *Control in time and space: Tramtrack69 cooperates with Notch and Ecdysone to repress ectopic fate and shape changes during Drosophila egg chamber maturation.* Development, 2009. **136**(24): p. 4187-4197.

201. Dietzl, G., et al., *A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila.* Nature, 2007. **448**(7150): p. 151-U1.

202. Mahr, A. and H. Aberle, *The expression pattern of the Drosophila vesicular glutamate transporter: A marker protein for motoneurons and glutamatergic centers in the brain.* Gene Expression Patterns, 2006. **6**(3): p. 299-309.

203. Nojima, T., et al., *Neuronal Synaptic Outputs Determine the Sexual Fate of Postsynaptic Targets.* Current Biology, 2010. **20**(9): p. 836-840.

204. Loveall, B.J. and D.L. Deitcher, *The essential role of bursicon during Drosophila development.* Bmc Developmental Biology, 2010. **10**.

205. Elf, J., G.W. Li, and X.S. Xie, *Probing transcription factor dynamics at the single-molecule level in a living cell.* Science, 2007. **316**(5828): p. 1191-1194.

206. Cao, Y., et al., *Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming.* Developmental Cell, 2010. **18**(4): p. 662-674.

207. Cao, Y., et al., *Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters.* Embo Journal, 2006. **25**(3): p. 502-511.

208. MacArthur, S., et al., *Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions.* Genome Biology, 2009. **10**(7).

209. Shannon, P., et al., *Cytoscape: A software environment for integrated models of biomolecular interaction networks.* Genome Research, 2003. **13**(11): p. 2498-2504.

Appendix 1

Original DamID datasets (See accompanying CD)

Appendix 2

Genelist data for heatmap (See accompanying CD)

Appendix 3

Original Venn Diagram data (See accompanying CD)

Apendix 4

Original GO p-value data (See accompanying CD)