



Developmental and Molecular Analysis of
Sperm in *Drosophila pseudoobscura*

Fiona Rachel Messer

March 2022

This thesis is submitted for the degree of Doctor of Philosophy

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council-funded South West Biosciences Doctoral Training Partnership [Training Grant BB/M009122/1].

Acknowledgements

There are many people, without whom, this work would not have been possible. I would like to first thank my supervisor, Helen White-Cooper, for all of her guidance, knowledge and kindness throughout this project. I would also like to thank Araxi Urrutia and Sonia Lopez de Quinto for their additional supervision on this project, and all of the advice they have given me over the past four years.

I must also thank all of the current and former members of the White-Cooper Lab who have helped me during my time at Cardiff – Katia Jindrich, Rob Mitchell, Shrinivas Dighe, Cristina Fernández Garcia, Dána Jackson, Sabrina Williams, Saurabh Chaudhary, Shannon Cole and Holly Everest. I would like to thank the PIs and researchers of the Cardiff Drosophila Group, and the School of Biosciences research community.

I would particularly like to thank my friends, at Cardiff, and elsewhere. In particular, my invaluable coffee buddies Sean, Nicola and Rakhee, my oldest friend and former housemate Libby, and my lockdown bubble Hannah and Shane.

I would like to especially thank my partner, Robbie, for everything that he has done supporting me throughout my whole PhD. Finally, I would like to thank my family, Mum, Dad and Rowan, who have been a constant source of love and have always been a great support. Thank you.

Summary

Drosophila pseudoobscura produce three distinct sperm morphs: a long fertilising morph, the eusperm, and short and medium non-fertilising morphs, parasperm 1 and 2. Parasperm protect the eusperm from female-derived spermicides in the female reproductive tract.

Drosophila spermatogenesis follows a well-characterised pattern of differentiation, mitosis, meiosis, elongation and individualisation. The majority of transcription of genes whose products are required during meiosis and post-meiosis occurs during the pre-meiotic primary spermatocyte stage.

Prior to this work, little was known regarding the specific molecular and developmental processes required for the production of multiple sperm morphs in *D. pseudoobscura*. I hypothesised that transcriptional variation would be present between sub-sets of primary spermatocyte cysts, which would contribute to development of the sperm morphs. RNA-seq analysis of single spermatocyte cysts showed transcriptional differences between sub-sets of cysts, prior to the onset of meiosis. Over 1000 genes were identified as differentially expressed between primary spermatocyte cysts. RNA-seq analysis of post-meiotic spermatid cysts suggested that transcriptional differences between cyst types are also present during elongation and individualisation, identifying around 1400 genes.

Analysis of cyst RNA-seq data, and subsequent validation by *in situ* hybridisation, revealed differentially expressed genes with potential functions in transcription, spermatogenesis and spermiogenesis, notably components of the testis meiotic arrest complex (tMAC) and the tMAC regulator *kumgang* (*kmg*). A Kmg-GFP fusion revealed that the Kmg protein is also differentially expressed in *D. pseudoobscura* spermatocytes and may contribute to morph differentiation.

I have used immunofluorescence to characterise the structure of the hub and apical proliferation centre in *D. pseudoobscura* testes, and propose an updated model of hub structure in this species. I have also developed *D. pseudoobscura* lines expressing endogenous *cas9*, and describe the results of validation experiments.

In this work, I have identified genetic components contributing to the development of the multiple sperm morphs in *D. pseudoobscura*.

Table of Contents

1	Introduction	1
1.1	<i>Drosophila pseudoobscura</i>, a Sperm Heteromorphic Species	1
1.1.1	What is sperm heteromorphism?	1
1.1.2	Heteromorphic sperm are observed in a diverse range of taxa	1
1.1.2.1	Lepidoptera produce eupyrene and apyrene sperm	2
1.1.2.2	Heteromorphic sperm in Hemiptera vary in chromosome number	2
1.1.2.3	Hymenoptera produce sperm heteromorphic for a coiled head structure	2
1.1.2.4	Coleoptera sperm bundle dimorphism	3
1.1.2.5	Do mammals exhibit sperm heteromorphism?	3
1.1.2.6	Convergent evolution of sperm heteromorphism in Diptera	3
1.1.3	Sperm heteromorphism in <i>Drosophila spp.</i>	4
1.1.3.1	Characterisation of sperm heteromorphism in <i>Drosophila pseudoobscura</i>	6
1.2	Functions of Heteromorphic Sperm	7
1.2.1	Hypotheses for the functions of heteromorphic sperm in <i>D. pseudoobscura</i>	8
1.2.1.1	Parasperm are not fertilisation competent	8
1.2.1.2	Evaluation of the Non-Adaptive Hypothesis: parasperm are adaptive and evolve independently of eusperm	9
1.2.1.3	The Provisioning Hypothesis: parasperm do not provide nutritional provisioning to the female	10
1.2.1.4	The Facilitation Hypothesis: parasperm increase survival of eusperm within the female reproductive tract	10
1.2.1.5	The Competition Hypothesis: parasperm may displace stored sperm but do not appear to delay female re-mating or provide an honest signal of male quality	11
1.2.1.6	Summary of eusperm and parasperm function in <i>D. pseudoobscura</i>	13
1.3	Sperm Development in <i>Drosophila</i>	13
1.3.1	General patterns of sperm development and testis structure in <i>Drosophila spp.</i>	13
1.3.1.1	<i>Drosophila</i> testis structure	14
1.3.1.2	Germline stem cell maintenance and differentiation	15
1.3.1.3	Encapsulation of the gonialblast by cyst cells	16
1.3.1.4	Spermatogonium mitotic divisions	16
1.3.1.5	Spermatocyte cell growth, transcription and meiosis	16
1.3.1.6	Spermatid elongation and individualisation	17
1.3.2	Genetic control of sperm development	18
1.3.2.1	Germline stem cell maintenance is dependent on hub signalling	18

1.3.2.2	Regulation of mitosis and the transition to meiosis in spermatogonia	20
1.3.2.3	Transcription control in spermatocytes	23
1.3.2.3.1	Regulation of transcription in spermatocytes by the testis meiotic arrest complex (tMAC) and testis specific paralogues of the TBP-associated factors	23
1.3.2.3.2	Maintenance of germline specific gene expression by regulation of tMAC	25
1.3.2.4	Transcription and translation in spermatids	27
1.3.2.4.1	Translational repression of spermatocyte expressed genes required in later stages	27
1.3.2.4.2	Post-meiotic transcription in spermatids	27
1.4	Heteromorphic Sperm Development in <i>D. pseudoobscura</i> and the <i>obscura</i> Species Group	28
1.4.1	Current research has not addressed the genetic control of heteromorphic sperm development	29
1.4.2	Evolution of testis genes; a role in heteromorphic sperm development?	30
1.4.2.1	'Out of the Testis' – testis genes are fast evolving	30
1.4.2.2	Duplicated genes in <i>D. pseudoobscura</i>	31
1.4.3	Sperm development as a model system	32
1.5	Techniques for the Study of Sperm Development in <i>Drosophila</i>	32
1.5.1	Genome assemblies and sequencing methods	32
1.5.2	Staining methods can be used to temporally and spatially localise transcripts and proteins	33
1.5.3	Gene editing techniques	33
1.5.4	Genetic tools for investigating gene function in <i>Drosophila</i> spermatogenesis	33
1.6	Aims and Key Findings	35
2	Materials and Methods	37
2.1	Fly Stock Maintenance	37
2.1.1	Laying pots	37
2.2	Fly Work and Molecular Biology Methods	38
2.2.1	Testis dissection	38
2.2.2	Testis RNA extraction and cDNA synthesis	38
2.2.2.1	Testis RNA extraction	38
2.2.2.2	cDNA synthesis from testis RNA	38
2.2.3	DNA extraction	39
2.2.3.1	General DNA extraction method	39
2.2.3.2	Single fly DNA extraction	40
2.2.4	PCR	40
2.2.4.1	GoTaq® G2 Master Mix (Promega)	40
2.2.4.2	Taq DNA Polymerase (NEB)	41

2.2.4.3	Q5® High-Fidelity Master Mix (NEB)	41
2.2.5	PCR purification	42
2.2.6	Restriction digest	43
2.2.7	Agarose gel electrophoresis	44
2.2.8	Agarose Gel Extraction	44
2.2.9	DNA ligation by T4 DNA Ligase (NEB)	45
2.2.10	LB agar plates	45
2.2.11	DH5α competent cell transformation	45
2.2.12	Plasmid extraction and purification	46
2.2.12.1	Miniprep of 1-5mL culture	46
2.2.12.2	Maxiprep of 100-250mL culture	46
2.2.13	Glycerol stocks for long-term storage of transformants	47
2.3	CRISPR Cas9 Gene Editing	47
2.3.1	Development of a <i>D. pseudoobscura cas9</i> integrated line	47
2.3.2	Embryo DNA staining to establish timing of developmental stages	47
2.3.3	Generating a pBluescript construct for <i>cas9</i> cassette assembly	48
2.3.4	piggyBac-3xP3-AmCyan-nos-Cas9 construct assembly	51
2.3.5	piggyBac-3xP3-AmCyan-nos-Cas9 Plasmid Injection in <i>D. pseudoobscura</i>	53
2.3.6	Transgenic adult collection and crosses	53
2.3.7	Inverse PCR and sequencing to determine piggyBac-nos-Cas9 insertion sites	54
2.3.8	<i>cas9</i> expression in <i>D. pseudoobscura</i> ovaries	54
2.3.9	Testing <i>D. pseudoobscura</i> integrated Cas9 by targeting marker genes <i>white</i> and <i>yellow</i>	55
2.3.9.1	gRNA expression construct assembly	56
2.3.9.2	pCFD3-U6:3-gRNA expression vector injection into <i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9	57
2.3.9.3	PCR and high resolution melt analysis (HRMA)	58
2.4	<i>D. pseudoobscura</i> Spermatocyte Cyst RNA Sequencing	60
2.4.1	Spermatocyte cyst dissection	60
2.4.2	Library preparation	60
2.4.3	RNA-Seq 1: Initial sequencing of ten spermatocyte cysts	60
2.4.4	RNA-Seq 2: further spermatocyte cyst sequencing and spermatid cyst sequencing	61
2.4.5	Bioinformatic analysis	61
2.4.6	Statistical analysis: clustering and differential gene expression analysis	62
2.4.7	Identification of target genes for further investigation	62
2.5	<i>In Situ</i> Hybridisation of <i>D. pseudoobscura</i> Testes	62
2.5.1	RNA probe design for <i>in situ</i> hybridisation	62
2.5.2	RNA probe synthesis	69

2.5.3	Dissection and preparation of <i>Drosophila</i> testes	70
2.5.4	<i>In situ</i> hybridisation to target mRNA transcripts in <i>D. pseudoobscura</i> testes	70
2.6	Immunofluorescence and Staining to Characterise <i>D. pseudoobscura</i> Testis Structure	71
2.6.1	Immunofluorescence for comparison of testis germinal proliferation centre structure of <i>D. pseudoobscura</i> and <i>D. melanogaster</i>	71
2.6.2	Dissection and DNA staining of whole testis and seminal vesicle	73
2.6.3	Dissection and DNA staining of mature <i>D. pseudoobscura</i> sperm	73
2.7	Generating Tagged Proteins of Differentially Expressed Genes to Investigate Localisation and Function in <i>D. pseudoobscura</i> Spermatogenesis	74
2.7.1	Kmg protein sequence analysis	74
2.7.2	Protein-GFP fusion for genome insertion with the piggyBac transposon system	74
2.7.2.1	piggyBac Kmg N- and C-terminal GFP Tag Construct Designs and Assembly	74
2.7.2.2	Collection of F0 injection survivors	80
2.7.2.3	Screening for transgenic F1 from F0 survivor crosses	80
2.7.2.4	F2 transgenic collection and strategy for generating stable Kmg-GFP fusion lines	80
3	Developing Integrated <i>cas9</i> Lines for CRISPR <i>Cas9</i> Gene Editing in <i>D. pseudoobscura</i>	81
3.1	Analysis of <i>D. pseudoobscura</i> Embryo Development	82
3.2	Generating <i>D. pseudoobscura</i> Lines Expressing <i>cas9</i> Under the <i>nanos</i> Promotor	86
3.2.1	<i>Cas9</i> construct for genome integration and expression: assembly and verification	86
3.2.2	Embryo injection with piggyBac-nos- <i>Cas9</i> construct	86
3.2.3	Screening and collection of transgenic F1	86
3.2.4	F2 and F3 crosses generated stable <i>D. pseudoobscura</i> 3xP3-AmCyan-nos- <i>Cas9</i> lines	89
3.2.5	Locations of piggyBac insertion in F2 transgenic <i>D. pseudoobscura</i>	91
3.2.5.1	F1 transgenic male 27A	92
3.2.5.2	F1 progeny of transgenic F0 29	92
3.2.5.3	F1 progeny of transgenic F0 49	92
3.2.5.4	F1 transgenic male 55A	93
3.3	Validation of CRISPR <i>Cas9</i> Gene Editing in <i>D. pseudoobscura</i>	94
3.3.1	Confirmation of <i>cas9</i> expression in <i>D. pseudoobscura</i> ovaries	94
3.3.2	Assessment of <i>Cas9</i> function in <i>D. pseudoobscura</i> lines expressing endogenous <i>cas9</i> by targeting <i>white</i> and <i>yellow</i> marker genes for NHEJ mutation	95
3.3.3	High resolution melt analysis of F0 <i>white</i> and <i>yellow</i> gRNA injection survivors	96
4	Gene Expression Analysis of <i>D. pseudoobscura</i> Spermatogenesis	103

4.1	RNA-Seq 1: Differential Gene Expression Analysis of <i>D. pseudoobscura</i> Spermatocyte	
	Cysts (Clontech)	104
4.1.1	Sequence quality	104
4.1.2	Hierarchical cluster analysis of primary spermatocyte cyst transcription data	105
4.1.3	Differential gene expression analysis of spermatocyte cyst transcription data	108
4.2	RNA-Seq 2: Differential Gene Expression Analysis of <i>D. pseudoobscura</i> Spermatocyte	
	Cysts	113
4.2.1	Spermatocyte cyst dissection for cDNA library preparation	113
4.2.2	Sequence quality	114
4.2.3	Hierarchical cluster analysis of primary spermatocyte cyst transcription data	117
4.2.4	Differential gene expression analysis of primary spermatocyte cysts	123
4.3	Spermatid Cyst Differential Gene Expression Analysis	127
4.3.1	Spermatid cyst dissection for cDNA library preparation	127
4.3.2	Sequence quality	128
4.3.3	Hierarchical cluster analysis did not distinguish between long and short spermatid cyst morphs	130
4.3.4	Differential gene expression analysis of long and short spermatid cysts	131
4.4	Comparison of Spermatocyte Cyst RNA-Seq Datasets	133
4.4.1	RNA-seq 1 vs. RNA-seq 2	133
4.4.2	Unique expression in spermatid cysts	134
4.5	Further Investigation of Differential Gene Expression by RNA <i>In Situ</i> Hybridisation	
	Staining of <i>D. pseudoobscura</i> Testes: Prioritisation of Candidate Genes	135
4.5.1	Differential gene expression of <i>GA18735</i> , an orthologue of <i>kumgang</i> .	135
4.5.2	Some components of the testis meiotic arrest complex show differential gene expression.	136
4.5.2.1	<i>GA21345 (achi/vis)</i> differential gene expression is isoform dependent	136
4.5.2.2	<i>D. pseudoobscura Caf1A (GA18051)</i> and <i>Caf1B (GA26389)</i> – A duplicated pair of the <i>D. melanogaster</i> tMAC component; <i>Caf1</i>	139
4.5.2.3	Other tMAC components	140
4.5.3	Transcription factor genes	142
4.5.4	Meiosis genes <i>twine</i> and <i>boule</i> show differential gene expression	143
4.5.5	Axoneme assembly gene <i>Dynein intermediate chain 61B (dic61B)</i>	145
4.5.6	<i>milkah (mil)</i> is required for localisation and shaping of the nucleus in the elongating spermatid	146
4.5.7	Germline stem cell maintenance and proliferation genes	147
4.5.8	Post-meiotic expression in spermatid cysts	148
4.5.9	Splicing Regulators	149

4.5.10	Other genes with known testis functions _____	150
4.5.11	Unknown functions _____	152
4.6	Summary _____	153
5	<i>Localisation of GFP-Tagged Kumgang in D. pseudoobscura Spermatogenesis</i> _____	157
5.1	Protein Sequence Analysis of <i>kumgang</i>, a Regulator of Germline-Specific Gene Expression _____	157
5.2	Assembly of piggyBac-3xP3-AmCyan-Kmg-GFP for Genomic Insertion of Fluorescently Tagged <i>D. pseudoobscura</i> Kumgang _____	159
5.3	Collection of Injection Survivors, Crosses and Screening for Transgenic F1 and F2 _____	163
5.4	Expression of GFP-Tagged Kumgang in Spermatocyte Cysts _____	166
5.4.1	Primary spermatocyte cysts show differential expression of the GFP-tagged Kumgang protein	166
5.4.2	Kmg is translated during the mitotic divisions and is maintained at a high level in some, but not all cysts	170
5.4.3	GFP fluorescence decreased in meiotic cysts and was undetectable in elongating spermatids	173
5.4.4	Lattice lightsheet imaging revealed localisation of Kmg-expressing cysts within the testes _____	177
6	<i>Characterisation of D. pseudoobscura Testis Structure by Immunofluorescence and DNA Staining</i> _____	183
6.1	Immunostaining <i>D. melanogaster</i> and <i>D. pseudoobscura</i> Testes for Hub and Germline Markers _____	183
6.1.1	Comparison between <i>D. melanogaster</i> and <i>D. pseudoobscura</i> germline stem cell populations with germline stem cell marker Vasa _____	183
6.1.2	Comparison between <i>D. melanogaster</i> and <i>D. pseudoobscura</i> hub size and structure by immunostaining for Fasciclin III _____	185
6.1.3	Further investigation of <i>D. pseudoobscura</i> testis hub structure by immunostaining for hub markers _____	187
6.1.3.1	Neural Cadherin showed a large hub structure in <i>D. pseudoobscura</i> testes _____	187
6.1.3.2	Escargot indicated the presence of substructures within the <i>D. pseudoobscura</i> testis hub region	188
6.1.3.3	Combined immunostaining for Hu-li tai shao and DE-Cadherin _____	189
6.1.3.4	Armadillo and Discs large are not markers of the <i>D. pseudoobscura</i> testis hub _____	191
6.1.4	Double staining for adhesins and FasIII did not reveal hub sub-structures _____	193
6.2	DNA Staining of <i>D. pseudoobscura</i> Seminal Vesicle and Ejaculatory Duct _____	196
6.3	Summary _____	198

7	Discussion	200
7.1	Project Overview: Prior State of Knowledge	200
7.1.1	Functions of heteromorphic sperm in <i>D. pseudoobscura</i>	200
7.1.2	Sperm development in <i>Drosophila</i>	200
7.1.3	Transcription in <i>D. melanogaster</i> spermatogenesis	201
7.1.4	Questions raised and project aims	202
7.2	Genetic Tools Are Lacking in <i>D. pseudoobscura</i>	203
7.2.1	Development of <i>D. pseudoobscura</i> Cas9 lines	203
7.2.2	Mutagenesis was unsuccessful in <i>D. pseudoobscura</i> Cas9 lines	204
7.2.3	Alternative approaches for CRISPR/Cas9 gene editing in <i>D. pseudoobscura</i>	206
7.3	Characterisation of Gene Expression in <i>D. pseudoobscura</i> Spermatocytes and Spermatids	206
7.3.1	RNA-seq of primary spermatocyte and spermatid cysts	207
7.3.1.1	Two alternative methods for library preparation for low-input RNA-seq	207
7.3.1.2	RNA-seq of primary spermatocyte cysts identified transcriptome variation prior to meiosis	207
7.3.1.3	RNA-seq of elongating spermatid cysts shows transcriptome variation relating to spermiogenesis and post-meiotic transcription	209
7.3.1.4	Limitations of single cyst RNA-seq analysis	210
7.3.1.5	Improvements to RNA-seq analysis based on the developments in this work	211
7.3.2	Validation of RNA-seq data by RNA <i>in situ</i> hybridisation	212
7.3.2.1	Genes identified by RNA-seq were expressed in the testes	213
7.3.2.2	Identification of genes of interest in heteromorphic sperm development; a role for tMAC and tMAC regulation	213
7.3.2.2.1	<i>D. pseudoobscura kmg, GA18735</i> , is differentially expressed in spermatocyte cysts	213
7.3.2.2.2	<i>GA21345</i> , the <i>D. pseudoobscura</i> orthologue of the <i>D. melanogaster achi</i> and <i>vis</i> paralogues, shows isoform dependent differential expression in spermatocyte cysts	214
7.3.2.2.3	<i>D. pseudoobscura</i> orthologues of <i>Caf1, GA18051</i> and <i>GA26389</i> , show stage-specific transcript localisation and morph-specific enrichment in spermatocyte cysts	214
7.3.2.2.4	The Multiple tMACs Hypothesis: do Caf1 paralogues and Achi/Vis isoforms contribute to morph-specific versions of tMAC?	215
7.3.2.3	Meiosis regulators – morph-specific or stage-specific expression?	217
7.3.2.4	Spermiogenesis Genes: eusperm and parasperm-specific morphogenesis?	218
7.3.2.5	Limitations of inferring functions between species	220
7.3.2.6	Examples of genes which may be post-meiotically expressed in <i>D. pseudoobscura</i>	220

7.4	Further Investigation of <i>D. pseudoobscura</i> kmg by piggyBac Insertion of GFP-Tagged Kmg for Endogenous Expression	221
7.4.1	Analysis of GFP-tagged Kmg expression in spermatocyte cysts confirms differential transcription and translation of <i>D. pseudoobscura</i> kmg	221
7.4.2	<i>D. melanogaster</i> and <i>D. pseudoobscura</i> Kmg alignments show similar protein structure	222
7.4.3	Hypotheses for Kmg functions in <i>D. pseudoobscura</i> heteromorphic spermatogenesis	223
7.4.3.1	Hypothesis 1A: Kmg is more abundant in eusperm, and blocks transcription.	223
7.4.3.2	Hypothesis 1B: Kmg is more abundant in eusperm, and promotes transcription.	224
7.4.3.3	Hypothesis 2A: Kmg is more abundant in parasperm, and blocks transcription.	224
7.4.3.4	Hypothesis 2B: Kmg is more abundant in parasperm, and promotes transcription.	224
7.4.3.5	<i>D. pseudoobscura</i> Kmg and the Multiple tMACs Hypothesis	224
7.5	Comparison of Hub Morphology between <i>D. pseudoobscura</i> and <i>D. melanogaster</i>	227
7.5.1	The ‘General Hubby Area’ model of the <i>D. pseudoobscura</i> testis niche	227
7.5.2	Sub-Hubs, multiple GSCs, and cyst cell contributions to heteromorphic sperm development?	229
7.6	Summary: How does <i>D. pseudoobscura</i> control production of multiple sperm morphs?	229
7.7	Future Directions	230

List of Figures

Figure 1.1: Phylogeny of Diptera based on morphological evidence, reprinted from Lambkin et al. (2013)	4
Figure 1.2: Model of testes structure in (A) <i>D. melanogaster</i> and (B) <i>D. pseudoobscura</i>	15
Figure 1.3: Overview of JAK-STAT and BMP signalling pathways in the <i>Drosophila</i> testis hub, based on de Cuevas and Matunis (2011)	20
Figure 1.4: Overview of spermatogonia mitotic divisions of the transit amplifying stage prior to differentiation of the primary spermatocytes	22
Figure 1.5: Models of transcriptional control in the primary spermatocytes	26
Figure 2.1: pBluescript (KS+) construct design for the assembly of a nos-Cas9 cassette into pBluescript prior to ligation into piggyBac	50
Figure 2.2: (A) piggyBac-3xP3-AmCyan-sv40-FRTlox-AeHex-ttAV2-K10 plasmid (from David Navarro) used to retrieve the piggyBac-3xP3-AmCyan backbone for ligation with nos-cas9. (B) piggyBac-3xP3-AmCyan-nos-Cas9 construct design for the integration of cas9 into the <i>D. pseudoobscura</i> genome under the nos germline promotor with 3xP3-AmCyan fluorescent eye marker	52
Figure 2.3: Construct designs for piggyBac mediated insertion of GFP tagged Kmg in <i>D. pseudoobscura</i>	77
Figure 2.4: GFP PCR product in pGEM®-T Easy (Promega) for construction of piggyBac GFP tag constructs	78
Figure 3.1: <i>D. pseudoobscura</i> SLOB3 wild type embryos stained for DNA (Hoechst 33258). Embryos were fixed at regular time points to establish timing of development stages	85
Figure 3.2: <i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9 and <i>D. melanogaster</i> AeHex1g-tTAV2 exhibiting cyan fluorescence from AmCyan eye marker	88
Figure 3.3: Crosses to generate stable transgenic <i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9 lines. Cyan eye indicates transgenic	91
Figure 3.4: Reverse transcription PCR products of cas9 mRNA expressed in ovaries of transgenic <i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9 shown on agarose gel	95
Figure 3.5: Examples of normalised melt curves from HRMA of yellow gene in <i>D. pseudoobscura</i> injected with Yellow 1 gRNA	97
Figure 4.1: Hierarchical cluster analysis of <i>D. pseudoobscura</i> A primary spermatocyte cyst normalised transcription data	107
Figure 4.2: Hierarchical cluster analysis of <i>D. pseudoobscura</i> A primary spermatocyte cyst normalised transcription data	108
Figure 4.3: Primary spermatocyte cysts dissected from two <i>D. pseudoobscura</i> SLOB3 wild-type males. Cysts were collected into individual tubes and RNA extracted for preparation of cDNA libraries	114
Figure 4.4: Hierarchical cluster analysis of <i>D. pseudoobscura</i> M primary spermatocyte cyst normalised transcription data	119
Figure 4.5: <i>D. pseudoobscura</i> M primary spermatocyte cyst normalised transcription data hierarchical cluster analysis	120

Figure 4.6: <i>D. pseudoobscura</i> R primary spermatocyte cyst normalised transcription data hierarchical cluster analysis	121
Figure 4.7: <i>D. pseudoobscura</i> R primary spermatocyte cyst normalised transcription data hierarchical cluster analysis	122
Figure 4.8: Hierarchical cluster analysis of combined M and R primary spermatocyte cyst normalised transcription data	123
Figure 4.9: Spermatid cysts dissected from six <i>D. pseudoobscura</i> SLOB3 wild-type males. Cysts were collected into individual tubes and RNA extracted for preparation of cDNA libraries	128
Figure 4.10: Hierarchical cluster analysis dendrogram of <i>D. pseudoobscura</i> SLOB3 wild type spermatid cyst RNA sequencing data	131
Figure 4.11: Hierarchical cluster analysis dendrogram of <i>D. pseudoobscura</i> SLOB3 wild type spermatid cyst RNA sequencing data	131
Figure 4.12 A-E: In situ hybridisation of GA18735, orthologous to the <i>D. melanogaster</i> gene <i>kumgang</i> (<i>kmg</i>)	136
Figure 4.13: Alternative splicing of <i>D. melanogaster</i> <i>achintya</i> (<i>achi</i>) and <i>vismay</i> (<i>vis</i>), a duplicate pair of meiotic arrest genes, components of the testis meiotic arrest complex (<i>tMAC</i>)	138
Figure 4.14: In situ hybridisation of GA21345, orthologous to the <i>D. melanogaster</i> genes <i>achintya</i> and <i>vismay</i> (<i>achi/vis</i>)	139
Figure 4.15: In situ hybridisation of GA18051 and GA26389, orthologous to the <i>D. melanogaster</i> gene Chromatin assembly factor 1 (<i>Caf1</i>)	140
Figure 4.16: In situ hybridisation of <i>tMAC</i> component genes, orthologous to <i>D. melanogaster</i> <i>tMAC</i> genes	141
Figure 4.17: In situ hybridisation of transcription factor gene	143
Figure 4.18: In situ hybridisation of GA18412, orthologous to <i>D. melanogaster</i> meiosis gene <i>boule</i> (<i>bol</i>)	144
Figure 4.19: In situ hybridisation of GA18558, orthologous to <i>D. melanogaster</i> meiosis gene <i>twine</i> (<i>twe</i>)	145
Figure 4.20: In situ hybridisation of GA20060, orthologous to <i>D. melanogaster</i> spermatogenesis gene Dynein intermediate chain at 61B (<i>Dic61B</i>)	146
Figure 4.21: In situ hybridisation of GA27003, orthologous to <i>D. melanogaster</i> spermatogenesis gene <i>milkah</i> (<i>mil</i>)	147
Figure 4.22: In situ hybridisation of germline stem cell maintenance genes	148
Figure 4.23: In situ hybridisation of spermiogenesis genes	149
Figure 4.24: In situ hybridisation of spliceosome genes	150
Figure 4.25: A: GA28347, <i>exuperantia 2</i> (<i>exu2</i>). B: GA18272, Nucleoporin 154 (<i>Nup154</i>). C: GA19370, P-element induced wimpy testis (<i>piwi</i>)	152
Figure 4.26: In situ hybridisation of GA23025, a <i>D. pseudoobscura</i> gene with no known orthologue in <i>D. melanogaster</i>	152
Figure 5.1: Alignment and analysis of <i>D. melanogaster</i> and <i>D. pseudoobscura</i> <i>Kmg</i> protein sequences	158
Figure 5.2: <i>Kumgang</i> -GFP C-terminal tag construct components in pGEM-T Easy, for subsequent ligation into piggyBac vector	160

Figure 5.3: Kumgang-GFP N-terminal tag construct components in pGEM-T Easy, for subsequent ligation into piggyBac vector	161
Figure 5.4: Construct assemblies and sequence alignments to predicted construct sequence for piggyBac-3xP3-AmCyan-Kmg-GFP C- and N-terminal constructs	162
Figure 5.5: Mutations present in <i>D. pseudoobscura</i> kmg sequence for piggyBac-3xP3-AmCyan-Kmg-GFP C terminal tag construct	163
Figure 5.6: Primary spermatocyte cysts dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735)	167
Figure 5.7: Primary spermatocyte cysts dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735) (green) and Hoechst 33342 indicating DNA (blue)	170
Figure 5.8: Spermatogonia and spermatocyte cysts dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy. Cysts shown at stages through spermatogonia mitotic divisions	171
Figure 5.9: Spermatogonia and spermatocyte cysts dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735)	172
Figure 5.10: Primary spermatocyte to spermatid stages dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-3xP3-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735)	175
Figure 5.11: Early spermatid cysts dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-3xP3-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy	176
Figure 5.12: Spermatid cysts dissected from <i>D. pseudoobscura</i> SLOB3 piggyBac-3xP3-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy	177
Figure 5.13: Lattice Lightsheet 7 (Zeiss) maximum intensity image showing GFP expression of Kumgang-GFP fusion in whole <i>D. pseudoobscura</i> Kmg-GFP testis	178
Figure 5.14: Lattice Lightsheet 7 (Zeiss) maximum intensity image showing GFP expression of Kumgang-GFP fusion in whole <i>D. pseudoobscura</i> Kmg-GFP testis	180
Figure 6.1: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub and germline	184
Figure 6.2: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub stained for FasIII and counterstained for DNA	186
Figure 6.3: <i>D. pseudoobscura</i> and <i>D. melanogaster</i> testes imaged by differential interference contrast microscopy, immunostained for FasIII	186
Figure 6.4: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testes immunostained for Neural Cadherin (N-Cad) and counterstained for DNA	187
Figure 6.5: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub immunostained for Escargot (Esg)	189
Figure 6.6: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub immunostained for Hu-li tai shao (Hts) (green) and DE-Cadherin (DE-Cad) (blue)	190

Figure 6.7: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub immunostained for Hu-li tai shao (Hts) (cyan) and counterstained for DNA (red)	191
Figure 6.8: <i>D. pseudoobscura</i> primary spermatocyte cysts immunostained for Hu-li tai shao (Hts) (cyan) and counterstained for DNA (red)	191
Figure 6.9: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub immunostained for Armadillo (Arm) (cyan) and counterstained for DNA (red)	192
Figure 6.10: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub immunostained for Discs large (Dlg) (cyan) and counterstained for DNA (red)	193
Figure 6.11: A: <i>D. pseudoobscura</i> testis hub immunostained for Fasciclin III (FasIII) (green), DE-Cadherin (DE-Cad) (red) and counterstained for DNA (blue)	194
Figure 6.12: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> testis hub immunostained for N-Cadherin (N-Cad) (red), Fasciclin III (green) and counterstained for DNA (blue)	196
Figure 6.13: A-B: <i>D. melanogaster</i> and <i>D. pseudoobscura</i> seminal vesicle stained for DNA (Hoechst 33258)	197
Figure 6.14: Sperm released from <i>D. pseudoobscura</i> seminal vesicle dissected in testis buffer with 2µg/mL Hoechst 33342	198
Figure 7.1: The Multiple tMAC Hypothesis: Caf1A and Caf1B paralogues and achi/vis alternate isoforms may contribute to multiple versions of tMAC	217
Figure 7.2: Nuclei of mature sperm extracted from <i>D. pseudoobscura</i> seminal vesicles, stained for DNA with Hoechst 33258	220
Figure 7.3: Models for potential interactions between Kmg and tMAC in <i>D. pseudoobscura</i> , under a 'Multiple tMAC' model	226
Figure 7.4: An updated model of the <i>D. pseudoobscura</i> testis hub structure	228

List of Tables

Table 1.1: Summary of sperm heteromorphism in <i>Drosophila</i> species	6
Table 1.2: Post-meiotically expressed comet and cup genes. Table from Barreau et al. (2008a)	28
Table 2.1: Components required for fly food. All flies were maintained on food in vials or bottles	37
Table 2.2: Template mixture for cDNA synthesis from testis RNA	39
Table 2.3: Reaction mix for cDNA synthesis from testis RNA	39
Table 2.4: Components required for PCR reaction mix using GoTaq® Master Mix (Promega)	40
Table 2.5: Reaction conditions for GoTaq® Master Mix (Promega) PCR	41
Table 2.6: Components required for PCR reaction mix using Taq DNA Polymerase (NEB)	41
Table 2.7: Reaction conditions for Taq DNA Polymerase (NEB) PCR	41
Table 2.8: Components required for PCR reaction mix using Q5® High-Fidelity Master Mix (NEB)	42
Table 2.9: Reaction conditions for Q5® High-Fidelity 2X Master Mix (NEB) PCR	42
Table 2.10: Reaction mix for restriction digestion with either one or two restriction enzymes in the same buffer (New England BioLabs)	43

Table 2.11: Restriction endonucleases, conditions for restriction digest of DNA (buffers, incubation temperature), and method of enzyme inactivation (heat inactivation, enzyme removal)	44
Table 2.12: Components required for 20 μ L reaction volume ligation. Vector and insert volumes were calculated according to molecular weight and concentration	45
Table 2.13: Primers for amplification of nos promotor + 5' UTR and 3' UTR fragments from synthesised DNA extended to include restriction sites EcoRI on the 3' end of nos 5' UTR and HindIII on the 5' end of nos 3' UTR, enabling ligation to cas9 digested with EcoRI and HindIII	49
Table 2.14: Ligation reaction with T4 DNA Ligase combining nos 5' UTR, cas9 and nos 3' UTR into pBluescript (KS+)	50
Table 2.15: Primers for inverse PCR and product sequencing to identify the location of piggyBac insertion sites. Primer sequences were retrieved from (Labbe et al. 2010)	54
Table 2.16: Primers for ovary cDNA PCR of cas9 expressed in transgenic <i>D. pseudoobscura</i> lines	55
Table 2.17: Guide RNA oligo sequences for targeting white and yellow marker genes in <i>D. pseudoobscura</i>	55
Table 2.18: Reagents required for annealing reaction mix, to anneal gRNA sense and anti-sense oligos prior to ligation into pCFD3 expression vectors (Port et al. 2014)	56
Table 2.19: gRNA oligo annealing thermocycler program	56
Table 2.20: Reagents required for ligation of annealed gRNA oligos into pCFD3-U6:3 expression vector	57
Table 2.21: <i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9 lines tested by injection of gRNA expression constructs targeting white and yellow markers	58
Table 2.22: F1 offspring genotypes from F0 transgenic crosses for white or yellow markers	58
Table 2.23: Primer sequences for PCR and melt analysis of white and yellow gRNA target sites	59
Table 2.24: Reagents required for high resolution melting master mix (Roche) for HRMA of white and yellow gRNA target sites	59
Table 2.25: LightCycler 96 program for high resolution melt analysis of white and yellow gRNA target sites	59
Table 2.26: Genes investigated by in situ hybridisation. Candidate genes were identified as differentially expressed in RNA-seq of spermatocyte and spermatid cysts	68
Table 2.27: Reagent required for RNA probe synthesis mix for in situ hybridisation	69
Table 2.28: Primary antibodies used in immunohistochemistry of squashed testes, primary antibody dilutions and host species	73
Table 2.29: Secondary antibodies and dilutions for immunohistochemistry	73
Table 2.30: Primer sequences for PCR of kmg and GFP fragments for GFP fusion constructs assembled in piggyBac	79
Table 2.31: Restriction sites for ligation of kmg fragments and GFP sequence into piggyBac vector	80
Table 2.32: Primer sequences for sequencing gene fragments and purified ligation products	80
Table 3.1: <i>D. pseudoobscura</i> piggyBac 3xP3-AmCyan-nos-cas9 F0 injection survivors collected which subsequently produced transgenic F1 progeny	87
Table 3.2: Progeny from F1 transgenic crosses. Progeny were screened for 3xP3-AmCyan eye marker	90
Table 3.3: <i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9 transgenic F1 inverse PCR sequence BLAST hits	94

Table 3.4: F0 adult survivors of white and yellow gRNA injections _____	101
Table 4.1: Paired-end read counts of raw <i>D. pseudoobscura</i> SLOB3 wild type single primary spermatocyte cyst RNA sequence data _____	105
Table 4.2: Counts of mapped and unmapped reads from <i>D. pseudoobscura</i> SLOB3 wild type single spermatocyte cyst RNA sequencing _____	105
Table 4.3: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq A based on hierarchical cluster analysis clustering into two cyst groups	111
Table 4.4: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq A based on hierarchical cluster analysis clustering into three cyst groups _____	113
Table 4.5: Stage two RNA-seq analysis of primary spermatocyte cysts _____	116
Table 4.6: Counts of mapped and unmapped reads from <i>D. pseudoobscura</i> SLOB3 wild type single spermatocyte cyst RNA sequencing _____	116
Table 4.7: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq based on hierarchical cluster analysis clustering into two cyst groups _	126
Table 4.8: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq based on hierarchical cluster analysis clustering into three cyst groups	127
Table 4.9: RNA-seq analysis of <i>D. pseudoobscura</i> SLOB3 wild type spermatid cysts _____	129
Table 4.10: Counts of mapped and unmapped reads from <i>D. pseudoobscura</i> SLOB3 wild type single spermatid cyst RNA sequencing _____	129
Table 4.11: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of spermatid cyst RNA-seq based on long and short cyst groups _____	133
Table 4.12: Summary of gene expression analysis in <i>D. pseudoobscura</i> testes _____	156
Table 5.1: <i>D. pseudoobscura</i> F0 adult survivors of piggyBac-3xP3-AmCyan-Kmg-GFP injection and offspring from F0 crosses with <i>D. pseudoobscura</i> SLOB3 wild-type _____	164
Table 5.2: F2 offspring from <i>D. pseudoobscura</i> piggyBac-3xP3-AmCyan-Kmg-GFP transgenic F1 adults crossed with SLOB3 wild-type _____	165

List of Appendices

Appendix 1: Kumgang DNA and protein sequences for *D. melanogaster* and *D. pseudoobscura*.

Appendix 2: Sequences of Cas9 constructs. Sequencing results of Cas9 constructs and inverse PCR.

Appendix 3: RNA-seq spermatocytes 1 – Raw and normalised RNA-seq count data.

Appendix 4: RNA-seq spermatocytes 1 – Analysed data, fold-changes, DGE-analysis p-values.

Appendix 5: RNA-seq spermatocytes 2 – Raw and normalised RNA-seq count data.

Appendix 6: RNA-seq spermatocytes 2 – Analysed data, fold-changes, DGE-analysis p-values.

Appendix 7: RNA-seq spermatids – Raw and normalised RNA-seq count data.

Appendix 8: RNA-seq spermatids – Analysed data, fold-changes, DGE-analysis p-values.

Appendix 9: List of genes detected in spermatid dataset only.

Appendix 10: Sequences of Kmg-GFP constructs. Sequencing results of Kmg construct assembly.

*Appendix 11: Materials and methods for spatial gene expression analysis (10x Genomics) of *D. pseudoobscura* and *D. melanogaster* testes and cysts.*

*Appendix 12: Materials and methods for construct assembly of *achi/vis* and *caf1A* fluorescent tag insertions.*

1 Introduction

1.1 *Drosophila pseudoobscura*, a Sperm Heteromorphic Species

1.1.1 What is sperm heteromorphism?

Sperm heteromorphism is defined as the production of two or more distinct sperm morphs by a single male (Joly and Lachaise 1994). In the *obscura* species group of *Drosophila* flies, mature sperm of three distinct size classes are produced: a form of sperm heteromorphism known as polymegaly. The *Ohomopterus* and *Scarites* ground beetles have dimorphic formation of sperm bundles. In the Lepidoptera, sperm morphs vary in size and cell contents, with 'apyrene' morphs lacking a nucleus.

While variations in the morphology of sperm from single males is common in many species, including humans, the production of non-fertile morphs in sperm heteromorphism is distinct from the production of deformed or non-functional sperm by aberrations in development (Harcourt 1991; Swallow and Wilkinson 2002; Pitnick *et al.* 2009; Mossman *et al.* 2013; van der Horst and Maree 2014). Sperm heteromorphism can also be defined in distinction from sperm polymorphism, where sperm polymorphism refers to variation in sperm morphology between sympatric males (Joly and Lachaise 1994). In sperm-heteromorphic species, non-fertile sperm morphs are adaptive, developmentally programmed, and show distinct morphology, with variation between morphs greater than within-morph variation (Harcourt 1989, 1991).

One well described example of sperm heteromorphism is that of *Drosophila pseudoobscura*, of the *obscura* species group (Frolova and Astaurov 1929). *D. pseudoobscura*, along with the other *obscura* group species, exhibits a form of sperm heteromorphism termed 'polymegaly' (Beatty and Sidhu 1969), in which two categories consisting of three morphs are produced, one eusperm morph and two parasperm morphs (Beatty and Sidhu 1969; Alpern *et al.* 2019). Eusperm are capable of fertilisation, whereas parasperm are not capable of fertilisation (fertilisation competence) (Snook *et al.* 1994; Snook and Karr 1998; Snook and Markow 2002).

1.1.2 Heteromorphic sperm are observed in a diverse range of taxa

There have been multiple instances of evolution of sperm heteromorphism in a diverse range of taxa in the Insecta alone, including Lepidoptera, Hemiptera, Hymenoptera and Diptera (reviewed in Swallow and Wilkinson 2002; Chapman 2008). There is also evidence of sperm heteromorphism in fish (Hayakawa 2007). There has been some debate as to whether a form of sperm heteromorphism, resulting from aberrations in meiosis, exists in mammals, however the current consensus is that this is non-adaptive variation (Baker and Bellis 1988; Baker and Bellis 1989; Harcourt 1989, 1991; Swallow

and Wilkinson 2002; Till-Bottraud *et al.* 2005). It has also been argued that sperm heteromorphism in animals and pollen heteromorphism in plants are an example of convergent evolution across kingdoms (Till-Bottraud *et al.* 2005).

1.1.2.1 Lepidoptera produce eupyrene and apyrene sperm

Sperm heteromorphism in the Lepidoptera (butterflies and moths) has been comparatively well-studied and characterised in terms of sperm morphology and development. Sperm dimorphism in Lepidoptera takes the form of two sperm morphs, eupyrene and apyrene, only the former of which is fertilisation competent (Friedländer and Gitay 1972). Apyrene sperm are not capable of fertilisation, and do not contain a nucleus (Friedländer and Gitay 1972; Meves 1903, cited in Friedländer *et al.* 2005). While many studies have investigated the function of dimorphic sperm in the Lepidoptera, the strongest evidence points to the use of apyrene sperm as a 'cheap filler' to delay female re-mating (Silberglied *et al.* 1984; Cook and Wedell 1999; Swallow and Wilkinson 2002). Dichotomous sperm production in Lepidoptera is sequential, rather than simultaneous. Eupyrene sperm are produced only in the larval stages. There is then a switchover to apyrene spermatogenesis in the pupa, induced by a hormone active in the haemolymph; 'apyrene-spermatogenesis-inducing-factor' (ASIF) (Friedländer 1997; Friedländer *et al.* 2005).

1.1.2.2 Heteromorphic sperm in Hemiptera vary in chromosome number

Some species belonging to the family Pentatomidae, of the order Hemiptera (true bugs), exhibit a form of sperm heteromorphism in which both the size and the chromosome number of mature sperm varies, termed 'heteroploidy'. These species produce sperm containing chromosome numbers as low as one (always the sex chromosome), to more than 100 (Schrader 1960; Swallow and Wilkinson 2002). Development of these sperm with varying numbers of chromosomes results from irregular meiosis, and takes place in one lobe of the testis, termed the 'harlequin lobe' (Schrader 1960; de Souza and Itoyama 2010).

1.1.2.3 Hymenoptera produce sperm heteromorphic for a coiled head structure

Of the Hymenoptera (bees and wasps) the wasp *Dahlbominus fuscipennis* has been described as producing at least two distinct sperm morphs, approximately 190µm in length with a distinctive corkscrew-like head structure. The head structure appears to be dimorphic, showing either dextral or sinistral coiling (Lee and Wilkes 1965; Quicke *et al.* 1992). The function of these sperm morphs is unclear. Swallow and Wilkinson (2002) describe Lee and Wilkes' hypothesis that the sinistrally-coiled morph forms a plug in the micropyle, but does not enter the egg, resulting in the development of haploid males. However, they found no further research supporting this hypothesis, nor any other function for dimorphic sperm in this species.

1.1.2.4 Coleoptera sperm bundle dimorphism

The ground beetle genus *Ohomopterus* has sperm monomorphic in length, but dimorphic in the formation of sperm bundles, or 'spermatodesms' (Takami and Sota 2007). Sperm bundles consist of multiple mature spermatozoa, joined to a cap at the head, but with tails moving freely. Takami and Sota (2007) found that increased sexual selection via sperm competition promotes greater sperm bundle size, suggesting that larger sperm bundles may be able to migrate to the sperm storage organs more quickly, however the role of dimorphic sperm bundles in *Ohomopterus* was unclear.

Another ground beetle, *Scarites terricola*, shows similar dimorphism, producing a morph which forms sperm bundles, and a morph which does not form bundles and shows marked morphological differences from bundle spermatozoa (Sasakawa 2009). It is unclear whether both morphs are capable of fertilisation, and much like the *Ohomopterus* ground beetles, the functions of *S. terricola* dimorphic sperm bundles are unknown.

1.1.2.5 Do mammals exhibit sperm heteromorphism?

It has also been suggested that the presence of abnormal or deformed sperm in mammals is also similarly adaptive, potentially playing a role in the formation of mating plugs, or aiding the passage of fertilising sperm (Baker and Bellis 1988; Baker and Bellis 1989), however the supporting evidence is highly disputed (Harcourt 1989, 1991).

1.1.2.6 Convergent evolution of sperm heteromorphism in Diptera

Sperm heteromorphism has evolved multiple times in the Diptera (flies), occurring in *Drosophila*, *Anopheles* and Diopsidae (Figure 1.1) (Beatty and Sidhu 1969; Joly and Lachaise 1994; Presgraves *et al.* 1999; Klowden and Chambers 2004). Sperm heteromorphism in Diopsidae is the ancestral condition, although one genus, *Diasemopsis*, has since lost the trait and now exhibits sperm monomorphism, with a single class of extremely long spermatozoa (Presgraves *et al.* 1999). The other genera (*Sphyracephala*, *Diopsis*, *Teleopsis* and *Cyrtodiopsis*) show sperm polymegaly, with two sperm classes distinct in length.

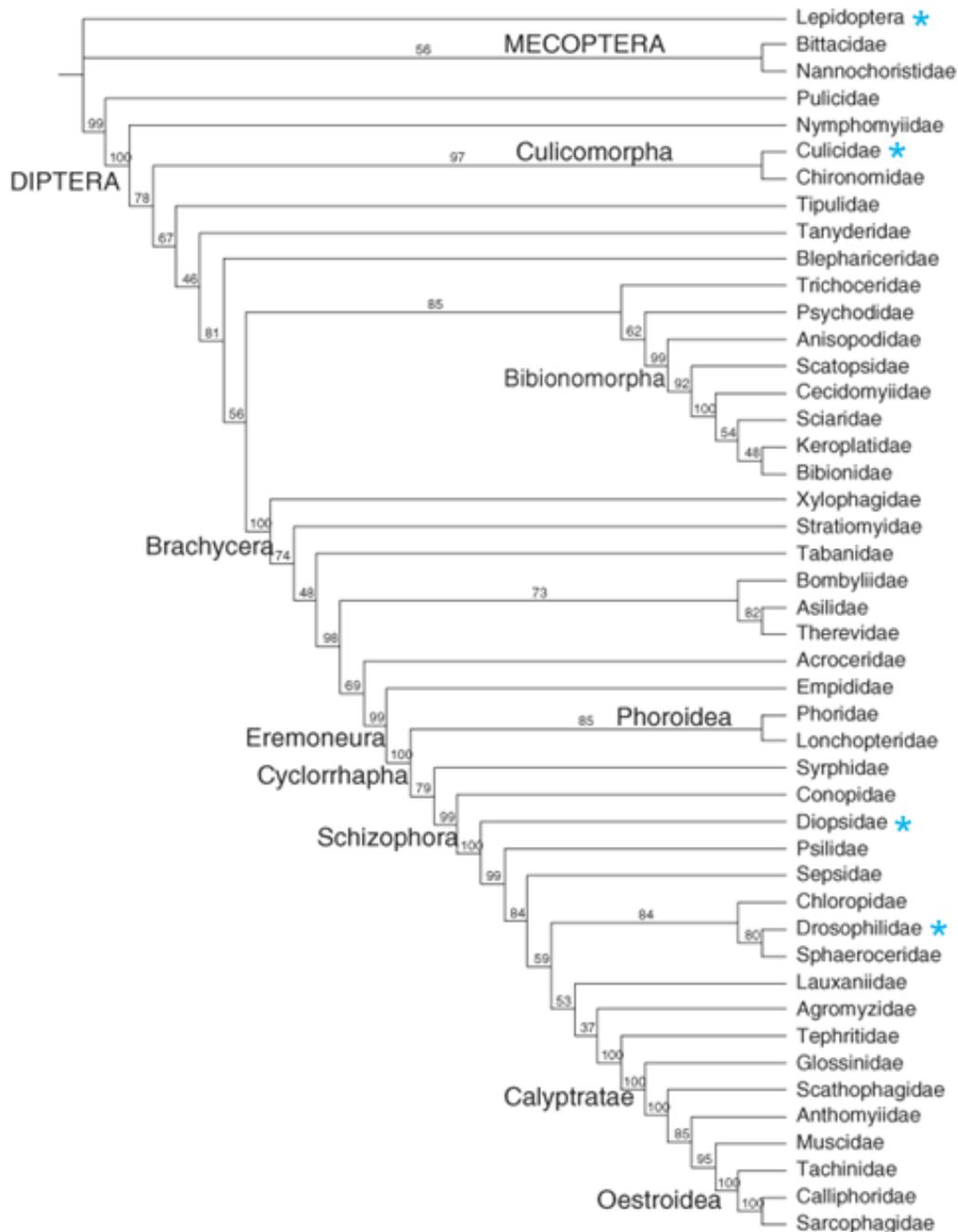


Figure 1.1: Phylogeny of Diptera based on morphological evidence, reprinted from Lambkin et al. (2013) (Wiley). Sperm heteromorphic clades are indicated by blue asterisks.

1.1.3 Sperm heteromorphism in *Drosophila* spp.

In *Drosophila*, heteromorphic sperm show two or more distinctive size classes, or ‘polymegaly’ (Beatty and Sidhu 1969). Thus far all species investigated of the *obscura* species group show polymegaly (Beatty and Sidhu 1969; Joly et al. 1989; Joly and Lachaise 1994). It is assumed to be a monophyletic trait of this species group, originating in a pre-*obscura* ancestor (Joly and Lachaise 1994). *Drosophila* species found to exhibit sperm heteromorphism, along with the number of sperm morphs present,

are summarised in Table 1.1. Sperm heteromorphism has also been observed in some populations of *D. teissieri* (Joly *et al.* 1991), a species of the *melanogaster* species group, having presumably arisen independently.

Within the *obscura* species group, there is considerable variation in the lengths of long morphs, and in the proportion of short sperm produced (Joly *et al.* 1989). The proportion of short sperm produced has also been shown to vary between laboratory and wild-caught populations, with wild populations producing and transferring more short sperm (Snook and Markow 2002). Lengths of both short and long sperm have been shown to vary between strains (Snook 1997).

Species	Number of Confirmed Sperm Morphs	Reference(s)
<i>D. subobscura</i> *	2	Joly <i>et al.</i> (1989); Joly and Lachaise (1994); Bircher and Hauschteck-Jungen (1997)
<i>D. affinis</i>		Sanger and Miller (1973); Joly <i>et al.</i> (1989); Joly and Lachaise (1994); Bircher and Hauschteck-Jungen (1997)
<i>D. helvetica</i>		Joly <i>et al.</i> (1989); Joly and Lachaise (1994)
<i>D. kitumensis</i>		
<i>D. microlabis</i>		
<i>D. bifasciata</i>		
<i>D. guanche</i>		
<i>D. madeiriensis</i>		
<i>D. obscura</i>		
<i>D. tristis</i>		Sanger and Miller (1973); Snook (1997)
<i>D. algonquin</i>		
<i>D. athabasca</i>		
<i>D. narragansett</i>		Sanger and Miller (1973)
<i>D. tolteca</i>	2 or 3	Beatty and Sidhu (1969); Joly <i>et al.</i> (1989); Joly and Lachaise (1994)
<i>D. persimilis</i>		
<i>D. ambigua</i>		
<i>D. azteca</i>	3	Beatty and Sidhu (1969)
<i>D. pseudoobscura</i>		Bircher and Hauschteck-Jungen (1997)
		Beatty and Sidhu (1969); Alpern <i>et al.</i> (2019)

Table 1.1: Summary of sperm heteromorphism in *Drosophila* species. *Beatty and Sidhu (1969) report up to four classes in *D. subobscura*; short, medium, long and, in some strains investigated, a possible 'giant-type', however subsequent analysis has confirmed the presence of short and long morphs only.

1.1.3.1 Characterisation of sperm heteromorphism in *Drosophila pseudoobscura*

D. pseudoobscura produces three sperm morphs; one eusperm and two parasperm morphs (Beatty and Sidhu 1969; Alpern *et al.* 2019). The presence of three sperm morphs was reported as early as 1969 (Beatty and Sidhu 1969), although evidence of the 'medium' class was disputed for many years, as the short and medium classes were often difficult to distinguish (Joly *et al.* 1989; Joly and Lachaise 1994; Snook *et al.* 1994; Snook 1997). Much of the literature only refers to two sperm morphs, and the two forms of parasperm are considered together in studies of parasperm function and evolution (Snook *et al.* 1994; Snook and Markow 1996, 2002; Holman and Snook 2008; Moore *et al.* 2013). The presence of the discrete 'medium' class and its function were re-confirmed in 2019 (Alpern *et al.* 2019).

Eusperm, which are fertilisation competent, are the longest morph, approximately 300µm in length (Joly *et al.* 1989; Alpern *et al.* 2019). Parasperm are incapable of fertilisation and are shorter than the fertilising eusperm (Snook and Karr 1998). Parasperm 1 and 2 are approximately 55µm and 100µm in length, respectively (Alpern *et al.* 2019). The parasperm 2 tail is spiral in form, and along with eusperm, has an associated spiral structure which dissociates from the tail when treated with acetic acid. The nature of this structure remains unknown; it does not contain DNA (Alpern *et al.* 2019), but it is unclear what the chemical composition of the structure could be beyond this. The nucleus length is positively correlated with flagellum length in each morph, and morphs can be distinguished based on nucleus length (Bircher and Hauschteck-Jungen 1997). Sperm morphs do not differ in their DNA content (Pasini *et al.* 1996). There are some suggestions that parasperm heads are wider than eusperm heads, to accommodate the equivalent DNA content, and that this wider head results in parasperm being unable to enter the egg micropyle, hence parasperm are not fertilisation competent (Snook and Karr 1998; Holman *et al.* 2008).

The percentage of the three size classes present in the seminal vesicle and the female reproductive tract of *D. pseudoobscura* was investigated by Beatty and Sidhu (1969), and was estimated to vary between 18-35% 'short', 22-38% 'medium' and 39-60% 'long', however it was also suggested in this study that these percentages may be strain-dependent. A subsequent study found the combined short and medium classes made up 45% of sperm extracted from the seminal vesicle (Joly *et al.* 1989) and Alpern *et al.* (2019) found that of sperm transferred to the female reproductive tract, 21% was

parasperm 1, 26% was parasperm 2, and 52% was eusperm. The proportions of sperm classes transferred are approximately equal to those produced (Snook *et al.* 1994).

Approximately 95% of total sperm from a given copulation is discarded by the female within thirty hours of mating, with only 5% stored in the spermathecae, the sperm storage organs (Bressac and Hauschteck-Jungen 1996). Females of sperm heteromorphic *Drosophila* preferentially store eusperm (Beatty and Sidhu 1969; Bircher *et al.* 1995; Bressac and Hauschteck-Jungen 1996). After 6 hours post mating, eusperm make up approximately 74% of sperm in storage, and after 48 hours, this increases to ~97% (Snook *et al.* 1994; Snook and Markow 1996). Alpern *et al.* (2019) observed all three sperm morphs in the sperm storage organs five days post-copulation, although the relative persistence of each morph was not measured in this study, and therefore it is unknown how long parasperm 2 remains in storage.

1.2 Functions of Heteromorphic Sperm

Theories for the function of heteromorphic sperm production have been the subject of discussion and experimentation over the decades since the phenomenon was first described. Hypotheses for heteromorphic sperm function vary across taxa and the form of sperm heteromorphism.

Schrader (1960) suggested a role for the large harlequin sperm of the Pentatomidae in providing nucleoproteins during fertilisation involving polyspermy, which could then be utilised by the developing embryo. The dimorphic coiled sperm of the wasp *Dahlbominus fuscipennis* was suggested to be involved in the determination of sex ratio, where males develop from unfertilised eggs (arrhenotokous parthenogenesis) (Lee and Wilkes 1965). In the Lepidoptera, apyrene sperm function as a 'cheap filler' to delay female re-mating (Friedländer and Gitay 1972; Silberglied *et al.* 1984; Friedländer 1997; Friedländer *et al.* 2005).

In the Diopsid flies, the length of long sperm is correlated with the length of the ventral receptacle and spermatheca (Presgraves *et al.* 1999). Presgraves *et al.* (1999) did not explore the function of the short morph in the Diopsidae, but suggested possible roles in delaying remating or preventing fertilisation from future mating.

Hypotheses for the functions of heteromorphic sperm in other orders of the Insecta are not necessarily also supported by evidence in *D. pseudoobscura*. The functions of the multiple sperm size classes in *D. pseudoobscura* and other *obscura* group species have been debated for many years. In the following section I will discuss the research exploring the potential function or functions of heteromorphic sperm, specifically non-fertilising parasperm, in *D. pseudoobscura*.

1.2.1 Hypotheses for the functions of heteromorphic sperm in *D. pseudoobscura*

Swallow and Wilkinson (2002) defined four categories of hypotheses for the functions of non-fertile sperm in Insecta:

1. Non-adaptive – parasperm have no function;
2. Provisioning – provision of nutrients to the female, ovum or eusperm;
3. Facilitation – including transportation of sperm morphs either from the testes or within the female, and capacitation (acquisition of motility);
4. Competition – including the elimination of stored sperm, blocking sperm entry, and delaying female re-mating.

Experimental studies examining the roles of eusperm and parasperm in fertilisation, provisioning, facilitation and sperm competition have demonstrated that *D. pseudoobscura* parasperm have roles in aspects of facilitation, and potentially in sperm competition.

1.2.1.1 Parasperm are not fertilisation competent

It was initially thought that both short and long morphs were fertilisation competent, and were used in different fertilisation scenarios; short sperm were used in immediate fertilisation, whereas long sperm, which persisted longer in storage, were used in delayed fertilisation (Bressac *et al.* 1991; Joly *et al.* 1991; Joly and Lachaise 1994). Joly *et al.* (1991) further theorised that sperm heteromorphism is an adaptation against last male sperm predominance.

It was later found that parasperm were not fertilisation competent (Snook *et al.* 1994; Snook and Karr 1998; Snook and Markow 2002). Notably, Snook *et al.* (1994) demonstrated that eggs oviposited up to 72 hours post-mating only contained eusperm, thus parasperm could not be used for immediate fertilisation. Therefore, sperm heteromorphism could not function in differing fertilisation scenarios as previously hypothesised by Bressac *et al.* (1991), Joly *et al.* (1991) and Joly and Lachaise (1994).

It is unclear why parasperm are not fertilisation competent, but may involve biochemical or physical sperm-egg interactions (Snook *et al.* 1994; Snook and Karr 1998). Both morphs contain acrosomes, so could theoretically enter the egg and initiate breakdown of the sperm plasma membrane (Pasini *et al.* 1996; Southern *et al.* 2018). It has been suggested that eusperm may contain surface receptors necessary for entry into the micropyle which parasperm lack. Alternatively, since the presence of the sperm tail within the egg is required to initiate embryogenesis, it is possible that only long sperm contain enough of the required components for embryogenesis (Karr 1991; Snook *et al.* 1994; Southern *et al.* 2018). A simpler explanation may be that the parasperm head is too wide to enter the egg through the micropyle (Snook *et al.* 1994; Holman *et al.* 2008).

1.2.1.2 Evaluation of the Non-Adaptive Hypothesis: parasperm are adaptive and evolve independently of eusperm

Since parasperm have been found to not be capable of fertilising eggs, it was unclear what, if any, function is performed by parasperm (Snook *et al.* 1994; Snook and Karr 1998). To establish whether parasperm did have a function, and were not simply a non-adaptive trait, eusperm and parasperm lengths were compared across the *obscura* group species, in order to measure phylogenetic correlation with trait variation.

Non-adaptive trait variation is expected to exhibit a particular pattern; closely related taxa should show greater similarity in a given trait, and as phylogenetic distance increases, similarity in trait values should decrease (high phylogenetic correlation). Where the trait has been subject to independent evolutionary processes, low phylogenetic correlation would be expected; short phylogenetic relationships show low similarity in a given trait and vice versa (Snook 1997). In the *obscura* species group, short sperm length does not show significant phylogenetic correlation, whereas 22% of variation in long sperm length is associated with phylogeny, which indicates that most to all trait variation in heteromorphic sperm length is the result of independent evolutionary processes and supports an adaptive function for sperm heteromorphism (Snook 1997). Furthermore, eusperm and parasperm lengths have evolved independently of one another, as indicated by the lack of a correlation between the lengths of the two morphs across the *obscura* species group (Snook 1997; Holman *et al.* 2008).

Flagellum lengths of both short and long sperm show much higher phenotypic variation than head lengths of both sperm types (Moore *et al.* 2013). Within each sperm morph, head and flagellum lengths show a positive genetic and phenotypic correlation (Moore *et al.* 2013). There is no genetic correlation between sperm morphs, and low phenotypic correlation, suggesting low genetic integration between sperm morphs, and likely therefore, little developmental integration (Moore *et al.* 2013). This is reflected in the separate development of eusperm and parasperm in different cysts within the testis (Beatty and Sidhu 1969).

Moore *et al.* (2013) also showed that parasperm have a greater capacity for evolution: ten times that of eusperm. It is important to note that this study assumed only one sperm morph was present, and as such some of the variation in parasperm length measured may therefore reflect the presence of two morphs, rather than a highly variable and therefore more evolvable single parasperm morph.

The evidence presented in these studies does not support the first hypothesis put forward by Swallow and Wilkinson (2002), that parasperm are non-adaptive. Parasperm morphology across the *obscura* species group did not match the characteristics of a non-adaptive trait, despite this sperm morph not

being fertilisation competent. Evidence also suggests that eusperm and parasperm evolve independently and are subject to different selection pressures.

1.2.1.3 The Provisioning Hypothesis: parasperm do not provide nutritional provisioning to the female

The provisioning hypothesis is that short sperm act as a nutrient source for the female. This hypothesis gives the prediction that when male-derived material is incorporated into female tissues, parasperm are the source, and are therefore acting to provide nutrition to the female.

Females mated to radiolabelled males were shown to incorporate labelled male-derived material prior to the 'disappearance' of short sperm from the sperm storage organs (6h post-mating), refuting the hypothesis that parasperm are a nutrient source for female tissues (Snook and Markow 1996). They suggest that the incorporated material could originate from accessory gland secretions in the ejaculate, or oral drops used in courtship feeding.

This evidence is therefore contrary to the provisioning hypothesis outlined by Swallow and Wilkinson (2002), suggesting that the function of parasperm is not as a nutrient source.

1.2.1.4 The Facilitation Hypothesis: parasperm increase survival of eusperm within the female reproductive tract

Parasperm have an adaptive function, and are subject to different selection pressures than eusperm (Snook 1997; Holman *et al.* 2008; Moore *et al.* 2013). Parasperm do not have a function in fertilisation, nor do they increase the likelihood of successful fertilisation by providing nutrients to the female (Snook *et al.* 1994; Snook and Markow 1996; Snook and Karr 1998). The facilitation hypothesis provides an alternative role for parasperm, in aiding the transportation of eusperm from the testes to the sperm storage organs (Swallow and Wilkinson 2002). The example given by Swallow and Wilkinson is that parasperm facilitate eusperm acquisition of motility. An alternative version of this hypothesis was subsequently considered by Holman *et al.* (2008); that parasperm protect eusperm from female-derived spermicide, therefore increasing eusperm survival in the female reproductive tract (Holman and Snook 2006; Holman *et al.* 2008; Holman and Snook 2008).

Holman *et al.* (2008) found a correlation between eusperm length and the proportion of parasperm produced; species with longer eusperm produced fewer eusperm and more parasperm. This indicated that parasperm may protect eusperm from spermicide since parasperm proportion was previously predicted to increase as eusperm became more vulnerable to spermicide, i.e. by increased length (Holman and Snook 2006).

Holman and Snook (2008) found lower eusperm viability where sperm was removed from the female reproductive tract 30 minutes after copulation, compared to sperm removed immediately after copulation. Furthermore, sperm extracted from the seminal vesicle had lower eusperm viability when incubated with extracted tissue from female reproductive tract compared to thoracic muscle tissue or saline. They then tested whether the presence of parasperm protects eusperm from spermicide. Males which transferred a higher proportion of parasperm had higher eusperm viability 30 minutes after copulation, and experimental manipulation of parasperm proportion showed a strong positive relationship with eusperm viability when incubated with female reproductive tract tissue extract. These results give strong evidence for the hypothesis that spermicides are present in the female reproductive tract, and that parasperm protect eusperm from female-mediated spermicide (Holman and Snook 2008). Alpern *et al.* (2019) further confirmed Holman *et al.*'s findings, showing that both parasperm morphs 1 and 2 protect eusperm from female-mediated spermicide.

The mechanisms of female-mediated spermicide, and parasperm protection from spermicides, remain unclear. The term 'female-mediated spermicide' covers a broad range of biological actions, which may include sperm death as the result of phagocytosis, exposure to adverse pH, cytotoxic compounds, enzymes or the immune system (Greeff and Parker 2000; Holman and Snook 2006). Holman and Snook (2008) found that female reproductive tract tissue remained spermicidal after homogenisation and removal of insoluble compounds, suggesting the presence of cytotoxic compounds in or around the female reproductive tract cells. Parasperm may not need to have specific adaptations or structures against spermicidal compounds; the presence of parasperm may be enough in and of itself to protect eusperm from spermicidal compounds, simply by dilution (Alpern *et al.* 2019).

The function of female-mediated spermicide is also unclear. It may protect against infection, prevent polyspermy (where multiple sperm enter the egg), aid in the facilitation of female control of sperm storage or sperm competition, or alternatively may be a non-adaptive by-product of mating-induced immune response (reviewed in Holman and Snook 2006).

Parasperm function in *D. pseudoobscura* can be considered a mechanism of facilitation, increasing the survival of eusperm, thus increasing the likelihood of fertilisation.

1.2.1.5 The Competition Hypothesis: parasperm may displace stored sperm but do not appear to delay female re-mating or provide an honest signal of male quality

Sperm competition arises where sperm from multiple males could be used for fertilisation by a single female. *D. pseudoobscura* females mate with multiple males, and store sperm in the spermathecae (Bressac and Hauschteck-Jungen 1996; Snook 1998a; Civetta 1999; Pitnick *et al.* 1999).

D. pseudoobscura have been found to respond behaviourally to increased risk of sperm competition (Snook 1998a; Dhole and Pfennig 2014). Males mated to previously mated females show a reduction in copulation duration compared to males mated to virgin females, and previously mated males also reduce copulation duration compared to virgin males (Snook 1998a). The risk of sperm competition was not found to influence the proportions of sperm morphs transferred in this study, suggesting that parasperm do not have a function in sperm competition.

An alternative method of varying the perceived risk of sperm competition resulted in the opposite result to the Snook (1998a) study. Price *et al.* (2012) exposed male *D. pseudoobscura* to rival conspecific males, finding that these males mate for longer, produce more offspring and increase the number of eusperm transferred to the female, than those kept alone or exposed to non-rival males (Price *et al.* 2012). Parasperm transfer did not significantly increase, although did show the same trend as eusperm transfer. This supported a role for eusperm, but not parasperm, in sperm competition.

There has also been debate about a potential role for parasperm in the displacement of sperm from previous matings with competing males, stored in the spermathecae. Two studies in particular show contradictory results regarding the function of parasperm in sperm competition. Snook (1998a) found that males did not modulate the proportion of parasperm transferred to old or non-virgin females, as would be expected where parasperm have a role in sperm displacement. Snook (1998a) further suggested that since short and long sperm were transferred simultaneously, short sperm were not acting as 'front runners' which arrive at the sperm storage organs first to displace stored sperm from previous males. In their 2019 study, Alpern *et al.* revisited parasperm function in light of the confirmation of a second parasperm morph. They found that in more competitive environments males lowered the proportion of parasperm 1 and eusperm but increased the proportion of parasperm 2, suggesting parasperm 2 has role in sperm competition, such as displacement of competitor sperm from the reproductive tract and sperm storage organs (Alpern *et al.* 2019).

Another aspect of sperm competition is whether and how quickly females will re-mate, with greater numbers of mates and shorter re-mating times increasing sperm competition. Males may try to delay female re-mating, for example by 'tricking' the female that their sperm storage organs contain higher numbers of fertilising sperm than are present, in other words, parasperm act as 'cheap fillers' (Silberglied *et al.* 1984). Females less receptive to re-mating have been shown to have a higher proportion of short sperm in the ventral receptacle compared to receptive females, however there was no difference between receptive and non-receptive females in the proportion of short sperm in the spermathecae (Snook 1998a), giving conflicting support for the 'cheap filler' hypothesis. Female receptivity to re-mating was instead linked to female fecundity and the absence of an egg in the

uterus, rather than the numbers of either sperm type in storage (Snook 1998a). Holman *et al.* (2008) found that eusperm and parasperm lengths in *obscura* species are not correlated with the size of sperm storage organs, further refuting a cheap filler function for parasperm.

An interesting suggested hypothesis is that parasperm functions as an 'honest signal' of quality to the female, such that she might evaluate the quality of the ejaculate and subsequently decide whether to retain the sperm of the previous male or dispose of it in favour of the new male (Swallow and Wilkinson 2002). In this scenario, parasperm acts as an 'honest signal', demonstrating that the male is of quality proportional to the amount of energy he is able to waste in the production of non-fertilising sperm, much like the tail of a peacock. However, Crudgington *et al.* (2009) did not find any evidence to support the evolution of the anti-spermicidal sperm caste by sexual selection in *D. pseudoobscura*, indicating that parasperm are not considered a 'desirable' trait.

1.2.1.6 Summary of eusperm and parasperm function in *D. pseudoobscura*

To summarise, both eusperm and parasperm have functions in *D. pseudoobscura*. Eusperm fertilise eggs, whereas parasperm are not capable of fertilisation (Snook *et al.* 1994; Snook and Karr 1998). The roles of parasperm appear to be morph dependent, with both morphs protecting eusperm from female-mediated spermicides, and parasperm 2, the medium morph, also having a potential role in the removal of competing sperm from the spermathecae (Holman and Snook 2008; Alpern *et al.* 2019). These functions are linked to independent evolution of morph lengths, indicating distinct genetic modulation of the structures of the different sperm morphs.

The functions of heteromorphic sperm in *D. pseudoobscura* have received considerable attention over the past thirty years, and are now reasonably well understood. The formation of these sperm morphs is regulated, adaptive, and genetically controlled, however there is almost nothing in the literature regarding the specific mechanisms involved. In the next section I will discuss the current understanding of sperm development in *Drosophila*, and importantly, the lack of research into sperm development in sperm heteromorphic species.

1.3 Sperm Development in *Drosophila*

1.3.1 General patterns of sperm development and testis structure in *Drosophila* spp.

There are multiple reviews available describing the phases of sperm development in *Drosophila melanogaster* (Lindsley and Tokuyasu 1980; Fuller 1993), which I will summarise below. Broadly, germline stem cells (GSC) at the apical tip of the testis differentiate to generate a spermatogonium and are encapsulated by two somatic 'cyst' cells. Within the cyst, the spermatogonium undergoes

several rounds of division to generate a pre-meiotic primary spermatocyte cyst (Kurokawa and Hihara 1976; Hardy *et al.* 1979; Fuller 1993). The spermatocytes undergo a 25-fold increase in volume prior to entry into meiosis I. Completion of meiosis I generates a secondary spermatocyte cyst (Fuller 1993; Fuller 1998). Completion of meiosis II results in the spermatid cyst. Spermatids undergo elongation and individualisation before release from the cyst at the base of the testis (Tokuyasu *et al.* 1972b; Fabian and Brill 2012).

It is important to note that the vast majority of research in this area is of the sperm monomorphic model species *D. melanogaster*, and therefore does not take into account the additional complexities of heteromorphic sperm development. It is still useful to understand the general patterns by which sperm development in *Drosophila* occurs, as this provides a framework to understand the systems in which heteromorphic sperm also develop, which I discuss further in the following section.

1.3.1.1 *Drosophila* testis structure

The *Drosophila* testis resembles a tube, with an apical blind-end and a basal end connected to the seminal vesicle. The seminal vesicle and accessory glands are attached to the ejaculatory duct, through which sperm are transferred in mating. The testis includes a sheath of muscle and pigment cells, separated from the lumen by the basement membrane. The *D. melanogaster* testis is coiled in shape, and is approximately 2mm in length, and 0.1mm in width (Hardy *et al.* 1979). *D. pseudoobscura* testes are ellipsoidal, approximately 0.7-0.8mm in length (Baker 1935), and the sheath has a characteristic orange-red colour, compared to a creamy-yellow in *D. melanogaster* (Stern and Hadorn 1939).

Within the testes, the developmental stages follow a typical pattern of localisation. In *D. melanogaster*, the earliest stages (hub, GSCs) are located at the apical tip. Cyst development coincides with movement away from the hub, with cysts later in development further from the apical hub region (Figure 1.2; A). *D. pseudoobscura* testes show similar organisation, with later stages further from the apical hub (Figure 1.2; B). Spermatocyte cysts are located at the peripheral regions of the testis, closer to the basement membrane. Spermatid cysts elongate in the medial region of the testis, surrounded by the spermatocyte cysts (Njogu *et al.* 2010). The precise arrangement of germline cells that will generate the different sperm morphs within the testis is not known.

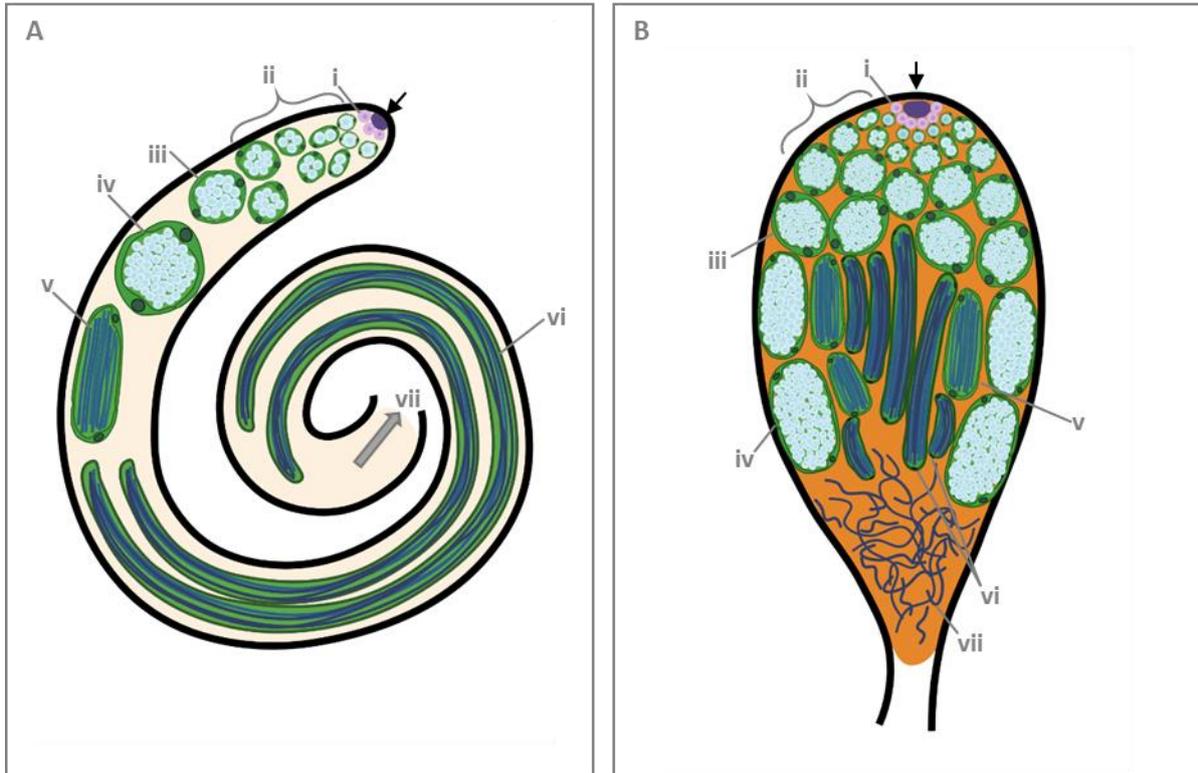


Figure 1.2: Model of testes structure in (A) *D. melanogaster* and (B) *D. pseudoobscura*. Black arrows indicate apical tip. Hub shown in purple. Germline cells shown in blue (dark blue indicates elongating/elongated cells). Cysts cells shown in green. (i) Germline stem cells, (ii) spermatogonia cysts, (iii) primary spermatocyte cysts, (iv) early spermatid cysts, (v) elongating spermatid cysts, (vi) elongated spermatid cysts, (vii) mature spermatozoa (*D. pseudoobscura* only).

1.3.1.2 Germline stem cell maintenance and differentiation

The germline stem cells proliferate from the apical tip of the testes, in a region termed the apical proliferation centre. The apical proliferation centre contains the testis hub, a cluster of approximately 10-15 somatic cells (Hardy *et al.* 1979) responsible for secretion of signalling molecules essential for the maintenance and differentiation of the stem cell niche (Tulina and Matunis 2001; Yamashita *et al.* 2003; Wang *et al.* 2006). The hub cells form a dome shape, with some cells adhered to the basement membrane, while others protrude into the testis lumen and provide a docking point for the germline stem cells (GSC) (de Cuevas and Matunis 2011; Matunis *et al.* 2012; Inaba *et al.* 2015). Surrounding the hub are the GSCs, of which there are approximately 6-9 in *D. melanogaster* (Hardy *et al.* 1979). GSC division is asymmetric, producing two daughter cells; one GSC which has maintained contact with the hub, and one gonialblast (also referred to as a primary spermatogonial cell) displaced away from the hub which will go on to form the spermatogonial cyst (Fuller 1998). The two daughter cells remain connected by a cytoplasmic bridge until the subsequent round of cell division (Hardy *et al.* 1979; Lindsley and Tokuyasu 1980).

1.3.1.3 Encapsulation of the gonialblast by cyst cells

GSCs are associated with the cyst progenitor stem cells (CySCs), which maintain contact with the hub by cytoplasmic projections. The division of two CySCs results in two daughter CySCs and two cyst cells, the latter of which encapsulate the gonialblast. The cyst cells encapsulate the developing germline cell as it goes through mitotic and meiotic division, cell growth and reshaping, until finally released as mature sperm at the base of the testis (Fuller 1993).

1.3.1.4 Spermatogonium mitotic divisions

Within the cyst, the spermatogonia undergo several rounds of mitotic divisions. Mitosis is incomplete, and the spermatogonia remain connected by intercellular bridges, the ring canals, through which extends the branched, spectrin-rich fusome structure (Tates 1971; Lindsley and Tokuyasu 1980; Fuller 1998). *D. melanogaster* spermatogonia undergo four rounds of mitosis resulting in a 16-cell cyst, whereas *D. pseudoobscura* undergo five rounds of mitosis, resulting in 32 cells (Dobzhansky 1934; Kurokawa and Hihara 1976; Scharer *et al.* 2008). The final round of mitotic division is followed shortly by a pre-meiotic S-phase to produce the 16- or 32-cell primary spermatocyte cyst (Dobzhansky 1934; Cross and Shellenbarger 1979; Fuller 1993).

1.3.1.5 Spermatocyte cell growth, transcription and meiosis

Primary spermatocyte development, consisting of extensive transcription and cell growth, takes place during an extended G2 phase (Olivieri and Olivieri 1965; Tates 1971; Fuller 1993). During this phase, the spermatocyte cells grow up to 25 times in volume (Lindsley and Tokuyasu 1980). The majority of transcription for gene products required in the later meiotic and post-meiotic phases occurs in the primary spermatocyte stage (Olivieri and Olivieri 1965; Gould-Somero and Holland 1974; Schafer *et al.* 1995; Li *et al.* 2022). Transcripts required at later stages are stored in cytoplasmic ribonucleoprotein bodies (Schafer *et al.* 1995).

Transcription is largely halted as the primary spermatocytes mature (Olivieri and Olivieri 1965; Gould-Somero and Holland 1974), the chromatin condenses and the primary spermatocytes enter into meiosis I. Completion of meiosis I results in a 32-cell secondary spermatocyte cyst (Cross and Shellenbarger 1979) (*D. pseudoobscura* = 64-cell). Interphase between meiosis I and II is short, as indicated by the relative rarity of secondary spermatocytes (Church and Lin 1985). Completion of meiosis II results in a 64-cell spermatid cyst (*D. pseudoobscura* = 128-cell) (Cross and Shellenbarger 1979; Fuller 1993; Scharer *et al.* 2008). Within the cyst, the spermatocytes remain interconnected throughout meiosis and into the spermatid phase. A new nuclear envelope consisting of two layers, the outer of which is derived from the endoplasmic reticulum, forms around the haploid nuclei (Tates 1971).

1.3.1.6 Spermatid elongation and individualisation

During the early spermatid stage, the centriole inserts into the base of the nucleus to form the basal body. During this time, the mitochondria migrate and aggregate near the basal body. The mitochondria fuse to produce the Nebenkern – a many layered mitochondrial derivative structure which has the appearance of an onion, hence this phase is referred to as the ‘onion stage’ (Tates 1971; Fuller 1993; Fabian and Brill 2012). During the onion stage, the Golgi bodies fuse to form the acroblast at the apical end of the spermatid nucleus. Assembly of the flagellar axoneme begins from the basal body, embedded in the nucleus.

Spermatid elongation involves a dramatic change in shape to reach full elongation; up to 1.8mm in *D. melanogaster* and 0.3mm in *D. pseudoobscura* eusperm (Lindsley and Tokuyasu 1980; Joly *et al.* 1989). Elongation involves the assembly of the flagellar axoneme, alongside which the unfurled Nebenkern elongates into two mitochondrial derivatives (Cross and Shellenbarger 1979). The axoneme and mitochondrial derivatives will ultimately elongate to the full length of the tail.

The flagellar axoneme is composed of a 9 + 2 formation of microtubules – nine outer microtubule doublets surround two inner microtubule singlets (Tokuyasu 1974a). The flagellar axoneme is surrounded by the axoneme sheath, derived from the endoplasmic reticulum (Lindsley and Tokuyasu 1980). Spermatid elongation is synchronous within the cyst, and polarisation is tightly controlled with respect to the testis axis; the elongating tails are positioned towards the apical end, and heads towards the base of the testis (Tokuyasu 1975a; Fabian and Brill 2012). The cyst cells are also polarised, with one over the heads, and the other over the tails (Fabian and Brill 2012). Spermatid elongation is also associated with nuclear elongation, resulting in a needle shaped nucleus (Tates 1971; Tokuyasu 1974b; Lindsley and Tokuyasu 1980). The chromatin is condensed from a histone-based configuration, via transition proteins, to a protamine-based configuration (Jayaramaiah Raja and Renkawitz-Pohl 2005; Rathke *et al.* 2007).

In each spermatid, an actin-rich structure, the investment cone, assembles over the nucleus then moves along the spermatid tail. Migration of this structure along the tails is coordinated between spermatids in the cyst and the cyst cells. The entire complex of the actin cones is known as the individualisation complex. As the individualisation complex moves from the nuclear end down the length of the spermatid bundle, excess cytoplasm and organelles are removed, and the intercellular membrane gaps are resolved such that each spermatid is encased in its own membrane. The individualisation complex is contained within a cystic bulge of waste material. The waste components in the cystic bulge pass into a ‘waste bag’ at the tail end (Tokuyasu *et al.* 1972a; Cross and Shellenbarger 1979; Fuller 1993; Fabian and Brill 2012).

Individualised spermatids then undergo a process of coiling. The somatic cyst cell encapsulating the heads of the spermatid bundle becomes embedded in the terminal epithelium at the base of the testis. The process of coiling draws the tails of the spermatid bundles into the base of the testes, and the tail cyst cell collapses (Tokuyasu *et al.* 1972b; Fuller 1993; Fabian and Brill 2012). The mature spermatozoa are released into the testis lumen and pass into the seminal vesicle, where they are stored prior to mating.

The ejaculate is composed of mature spermatozoa and seminal fluid, containing a mixture of lipids, enzymes, other proteins and protein aggregates, and carbohydrates (Gromko *et al.* 1984; Pitnick *et al.* 2020; Wigby *et al.* 2020). Seminal fluid components are produced by the accessory glands and ejaculatory duct (Gromko *et al.* 1984).

1.3.2 Genetic control of sperm development

Each stage of the process of sperm development requires significant control and synchronisation to ensure the correct timing and initiation of the many processes involved. Much of this control relies on inter- and intra-cellular signalling, initiation and regulation of transcription, and subsequent translational control of those transcription products.

Transcriptional control of the GSC maintenance and differentiation, spermatogenesis and spermiogenesis processes have been reviewed multiple times (Lin *et al.* 2000; White-Cooper 2010; White-Cooper and Davidson 2011; Lim *et al.* 2012), and are summarised below.

1.3.2.1 Germline stem cell maintenance is dependent on hub signalling

The two major signalling pathways identified as acting within the testis niche are the Janus kinase-signal transducer and activator of transcription (JAK-STAT) and Bone morphogenetic protein (BMP) signalling pathways (Figure 1.3) (Kiger *et al.* 2001; Tulina and Matunis 2001; Matunis *et al.* 2012). Regulation of GSC maintenance and differentiation is also dependent on insulin signalling, nutrition and epigenetic factors (Ueishi *et al.* 2009; de Cuevas and Matunis 2011; Matunis *et al.* 2012; Vidaurre and Chen 2021).

The testis hub is a source of the Unpaired (Upd) ligand, which binds to the STAT receptor to activate the JAK-STAT regulated transcription in adjacent stem cells (Tulina and Matunis 2001; Hombria and Brown 2002; Arbouzova and Zeidler 2006). JAK-STAT signalling promotes maintenance in the stem cells adjacent to the hub (Kiger *et al.* 2001; Tulina and Matunis 2001). As GSCs divide, the differentiating cells are displaced away from the hub, and so receive less Upd resulting in reduced activation of the JAK-STAT pathway and thus allowing differentiation in these cells (Tulina and Matunis 2001). JAK-STAT signalling is also required for cyst stem cell self-renewal (Leatherman and Dinardo

2008; Matunis *et al.* 2012; Sinden *et al.* 2012). JAK-STAT signalling contributes to the enrichment of E-cadherin, one of several components of the adherens junctions required for GSC adhesion to the hub (Lim *et al.* 2015; Kahney *et al.* 2019). Adhesion of the GSCs to the hub contributes to germline cell polarity and subsequent differentiation of the daughter cells displaced away from the hub (Matunis *et al.* 2012; Kahney *et al.* 2019).

The hub and cyst cells secrete the BMP ligands Dpp and Gbb, which bind to the receptor Mothers against dpp (Mad), and ultimately activating BMP pathway target genes in GSCs (de Cuevas and Matunis 2011). BMP pathway activation is required for the repression of *bag of marbles (bam)*, thereby promoting GSC maintenance over differentiation (Chen and McKearin 2003; Kawase *et al.* 2004; Leatherman and Dinardo 2010).

In addition to the typical system of GSC differentiation, there is a homeostatic mechanism to ensure replacement of GSCs via symmetric divisions, if a single GSC is lost (Tulina and Matunis 2001; Matunis *et al.* 2012). In extreme circumstances, GSCs can be regenerated by a process of dedifferentiation of the spermatogonia (Brawley and Matunis 2004). Breakdown of the spermatogonia cyst relies on cooperation between the cyst and the somatic stem cells and is also regulated by the JAK-STAT pathway (Brawley and Matunis 2004; Herrera and Bach 2018).

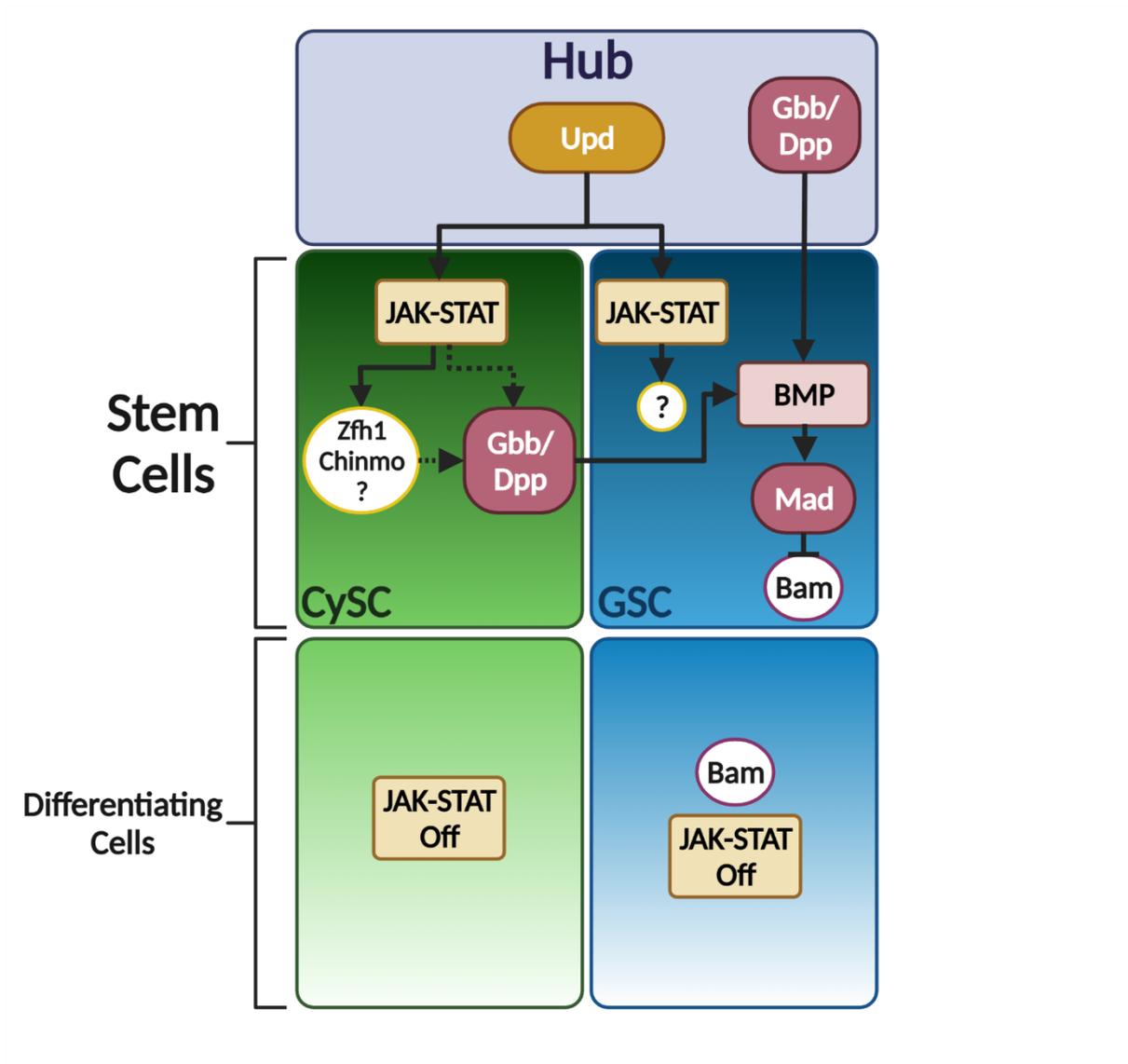


Figure 1.3: Overview of JAK-STAT and BMP signalling pathways in the *Drosophila* testis hub, based on de Cuevas and Matunis (2011). Purple: hub. Green: Cyst cell lineage. Blue: Germline cell lineage. Yellow: JAK-STAT pathway. Red: BMP pathway. The hub secretes Unpaired (Upd) which binds to the STAT receptors in GSCs and CySCs activating JAK-STAT regulated transcription. Glass bottomed boat (Gbb) and Decapentaplegic (Dpp) are produced by both the hub and CySC and regulated BMP signalling in the GSCs, which represses transcription of bag of marbles (bam). GSC division produces two daughter cells, one of which remains in contact with the hub, and the second is displaced away from the hub. The displaced cell receives less Upd resulting in lower activation of the JAK-STAT pathway. bam transcription increases, promoting differentiation.

1.3.2.2 Regulation of mitosis and the transition to meiosis in spermatogonia

The spermatogonia stage is characterised by a series of synchronous mitotic divisions prior to entry into meiosis (Fuller 1993; Fuller 1998). RNA synthesis occurs in the spermatogonia but to a lesser extent than in primary spermatocytes (Olivieri and Olivieri 1965). Characterisation of gene expression

and signalling pathways present in the spermatogonia has revealed roles in the regulation of the mitotic divisions and entry into meiosis (Figure 1.4).

Mitotic division in the spermatogonia is regulated by Bam and Benign gonial cell neoplasm (Bgcn); mutation of either *bam* or *bgn* results in accumulation of gonial cells which fail to undergo meiosis (Fuller 1998). Bam protein accumulates in the spermatogonia, eventually reaching a threshold at which point the gonial cells differentiate into spermatocytes (Gonczy *et al.* 1997; Insko *et al.* 2009; Lim *et al.* 2012). Translation of *bam* is repressed in spermatogonia by HOW RNA-binding protein (HOW) (Monk *et al.* 2010) and micro RNA-7 (miR-7), the latter of which is also repressed by the RNA-binding protein Maelstrom (Mael), thus allowing Bam accumulation in spermatogonia and eventual spermatogonia-spermatocyte transition (Pek *et al.* 2009). Bam and Bgcn form a protein complex which antagonises GSC self-renewal factors and promotes differentiation gene expression (Lim *et al.* 2012).

The transition to meiosis is dependent on the Epidermal growth factor (Egf) pathway (Lim *et al.* 2012). The Egf receptor (Egfr) ligand Spitz is processed by the transmembrane protease Stet, and activates Egfr signalling in cyst cells. Egfr signalling in cyst cells then signals back to the germline via an unknown mechanism, restricting spermatogonial division and promoting the mitosis to meiosis transition in germline cells (Kiger *et al.* 2000).

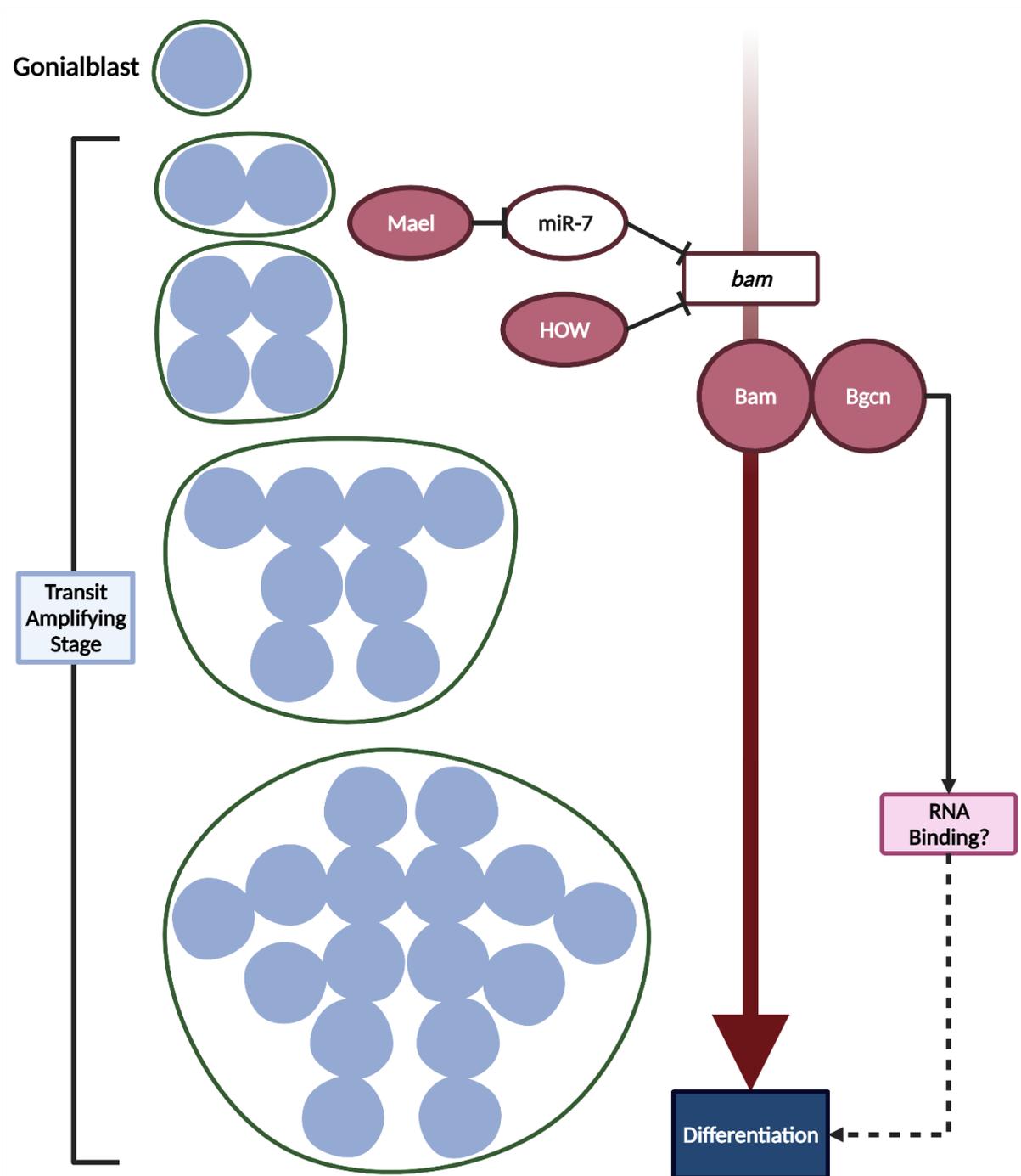


Figure 1.4: Overview of spermatogonia mitotic divisions of the transit amplifying stage prior to differentiation of the primary spermatocytes. Spermatogonia (blue) undergo four rounds of incomplete mitosis (*D. melanogaster*, five rounds in *D. pseudoobscura*) within the cyst (green). Mitotic division is regulated by the Bag of marbles (Bam) and Benign gonial cell neoplasm (Bgcn) (red). micro RNA 7 (miR-7) and HOW RNA-binding protein regulate bam translation. Maelstrom (Mael) represses miR-7 allowing Bam accumulation in spermatogonia (red arrow). Bam accumulation reaches a threshold triggering differentiation, possibly by promoting differentiation transcription by formation of a protein complex with Bgcn.

1.3.2.3 Transcription control in spermatocytes

The spermatocyte stage is initiated by completion of the S-phase and entry into the extended G2 phase (Olivieri and Olivieri 1965; Tates 1971; Fuller 1993). Spermatogenesis is characterised by extensive growth and transcription, essential for completion of meiosis and eventual spermatid differentiation processes (White-Cooper 2010; White-Cooper and Davidson 2011). More than 50% of annotated *D. melanogaster* genes are expressed in the testis, and about 25% of testis expressed genes are either testis-specific or testis-enriched (Andrews *et al.* 2000; Parisi *et al.* 2004; Chintapalli *et al.* 2007; White-Cooper 2010; Vedelek *et al.* 2018; Shi *et al.* 2020).

1.3.2.3.1 Regulation of transcription in spermatocytes by the testis meiotic arrest complex (tMAC) and testis specific paralogues of the TBP-associated factors

Transcription in spermatocytes is dependent on the meiotic arrest genes – so-called because mutants do not undergo meiosis, with cysts arrested at the G2/M transition phase (Lin *et al.* 1996; White-Cooper *et al.* 1998). Lin *et al.* (1996) first identified four of the meiotic arrest genes: *always early (aly)*, *meiosis I arrest (mia)*, *cannonball (can)* and *spermatocyte arrest (sa)*. They proposed that *can*, *mia* and *sa* had some function in the accumulation of transcripts required for both G2/M transition and spermatid differentiation, and that *aly* was upstream of the other three.

The meiotic arrest genes identified by Lin *et al.* (1996) were later shown to regulate entry into meiosis I by direct or indirect control of the RNA-binding protein Boule and the Cdc25-type phosphatase Twine (Alphey *et al.* 1992; Courtot *et al.* 1992; Eberhart *et al.* 1996; White-Cooper *et al.* 1998; Maines and Wasserman 1999). *twine* mRNA is expressed in the primary spermatocyte growth phase and translationally repressed until meiosis (Alphey *et al.* 1992; Courtot *et al.* 1992). Translocation of the Boule from the nucleus to the cytoplasm leads to expression of Twine protein (Maines and Wasserman 1999). However, Twine function is not required for the onset of spermatid differentiation; *twine* mutants fail to undergo meiosis, but do not arrest at this stage (Alphey *et al.* 1992; Courtot *et al.* 1992; Lin *et al.* 1996). The meiotic arrest genes were found to control expression of *twine* and *boule*, with *twine* transcription dependent on *aly*, and *twine* translation (via Boule) dependent on *can*, *mia* and *sa* (White-Cooper *et al.* 1998).

In addition to *twine* and *boule* regulation, transcriptional control of several spermatid differentiation genes by the meiotic arrest genes *aly*, *mia*, *can* and *sa* was identified (White-Cooper *et al.* 1998; White-Cooper *et al.* 2000). Further analysis revealed new meiotic arrest genes, including the tandem duplicated *achintya (achi)* and *vismay (vis)*, *cookie monster (comr)*, *matotopetli (topi)*, and *tombola (tomb)* (Ayyar *et al.* 2003; Jiang and White-Cooper 2003; Wang and Mann 2003; Perezgasga *et al.* 2004; Jiang *et al.* 2007).

aly, *comr*, *topi*, *tomb* and *achi/vis* make up the Aly-class meiotic arrest genes (White-Cooper *et al.* 1998). Of these proteins, Comr, Topi, Tomb and Achi/Vis have DNA binding domains, and all but Achi/Vis are testis-specifically expressed (White-Cooper *et al.* 2000; Ayyar *et al.* 2003; Jiang and White-Cooper 2003; Perezgasga *et al.* 2004; Jiang *et al.* 2007; White-Cooper 2010; White-Cooper and Davidson 2011; Lim *et al.* 2012). Purification of mip40-containing complexes from nuclear extracts revealed the presence of a protein complex, the testis meiotic arrest complex (tMAC), which contained Aly, Tomb, Topi, Comr, Myb-interacting protein 40 (Mip40) and Chromatin assembly factor 1 (Caf1) (Beall *et al.* 2007). Co-immunoprecipitation of Achi/Vis found a second version of tMAC also containing Aly and Comr (Wang and Mann 2003).

All of the proteins comprising tMAC localise to chromatin in wild-type spermatocytes (White-Cooper *et al.* 2000; Ayyar *et al.* 2003; Beall *et al.* 2007). Localisation and normal functioning of all tMAC components is interdependent (Jiang and White-Cooper 2003; Beall *et al.* 2007). For example, Aly protein does not localise to chromatin in *comr* mutants, and vice versa (Ayyar *et al.* 2003; Jiang and White-Cooper 2003). This interdependent localisation and similarity of mutant phenotypes is evidence that the components of tMAC function together to promote transcription in primary spermatocytes.

can, *mia*, *sa*, *no hitter (nht)* and *ryan express (rye)* make up the *can*-class meiotic arrest genes (White-Cooper *et al.* 1998; Hiller *et al.* 2004; Metcalf and Wassarman 2007). The *can*-class meiotic arrest genes encode testis-specific paralogues of the TBP-associated factors (TAF) which, along with TATA-binding protein (TBP), are predicted to make up a testis-specific version of the basal transcription factor complex TFIID (Hiller *et al.* 2001; Hiller *et al.* 2004; Metcalf and Wassarman 2007). The general TFIID transcription factor complex promotes transcription by facilitating the interaction between RNA polymerase II and the genes to whose promotor regions it binds (Walker *et al.* 2001; Chen and Hampsey 2002; Matangkasombut *et al.* 2004). It was suggested that the testis TFIID complex may have a similar function, with the testis TAFs (tTAFs) binding to the promoters of testis specifically expressed genes, resulting in testis-specific expression (Figure 1.5). However, tTAFs have been found to localise to the nucleolus, in addition to the chromatin, and are required for the localisation of the Polycomb repression complex 1 (PRC1) to the nucleolus (Chen *et al.* 2005; Metcalf and Wassarman 2007). The tTAFs could be acting as a repressor of another repressor, thereby allowing transcription of testis-specific genes, as well as an activator of testis-specific gene expression (Figure 1.5) (Chen *et al.* 2005; Metcalf and Wassarman 2007; White-Cooper 2010; White-Cooper and Davidson 2011).

Both tMAC and the tTAFs are responsible for a large proportion of gene expression in the spermatocytes. Evidence from microarrays suggests that tMAC is required for the expression of over 1000 genes in the spermatocytes, while the tTAFs, either functioning as a testis-specific TFIID complex

or as a repressor of PCR1, upregulate 337 genes in spermatocytes and are required for expression of around 150 genes (Figure 1.5) (White-Cooper 2010; Doggett *et al.* 2011; Laktionov *et al.* 2014; Laktionov *et al.* 2018). There is also evidence that a subset of these genes are dependent on both tMAC and tTAFs for full expression (Laktionov *et al.* 2014).

1.3.2.3.2 Maintenance of germline specific gene expression by regulation of tMAC

More recently, new mechanisms of regulating germline-specific transcription in *Drosophila* have been described. While tMAC is an activator of transcription in spermatocytes, other mechanisms are required to prevent aberrant transcription of non-germline genes by tMAC. Kim *et al.* (2017) described the activity of a testis-specific zinc finger protein Kungang (Kmg) in spermatocytes, showing that it is required for repression of somatic-expressed genes in the germline (Figure 1.5).

Mutation experiments identified that Kmg is required in primary spermatocytes for repression of approximately 440 genes, normally transcribed in somatic cell types. RNA-seq identified that these ectopically expressed transcripts typically initiated from different transcription start sites (TSS) than those used in somatic cells, suggesting that they were expressed by activation of usually silent 'cryptic' promoters. Kmg co-immunoprecipitated with dMi-2, which was also required for repression of the 440 genes found to be repressed by Kmg. This indicated that Kmg and dMi-2 function together to repress somatic expression. Further analysis by ChIP-seq showed enrichment of Kmg and dMi-2 just downstream of the TSS of cryptic promoters, and that knockdown of Kmg resulted in reduced dMi-2 signal at these genes, indicating that Kmg recruits dMi-2 to the target regions. Knockdown of Kmg resulted in an enrichment of Aly detected at the cryptic promoters previously found to be enriched for Kmg in *wild-type* testes, suggesting that Kmg and dMi-2 prevent Aly accessing cryptic promoters, thus restricting expression to germline-specific promoters (Kim *et al.* 2017).

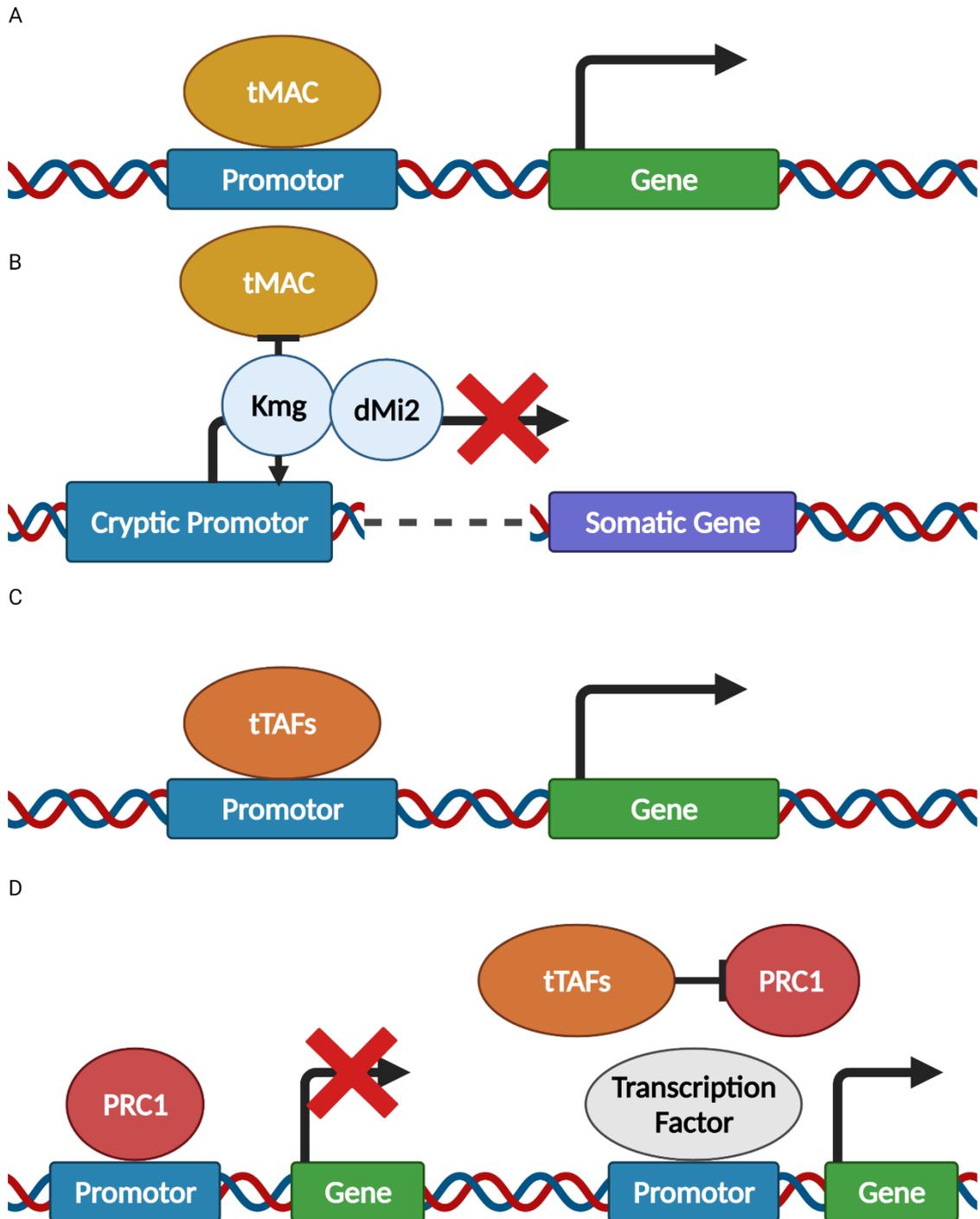


Figure 1.5: Models of transcriptional control in the primary spermatocytes. A: The testis meiotic arrest complex (tMAC) is required for germline expression of over 1000 genes. B: Kumgang (Kmg) maintains germline-specific expression by blocking tMAC from accessing cryptic promoters of somatic-expressed genes. Adapted from Kim et al. (2017). C-D: The testis paralogues of the TBP-associated factors (tTAFs) may contribute directly to germline-specific expression (activator), and block Polycomb repression complex 1 from repressing transcription of germline-expressed genes (repressor of a repressor).

1.3.2.4 Transcription and translation in spermatids

1.3.2.4.1 Translational repression of spermatocyte expressed genes required in later stages

tMAC and the tTAFs activate the transcription of over 1000 genes required in sperm development, of which some are required in the transition to meiosis (e.g. *twine*, *boule*), but also those required in post-meiotic spermatid differentiation (Lin *et al.* 1996; White-Cooper *et al.* 1998; White-Cooper *et al.* 2000; Ayyar *et al.* 2003; Jiang and White-Cooper 2003; Beall *et al.* 2007; Doggett *et al.* 2011). Translation of RNAs transcribed during spermatogenesis can be delayed until the gene product is required later in development (Kuhn *et al.* 1988; Schafer *et al.* 1995). RNA-binding proteins form ribonucleoprotein complexes with RNAs expressed during spermatogenesis, to repress translation. One such example is the RNA-binding protein Testis-specific RRM (Tsr), which is localised to the cytoplasm of spermatocytes and early spermatids. *tsr* mutants have reduced fertility, with spermatids showing defects in mitochondrial morphogenesis, and premature translation of some spermiogenesis proteins prior to meiosis (Haynes *et al.* 1997).

1.3.2.4.2 Post-meiotic transcription in spermatids

While the vast majority of transcription in sperm development occurs prior to meiosis (Olivieri and Olivieri 1965; Schafer *et al.* 1995), a small number of genes have been found to have post-meiotic expression, during elongation but prior to the replacement of histones with protamines during chromatin reconfiguration (Table 1.2) (Barreau *et al.* 2008a; Barreau *et al.* 2008b; Vibranovski *et al.* 2010). These post-meiotic expressed genes were termed the ‘comet’ and ‘cup’ genes, for their expression patterns in elongating spermatid bundles. Notably, the comet and cup transcripts localise to the distal ends of elongating spermatids, far from the nucleus (Barreau *et al.* 2008a; Barreau *et al.* 2008b). Of the comet and cup genes, at least one, *scotti* (*soti*), is required for male fertility; *soti* mutant spermatids fail to individualise. Post-meiotic transcription can therefore be essential for correct spermiogenesis.

Expression Pattern	Gene	Symbol	CG number
Comet	<i>schumacher-levy</i>	<i>schuy</i>	CG17736
	<i>hale-bopp</i>	<i>hale</i>	CG7570
	<i>sungrazer</i>	<i>sunz</i>	CG15179
	<i>solwind</i>	<i>sowi</i>	CG15178
	<i>borrelly</i>	<i>boly</i>	CG30362

	<i>comas sola</i>	<i>cola</i>	CG30363
	<i>hug-bell</i>	<i>hubl</i>	CG30364
	<i>spacewatch</i>	<i>spaw</i>	CG30365
	<i>whipple</i>	<i>whip</i>	CG34218
	<i>swift-tuttle</i>	<i>swif</i>	CG30366
	<i>scotti</i>	<i>soti</i>	CG8489
Cup	<i>calcutta-cup</i>	<i>c-cup</i>	CG15623
	<i>walker-cup</i>	<i>wa-cup</i>	CG10113
	<i>ryder-cup</i>	<i>r-cup</i>	CG10998
	<i>davis-cup</i>	<i>d-cup</i>	CG14387
	<i>presidents-cup</i>	<i>p-cup</i>	CG12993
	<i>world-cup</i>	<i>w-cup</i>	CG7363
	<i>stanley-cup</i>	<i>s-cup</i>	CG30044
	<i>tetleys-cup</i>	<i>t-cup</i>	CG31858
	<i>flyers-cup</i>	<i>f-cup</i>	CG9611
	<i>heineken-cup</i>	<i>h-cup</i>	CG6130
	<i>mann-cup</i>	<i>m-cup</i>	CG11896
	<i>oo18 RNA-binding protein</i>	<i>orb</i>	CG10868

Table 1.2: Post-meiotically expressed comet and cup genes. Table from Barreau et al. (2008a).

1.4 Heteromorphic Sperm Development in *D. pseudoobscura* and the *obscura* Species Group

The system most commonly used for the study of sperm development in *Drosophila* is *D. melanogaster*, and the vast majority of research into the genetic control of sperm development has been carried out using this species. Of the sperm heteromorphic species within the *obscura* group, there have been some studies of the testis and cyst structure, and of the structure of mature heteromorphic spermatozoa. A few studies have characterised a small number of genetic components

of spermatogenesis and spermiogenesis. However, the developmental and molecular processes of heteromorphic spermatogenesis remain almost entirely unknown.

1.4.1 Current research has not addressed the genetic control of heteromorphic sperm development

The general process of sperm development in *D. pseudoobscura* is assumed to be similar to that of *D. melanogaster*; germline stem cells and cyst cells are maintained in the testis niche by hub and soma-germline signalling, GSCs differentiate and enter a transit amplification stage of mitotically dividing spermatogonia, then differentiate into spermatocytes which undergo meiosis and subsequently differentiate into mature spermatozoa. However, there must be some differences in the developmental process to produce the three distinct sperm morphs.

The testis niche is the source of all germline cells in the testis. In *D. melanogaster*, the testis niche comprises the hub, approximately 10-15 somatic cells, surrounded by 6-9 germline stem cells interspersed with cyst stem cells (Figure 1.2) (Hardy *et al.* 1979). The structure of the testis niche appears to be somewhat different in *D. pseudoobscura* testes. Work in the Klaus lab has found that immunofluorescence staining of the testis for the hub marker Fasciclin III (FasIII) shows a much larger hub region in *D. pseudoobscura* and its sister species *D. persimilis*, compared to that of *D. melanogaster* (Beaury 2012; Mena 2012; Cardaci 2014).

Early research on sperm heteromorphism in *D. pseudoobscura* noted the presence of only one morph in each spermatid cyst (Beatty and Sidhu 1969). Subsequent studies have found this is also the case in other *obscura* group species (Hauschteck-Jungen and Maurer 1976; Bircher *et al.* 1995). It has been suggested that morph specialisation occurs early in development, either as GSCs, predicting the presence of multiple GSC populations within the testis niche, or shortly after GSC differentiation, prior to the onset of mitotic divisions (Moore *et al.* 2013).

The potential link between the large hub region observed in *D. pseudoobscura* testes and the development of multiple sperm morphs has not been previously discussed. There is no direct evidence indicating at which point morph specialisation occurs during sperm development. It is likely that morph specialisation occurs early in *D. pseudoobscura* spermatogenesis, and as such, it would be expected that the number of GSCs present in the *D. melanogaster* niche (6-9), would not be sufficient to sustain three distinct GSC populations over the whole lifetime of the fly, due to stochastic loss. *D. pseudoobscura* require a greater number of GSCs overall if there are three GSC populations. The number of GSCs in the *D. pseudoobscura* testis niche is not yet known. Alternatively, a single GSC

population may exist, with each GSC able to produce daughter cells contributing to any of the three morphs.

The mechanisms involved in spermatogenesis and spermiogenesis in *D. pseudoobscura* are not well understood. One recent study has investigated the role of actin related proteins in *D. pseudoobscura* spermiogenesis, specifically the case of *Actin related protein 2 (Arp2)* and a retroduplicated copy, *Arp2D*, which appears to have arisen in the *obscura* lineage. Schroeder *et al.* (2020) found that the *obscura*-specific Arp paralogues have evolved under positive selection, indicating some non-redundant function. *Arp2* has ubiquitous expression, whereas *Arp2D* is testis specific. *Arp2D*, tagged with GFP, was detected in meiotic and post-meiotic cysts, localising to cytoplasm. During spermatid individualisation *Arp2D* localised to the actin cones of both eusperm and parasperm cysts, indicating that it is not involved in the differential development of heteromorphic sperm (Schroeder *et al.* 2020).

There is an indication that parasperm are less energetically costly and require less time to develop. Bircher *et al.* (1995) found a higher frequency of short cysts and lower frequency of long cysts in *D. subobscura* males less than one day old, compared to four days old, suggesting that short cysts require less time for maturation. Males which mate frequently have been shown to produce more short sperm, indicating a higher cost for the production of long sperm (Snook 1998a; Alpern *et al.* 2019).

It is evident that there has been comparatively little research into sperm development in *D. pseudoobscura*. So far, there is no published data on the mechanisms and processes which take place during spermatogenesis and spermiogenesis which contribute specifically to the differential development of heteromorphic sperm in *D. pseudoobscura*.

1.4.2 Evolution of testis genes; a role in heteromorphic sperm development?

A question that arises from the phenomenon of *D. pseudoobscura* sperm heteromorphism is; how did sperm heteromorphism evolve in this and the other *obscura* species? One potential explanation lies in the mechanisms by which novel genes gain function in the testes.

1.4.2.1 'Out of the Testis' – testis genes are fast evolving

Testis-expressed genes have been found to be fast evolving compared to genes expressed in other tissues, as the testes are under strong selective pressures (Kaessmann 2010; Assis and Bachtrog 2013). Furthermore, the testes have been described as having more promiscuous expression of otherwise non-functional or non-transcribed genetic elements, compared to other tissues, which may result from open chromatin during meiosis (Kleene 2005; Kaessmann 2010). Young genes lacking regulatory elements (e.g. as a result of retrotransposition) may still have expression in the testes and gain exposure to selective forces (Assis and Bachtrog 2013; Kondo *et al.* 2017). This has been found in

studies of retroduplicated genes, with young duplicated genes showing testis-biased expression, and older duplicates showing more diverse expression, supporting a role for testis expression prior to functional diversification (Vinckenbosch *et al.* 2006; Bai *et al.* 2007; Kaessmann *et al.* 2009; Assis and Bachtrog 2013).

1.4.2.2 Duplicated genes in *D. pseudoobscura*

Studies of gene duplications in *D. pseudoobscura* have found a range of gene duplication events since the divergence from the most recent common ancestor with *D. melanogaster*. Meisel (2009a) identified 88 gene duplication events comprising a total of 101 genes, of which eight were the result of retrotransposition, and 46 were DNA duplications (Meisel 2009a, b). A subsequent study found 173 duplicate pairs in *D. pseudoobscura* (Assis and Bachtrog 2013). Classification of the evolutionary processes retaining these 173 duplicate pairs revealed that the majority were maintained by neofunctionalisation of the child copy. Conservation and specialisation were also found to contribute to retention of duplicated genes. Subfunctionalisation was rarely identified as contributing to the retention of duplicated genes in *D. pseudoobscura* (Assis and Bachtrog 2013).

Derived copies of duplicated genes in *D. pseudoobscura* tend to have testis-biased or testis-specific expression (Meisel *et al.* 2010). Three of the four RNAi pathway nuclease *Argonaute 2 (Ago2)* segmentally duplicated paralogues show testis-specific expression, with one paralogue demonstrating the ancestral condition of ubiquitous expression, although this paralogue does not appear to be the ancestral copy (Lewis *et al.* 2016; Crysnanto and Obbard 2019). As described previously, duplicated Arp2 proteins have been found to have specific functions in spermiogenesis (Schroeder *et al.* 2020). Gene duplications may have allowed the divergence of sperm morphology observed in the *obscura* group and may also take on roles in different sperm morphs (Schroeder *et al.* 2020), although there are no known examples of this.

Genes involved in sperm heteromorphism may have evolved function as a result of gene duplication, but may also have arisen as a result of adaptation of genes, pathways and mechanisms which were already present in the ancestral *obscura* lineage, for example by adaptation of transcriptional and translational control, or alternative splicing.

Examination of the genetic, molecular and developmental processes leading to heteromorphic sperm differentiation in *D. pseudoobscura* may shed further light on the processes by which sperm heteromorphism arose in the *obscura* species group.

1.4.3 Sperm development as a model system

Drosophila testis has been widely used as a model system. In particular, *Drosophila* spermatogenesis is a useful model of the mechanisms by which stem cells differentiate. The benefits of this model are clear: a single stem cell population gives rise to a single mature cell population with well-characterised, definable developmental stages in between. *Drosophila* testes are a useful model for the processes of cell division and cell cycle regulation, and of the mechanisms of gene activation and transcription in cell differentiation (White-Cooper 2004, 2010; Spradling *et al.* 2011; White-Cooper and Davidson 2011; Lim *et al.* 2012). The *Drosophila* germline stem cell niche can also be used as a model system for stem cell niches more widely, including those of mammals (Lin 2002).

The study of *D. pseudoobscura* spermatogenesis offers the opportunity to expand the model system, allowing investigation into the development of multiple similar, but distinct, mature cell populations within a tissue. Understanding the molecular mechanisms of sperm heteromorphism will allow insight into the processes of cell differentiation and the evolutionary processes resulting in novel cell morphology, in particular the contributions of gene duplication and evolution.

1.5 Techniques for the Study of Sperm Development in *Drosophila*

The techniques available for studying gene function in *Drosophila* spermatogenesis have been reviewed elsewhere, and are summarised in the following section (White-Cooper 2004; Singh and Hou 2008; White-Cooper 2009; Demarco *et al.* 2014). I will discuss the methods which can be adapted for use in other *Drosophila* species, for which there are fewer resources currently available.

1.5.1 Genome assemblies and sequencing methods

The *D. melanogaster* genome sequence was first published in 2000, and has been continually updated since (Adams *et al.* 2000; Celniker and Rubin 2003; Clark *et al.* 2007; dos Santos *et al.* 2015; Gramates *et al.* 2017; Thurmond *et al.* 2019; Larkin *et al.* 2021). A further 12 genome assemblies, including *D. pseudoobscura*, were published in subsequent years (Clark *et al.* 2007). More recently, genome sequencing using Oxford Nanopore has allowed an expansion in the number of species and lines for which genome assemblies are available (Kim *et al.* 2021). Tissue-specific gene expression data is available for *D. melanogaster*, including imaginal disc, central nervous system, head, salivary gland, digestive system, fat body, carcass, ovary and testis (Larkin *et al.* 2021). Tissue-specific gene expression data of male and female whole fly, head, carcass, ovary and testis is also available for previous *D. pseudoobscura* genome assembly releases (Gramates *et al.* 2017).

Readily available sequence data greatly enhances the number of techniques available for use in *D. pseudoobscura*. Sequencing techniques, such as RNA-seq, ChIP and spatial sequencing can be applied

to *D. pseudoobscura*, and analysed using the available reference genomes. It also simplifies PCR-based techniques, as primers and target regions can be easily identified.

1.5.2 Staining methods can be used to temporally and spatially localise transcripts and proteins

RNA *in situ* hybridisation can be used to examine transcript localisation within the testes, and in some cases, such as elongated spermatids, the subcellular localisation within cysts (Barreau *et al.* 2008a; Morris *et al.* 2009). Transcript localisation indicates the stages at which the target transcript is expressed, although there are limits to detection. Disappearance of staining in later stages is often well correlated with the point at which translation has occurred (Barreau *et al.* 2008a; Barreau *et al.* 2008b).

Immunofluorescence, using antibodies raised against the protein of interest, can show localisation of the proteins within the testes, and therefore the stages at which they function. Antibodies are readily available for multiple *D. melanogaster* protein markers of the hub, germline and somatic cells (Singh and Hou 2008). The relative similarity between *D. melanogaster* and *D. pseudoobscura* suggests that most of these reagents should also work in this related species (Beaury 2012; Mena 2012; Cardaci 2014).

1.5.3 Gene editing techniques

Gene editing techniques allow greater insight into gene function. The development of CRISPR/Cas9 gene editing has greatly expanded the ability to generate new, highly controlled mutants in *Drosophila* and many other species (Bassett *et al.* 2013; Gratz *et al.* 2013; Kondo and Ueda 2013; Ren *et al.* 2013; Gratz *et al.* 2014; Port *et al.* 2014; Wright *et al.* 2016). Gene editing can be used to generate new endogenous tags of genes of interest, complete or partial gene knockouts, DNA visualisation and modification of the epigenome, among many other potential applications (Pulecio *et al.* 2017; Bier *et al.* 2018; Bukhari and Muller 2019).

CRISPR gene editing methods are constantly being developed and optimised for use in a greater number of species, and have great potential in non-model species, in particular for the investigation of gene function.

1.5.4 Genetic tools for investigating gene function in *Drosophila* spermatogenesis

Many genetic tools are available for the study of gene function in *D. melanogaster*. These include UAS-Gal4 lines for control and analysis of tissue-specific gene function, RNAi-mediated gene expression

knockdown, and the wide availability of many transgenic lines developed with P-element- and phiC31-mediated insertion (Demarco *et al.* 2014). As CRISPR gene editing becomes more widely available, many new transgenic lines are constantly being developed.

There is far less published research regarding *D. pseudoobscura* compared to *D. melanogaster*, and as a result there is limited available data or previously tested and optimised techniques. Genome assemblies are available, as is tissue-specific transcription data. However, commonly used methods in *D. melanogaster*, such as RNAi and UAS-Gal4, are not available for use. The vast availability of mutant lines in *D. melanogaster* is not replicated in *D. pseudoobscura*, beyond development of *white* and *yellow* mutants by a few individual labs (Phadnis Lab, referenced in Schroeder *et al.* 2020)

Some general methods are easily adaptable for use in *D. pseudoobscura*, given genome sequences and tissue-specific transcriptome data are available. PCR-based methods, sequencing and RNA *in situ* hybridisation staining can be easily applied to *D. pseudoobscura* with little to no optimisation necessary. The use of piggyBac P-element insertion to generate tagged proteins has been applied in many non-model insect species, and more recently in *D. pseudoobscura* (Handler *et al.* 1998; Horn and Wimmer 2000; Holtzman *et al.* 2010; Tanaka *et al.* 2016; Schroeder *et al.* 2020). Immunostaining using antibodies raised against *D. melanogaster* proteins have been used to some success in *D. pseudoobscura*, and it may be possible to further develop this method (Beaury 2012; Mena 2012; Cardaci 2014).

It should also be possible to adapt methods used in *D. melanogaster* for use in *D. pseudoobscura*, of which CRISPR/Cas9 gene editing would be extremely useful. Early applications of CRISPR/Cas9 in *D. melanogaster* utilised expression vectors to inject *cas9* DNA or *in vitro* transcribed *cas9* mRNA directly into the posterior region of the embryo (Bassett *et al.* 2013; Gratz *et al.* 2013). Other approaches utilised lines expressing endogenous *cas9* in the germline. In this approach, *cas9* is integrated into the genome under a germline promotor, such as *vasa* or *nanos*, and embryos are then injected with an expression vector encoding the sgRNA (Ren *et al.* 2013; Sebo *et al.* 2014). The benefit of this approach is that mutagenesis in the somatic cells is limited, since the Cas9 is localised to the germline. It has also been found to have high efficiency, approximately 74% in *D. melanogaster*, and is both cost and time effective (Ren *et al.* 2013). The Phadnis Lab (University of Utah) has had success with CRISPR in *D. pseudoobscura* by direct injection of the Cas9 protein (Dean Castillo, pers. comm. 2021). The development of lines expressing endogenous Cas9 would be of great use, allowing easier generation of endogenous tags and mutants.

1.6 Aims and Key Findings

Sperm development in *Drosophila* is controlled by a combination of many developmental, genetic and molecular mechanisms. Previous research suggests that the formation of heteromorphic sperm in *D. pseudoobscura* is adaptive, regulated, and genetically controlled, however little is understood regarding specific mechanisms. In this project, I have aimed to answer the question – what are the developmental and molecular mechanisms contributing to the differential development of heteromorphic sperm in *D. pseudoobscura*?

The first aim of this project was to establish the role of transcription in differential sperm development. I have started with an investigation of gene expression in the primary spermatocyte cysts and post-meiotic spermatids of *D. pseudoobscura*, using a single cyst RNA sequencing approach. By analysing cyst RNA-seq data, I aimed to establish whether multiple spermatocyte cyst populations are identifiable by transcriptional similarities, and to identify genes that were significantly different in transcript abundance between cyst populations. The majority of gene expression occurs in the primary spermatocyte stage of *Drosophila* spermatogenesis, including that which is required for spermatid differentiation, post-meiosis (Olivieri and Olivieri 1965; Lindsley and Tokuyasu 1980; Fuller 1993; White-Cooper 2010; Lim *et al.* 2012). It was therefore of interest to identify mechanisms which may influence transcriptional control differentially between the sperm morphs, during the spermatocyte stage. Some gene expression occurs post-meiosis, and functions in spermatid elongation and individualisation (Barreau *et al.* 2008a). Given the differences between sperm morphs in length, it is also of interest to investigate the potential differences in post-meiotic gene expression and persistence of pre-meiotic transcription in the spermatid cysts of *D. pseudoobscura*.

RNA sequencing allowed identification of candidate genes of interest, which were investigated for RNA localisation within the *D. pseudoobscura* testis, using an RNA *in situ* hybridisation method, previously optimised for use in *D. melanogaster* testes, and used in our lab in *D. pseudoobscura* (Morris *et al.* 2009). *In situ* hybridisation revealed distinct patterns of differential gene expression in primary spermatocytes in *D. pseudoobscura* testes, including genes with *D. melanogaster* orthologues known to have functions in regulating transcription in primary spermatocytes, notably *kumgang*, *caf1* and *achi/vis*.

I hypothesised that Kmg protein is differentially abundant between spermatocyte cyst morphs. To test this, I have used a piggyBac insertion construct to insert a GFP-tagged copy of the *kmg* gene, generating lines expressing a Kmg-GFP fusion. Examination of the testes of Kmg-GFP *D. pseudoobscura* revealed that Kmg protein is present in the nuclei of all primary spermatocyte cysts, but is also differentially expressed between spermatocyte cysts. This suggests that Kmg has an essential function

in *D. pseudoobscura* spermatogenesis, and also contributes to specific heteromorphic sperm development of at least one morph. I discuss these findings in relation to the current understanding of Kmg function in *D. melanogaster*.

The overall structure of the *D. pseudoobscura* testis apical proliferation centre is poorly described. To further our understanding of this region I have examined the structure of the apical proliferation centre in *D. pseudoobscura*, using an immunofluorescence method to examine the size and substructure of the hub and testis niche in this species. Using immunofluorescence to stain hub, germline and somatic markers, I have compared the structures of the *D. melanogaster* and *D. pseudoobscura* apical proliferation centres, finding that the *D. pseudoobscura* hub is much larger in size but also appeared more diffuse, suggestive of a large flat epithelial hub structure to which GSCs and CySCs adhere.

A major challenge associated with the use of a non-model system is the lack of availability of tools for genetic analysis. To expand the genetic toolbox available for *D. pseudoobscura*, I aimed to develop *D. pseudoobscura* lines endogenously expressing Cas9, localising to the poleplasm of the embryos. Use of the piggyBac transposon system for insertion of transgenes was successful in *D. pseudoobscura*. I then aimed to validate the system with *white* and *yellow* *D. pseudoobscura* mutants, before then using it to generate endogenously tagged proteins of interest, and allowing easier generation of mutants in *D. pseudoobscura*.

2 Materials and Methods

2.1 Fly Stock Maintenance

Wild-type *D. pseudoobscura* SLOB3 (obtained from Tom Price, University of Liverpool) were maintained between 21-25°C. *D. pseudoobscura* maintained at 25°C in an incubator were kept on a 12:12h light:dark cycle. All stocks were maintained on food medium in vials and bottles (Table 2.1).

Fly Food Component	Amount per Litre of Food
Dextrose	13.75g
Maize	12.5g
Yeast	8.75g
Agar	7.5g
180mL Propionic Acid	2.25mL
10% Nipagen in Ethanol	19mL
H ₂ O (Sterile)	Adjust to 1L

Table 2.1: Components required for fly food. All flies were maintained on food in vials or bottles.

2.1.1 Laying pots

Laying pots were set up to collect embryos at 1-1.5 hour intervals. Laying plates containing a cranberry juice agar with yeast paste were attached to the bottom of a plastic laying pot. Cranberry juice agar was made using juice which contained additional sucrose, but not *Stevia*-based sweeteners, as some components of these have been implicated in reducing longevity in *Drosophila* (Baudier *et al.* 2014).

Cranberry juice agar was made by dissolving 6g sucrose in 60mL boiling cranberry juice, which was then cooled for at least ten minutes. 6g agarose was dissolved in 200mL boiling distilled water. 3mL 10% nipagen in ethanol was added to the cooled sucrose/cranberry juice solution. The agarose was allowed to cool until it could be held. The juice solution was added and mixed by swirling. The agarose was poured into 60mm petri dishes and allowed to set.

Yeast paste was made by mixing fresh yeast (Allinson) with water that had been boiled and cooled to approximately 40°C. Yeast paste was made in batches and stored at 4°C for up to two weeks.

2.2 Fly Work and Molecular Biology Methods

2.2.1 Testis dissection

Flies were anaesthetised by CO₂ and males removed to a petri dish lid containing several drops of testis buffer (183mM KCl, 47mM NaCl, 10mM Tris HCl, pH6.8). Under a Leica MZ10F dissection microscope (Leica Microsystems), flies were positioned with the abdomen at the edge of the testis buffer drop. The head and thorax were held with a pair of forceps, while a second pair of sharp forceps were used to pull the end of the abdomen towards the testis buffer. The testes were removed from the abdomen into the buffer.

2.2.2 Testis RNA extraction and cDNA synthesis

Testis RNA extraction and cDNA synthesis was carried out using the method described in Morris *et al.* (2009).

2.2.2.1 Testis RNA extraction

20 testes were dissected and transferred to the lid of a 1.5mL Eppendorf tube containing 200µL Trizol (Invitrogen), the tube was closed and inverted several times. 800µL Trizol was added and the testes homogenised by drawing up and down through a fine-gauge syringe needle. The homogenate was incubated at room temperature for five minutes, to allow dissociation of nucleoprotein complexes. 200µL chloroform was added and incubated for 2 minutes at room temperature. The sample was centrifuged at 12000 x *g* for 5 minutes, at 4°C. The upper (aqueous) phase containing the RNA was transferred to a new tube.

500µL isopropanol was added, mixed by inverting and the sample incubated for 10 minutes on ice. The sample was centrifuged at 12000 x *g* for 10 minutes, at 4°C. RNA formed a translucent white pellet at the bottom of the tube. The supernatant was discarded.

The pellet was washed in 75% ethanol, vortexed briefly, and centrifuged at 7500 x *g* for 5 minutes, at 4°C. The supernatant was removed and the RNA pellet air-dried for 10 minutes. The RNA pellet was resuspended in 20µL dH₂O (Invitrogen).

2.2.2.2 cDNA synthesis from testis RNA

cDNA was synthesised using the SuperScript III First Strand Synthesis System (Invitrogen) following the method described in Morris *et al.* (2009).

The template mixture was prepared as described in Table 2.2. Secondary structure was denatured by heating the template mixture to 65°C for 5 minutes followed by immediate cooling on ice for at least

1 minute. 6µL of reaction mix (Table 2.3) was added to the template mix, mixed, and incubated at 50°C for 2 minutes.

1µL SuperScript III reverse transcriptase enzyme was added to each reaction mix. The reaction mix was incubated at 50°C for 1 hour. The reaction was terminated by incubating at 70°C for 15 minutes.

Synthesised testis cDNA was stored at -80°C.

Component	Volume (µL)
Oligo (dT)₂₀ primer (50µM)	1
Total RNA	4
10mM dNTP mix	1
Nuclease-free water	7

Table 2.2: Template mixture for cDNA synthesis from testis RNA.

Component	Volume (µL)
5X RT buffer	4
100mM DTT	1
RNAse OUT (40 U µL⁻¹)	1

Table 2.3: Reaction mix for cDNA synthesis from testis RNA.

2.2.3 DNA extraction

2.2.3.1 General DNA extraction method

40 flies were homogenised in a tube containing 1mL of 0.1M Tris-HCl pH9.0, 0.1M EDTA pH8.1, 1% SDS and 1% DEPC. The sample was transferred to ice. 140µL 8M potassium acetate added and the sample vortexed to mix. The sample was centrifuged at 15000 x *g* for 5 minutes. The supernatant was removed to a fresh tube. 500µL isopropanol was added, and the sample mixed by inverting. The sample was incubated at room temperature for 5 minutes, then centrifuged at 15000 x *g* for 10 minutes at 4°C. The supernatant was discarded and the DNA pellet was washed in 70% ethanol. The ethanol was removed and the pellet air-dried for 10 minutes. The pellet was resuspended in 400µL TE buffer (100mM Tris pH7.5, 1mM EDTA).

400µL of 50:49:1 phenol:chloroform:isoamyl alcohol was added to the resuspended sample. The solution was vortexed, then centrifuged at 15000 x *g* for 1 minute. The upper phase was removed to a new tube. 400µL chloroform added and the sample mixed. The upper phase was removed to a new tube. To precipitate the DNA, a 1/10 volume of 3M sodium acetate and 3 volumes 100% ethanol were added to the sample and vortexed to mix, then centrifuged at 15000 x *g* for 10 minutes at 4°C.

The DNA pellet was washed in 70% ethanol and centrifuged at 15000 x *g* for 5 minutes. The supernatant was discarded. The DNA pellet was dried and resuspended in TE buffer.

2.2.3.2 Single fly DNA extraction

A single fly was anaesthetised and transferred to a 1.5mL Eppendorf tube. 50µL squishing buffer (10mM Tris pH 8.0, 1mM EDTA, 25mM NaCl, 0.2mg/mL proteinase K) was aspirated into a 200µL pipette tip. The fly was squished with the pipette tip, without expelling the squishing buffer. The squishing buffer was expelled and the sample mixed by pipetting. The sample was incubated at room temperature for 20 minutes. The reaction was stopped by incubation at 95°C for 2 minutes, then transferred to ice.

The sample was centrifuged at 15000 x *g* for 5 minutes. The supernatant was removed to a new tube and stored at -20°C (Evans Lab 2014).

2.2.4 PCR

Several protocols were used for PCR, according to the optimum conditions of the DNA polymerase and primers.

2.2.4.1 GoTaq® G2 Master Mix (Promega)

All reaction mixtures were prepared on ice. Samples were mixed by pipetting 5 times, then briefly centrifuged. The reaction mix was prepared according to the conditions in Table 2.4, in a 0.2mL PCR tube. The samples were transferred to a PCR machine and PCR run according to the conditions in Table 2.5.

Component	Volume (µL)
GoTaq G2 Green Master Mix 2X	12.5
10µM Forward Primer	1
10µM Reverse Primer	1
DNA Template	0.5-5
Nuclease Free Water	To final volume 25µL

Table 2.4: Components required for PCR reaction mix using GoTaq® Master Mix (Promega).

Step Number	Step	Temperature (°C)	Time
1	Denaturation	95	2 mins
2	Amplification	95	30s
3		55*	20s
4		72	1min/kb**
			x30 cycles

5	Final Extension	72	5 mins
----------	------------------------	-----------	---------------

Table 2.5: Reaction conditions for GoTaq® Master Mix (Promega) PCR. *Step 3 was conditional on the predicted melting temperature (T_m) of the primers, designed to be between 55-60°C, and within +/- 1°C for each primer pair. **Step 4 was conditional on the extension time required for the expected PCR product; for every 1 kilobase in length of PCR product, an extension of 60 seconds was required.

2.2.4.2 Taq DNA Polymerase (NEB)

The reaction mix was prepared according to the conditions in Table 2.6, in a 0.2mL PCR tube. Samples were mixed by pipetting 5 times, then briefly centrifuged. The samples were transferred to a PCR machine and PCR run according to the conditions in Table 2.7.

Component	Volume (μ L)
10x PCR Buffer	4
10mM dNTPs	0.8
10μM Forward Primer	0.8
10μM Reverse Primer	0.8
cDNA Template	2
Taq DNA Polymerase	0.4 μ L
dH₂O	To total volume 40 μ L

Table 2.6: Components required for PCR reaction mix using Taq DNA Polymerase (NEB).

Step Number	Step	Temperature (°C)	Time
1	Initial denature	95	2min
2	Denature	95	30s
3	Anneal	55*	30s
4	Extend	68	1min/kb**
5	Final extension	68	5mins

Table 2.7: Reaction conditions for Taq DNA Polymerase (NEB) PCR. *Step 3 was conditional on the predicted melting temperature (T_m) of the primers, designed to be between 55-60°C, and within +/- 1°C for each primer pair. **Step 4 was conditional on the extension time required for the expected PCR product; for every 1 kilobase in length of PCR product, an extension of 60 seconds was required.

2.2.4.3 Q5® High-Fidelity Master Mix (NEB)

The reaction mix was prepared according to the conditions in Table 2.8, in a 0.2mL PCR tube. Samples were mixed by pipetting 5 times, then briefly centrifuged. The samples were transferred to a PCR machine and PCR run according to the conditions in Table 2.9.

Component	Volume (μL)
Q5[®] High-Fidelity 2X Master Mix	12.5
10μM Forward Primer	1.25
10μM Reverse Primer	1.25
DNA Template	0.5-2
Nuclease Free Water	To final volume 25 μL

Table 2.8: Components required for PCR reaction mix using Q5[®] High-Fidelity Master Mix (NEB).

Step Number	Step	Temperature ($^{\circ}\text{C}$)	Time
1	Initial denature	98	30s
2	Denature	98	10s
3	Anneal	55*	20s
4	Extend	68	30s/kb**
5	Final extension	72	2mins

Table 2.9: Reaction conditions for Q5[®] High-Fidelity 2X Master Mix (NEB) PCR. *Step 3 was conditional on the predicted melting temperature (T_m) of the primers, designed to be between 55-60 $^{\circ}\text{C}$, and within $\pm 1^{\circ}\text{C}$ for each primer pair. **Step 4 was conditional on the extension time required for the expected PCR product; for every 1 kilobase in length of PCR product, an extension of 30 seconds was required.

2.2.5 PCR purification

When required, PCR products were purified with the QIAquick[®] PCR Purification Kit, according to the manufacturer's protocol.

5 volumes of Buffer PB were added to 1 volume of PCR reaction and mixed. The sample was applied to a QIAquick column with 2mL collection tube and centrifuged for 1 minute to bind the DNA to the column membrane. Flow through was discarded and the collection tube replaced.

DNA was washed by applying 750 μL Buffer PE to the QIAquick column and centrifuging for 1 minute. Flow through was discarded and the collection tube replaced. The sample was centrifuged for a further minute to remove residual wash buffer.

The QIAquick column was placed in a clean 1.5mL Eppendorf tube. 30 μL sterile water, preheated to 50 $^{\circ}\text{C}$, was added to the centre of the column membrane and centrifuged for 1 minute to elute the purified DNA.

2.2.6 Restriction digest

Restriction digestion of DNA was carried out according to a standard protocol. Incubation temperature and enzyme inactivation conditions were determined by the restriction enzyme(s) required by the protocol. Reaction mix composition was determined by DNA concentration and the number of restriction enzymes required (Table 2.10). The reaction mixture was added to a 1.5mL Eppendorf tube and incubated in a heat block for up to 1 hour. The enzymes were either inactivated by transfer to a 65/80°C heat block (as appropriate, Table 2.11), or the reaction mix purified with the QIAquick PCR Purification kit (Qiagen). Digested DNA was stored at -20°C.

Component	Volume
DNA	1µg (volume according to concentration)
Enzyme 1	1µL
Enzyme 2	1µL
10X NEB CutSmart Buffer	5µL
dH₂O	To total volume 50µL

Table 2.10: Reaction mix for restriction digestion with either one or two restriction enzymes in the same buffer (New England BioLabs).

Restriction Enzyme (NEB)	Buffer	Incubation Temperature (°C)	Inactivation Conditions
EcoRI-HF	CutSmart® (NEB)	37	65°C for 20 minutes
HindIII-HF	CutSmart® (NEB)	37	80°C for 20 minutes
KpnI-HF	CutSmart® (NEB)	37	DNA purification by QIAquick PCR Purification (Qiagen)
NotI-HF	CutSmart® (NEB)	37	65°C for 20 minutes
SacII	CutSmart® (NEB)	37	65°C for 20 minutes
TaqαI	CutSmart® (NEB)	65	DNA purification by QIAquick PCR Purification (Qiagen)
MspI	CutSmart® (NEB)	37	DNA purification by QIAquick PCR Purification (Qiagen)
HaeIII	CutSmart® (NEB)	37	80°C for 20 minutes

AvrII	CutSmart® (NEB)	37	DNA purification by QIAquick PCR Purification (Qiagen)
BbsI	NEBuffer™ r2.1: 100%	37	65°C for 20 minutes

Table 2.11: Restriction endonucleases, conditions for restriction digest of DNA (buffers, incubation temperature), and method of enzyme inactivation (heat inactivation, enzyme removal).

2.2.7 Agarose gel electrophoresis

Agarose gel electrophoresis was used to examine the products of PCR reactions and restriction digestion.

Agarose gels were made by dissolving agarose in 1x TAE buffer (diluted from 50x stock solution; 50mM EDTA, 2M Tris, 1M acetic acid, in dH₂O). Agarose concentration varied depending on the expected size of DNA product, e.g. 0.8-3% weight to volume. 5µL SafeView Nucleic Acid Stain (NBS Biologicals) was added for every 100mL of agarose.

The agarose was poured into a gel mould and allowed to set. The gel was transferred to an electrophoresis tank and covered with 1x TAE buffer. The DNA samples were loaded with 5:1 DNA:6x bromophenol blue dye. A molecular weight standard DNA ladder was also loaded into one well. Ladders used were dependent on the predicted base pair length of DNA, e.g. 100bp DNA Ladder (NEB) and 1kb Plus DNA Ladder (NEB). Electrophoresis was run until the bromophenol blue dye had run approximately half to two-thirds of the length of the gel.

2.2.8 Agarose Gel Extraction

Agarose gel extraction was used as a method of DNA purification, using the QIAquick® Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

The DNA fragment was excised from the agarose gel with a clean scalpel and transferred to a clean 1.5mL tube. 3 volumes of Buffer QG were added to 1 volume of excised gel (100mg = 100µL). The sample was incubated in a 50°C heat block for 10 minutes with regular mixing to dissolve the gel.

1 volume isopropanol was added to the sample and mixed by vortexing. The sample was applied to a QIAquick column with 2mL collection tube and centrifuged for 1 minute to bind the DNA to the column membrane. Flow through was discarded and the collection tube replaced. 500µL Buffer QG was added to the column and centrifuged for 1 minute. Flow through was discarded and the collection tube replaced. DNA was washed by applying 750µL Buffer PE to the QIAquick column and centrifuged for 1

minute. Flow through was discarded and the collection tube replaced. The sample was centrifuged for a further minute to remove residual wash buffer.

The QIAquick column was placed in a clean 1.5mL Eppendorf tube. 30µL sterile water, preheated to 50°C, was added to the centre of the column membrane and centrifuged for 1 minute to elute the purified DNA.

2.2.9 DNA ligation by T4 DNA Ligase (NEB)

Plasmid or construct fragments were ligated at restriction sites with T4 DNA Ligase (NEB), according to the manufacturer's protocol. Unless specified otherwise, a 1:3 vector to insert molar ratio was used.

The reagents were added to a microcentrifuge tube on ice (Table 2.12). The reaction mix was gently mixed and briefly centrifuged. The reaction mix was incubated at room temperature for 10 minutes. T4 DNA ligase enzyme was inactivated by incubation at 65°C for 10 minutes. The ligated construct was then transformed into competent cells.

Component	Volume Required for 20µL Reaction
T4 DNA Ligase Buffer	2µL
Vector DNA	0.020pmol
Insert DNA	0.060pmol
T4 DNA Ligase	1µL
Nuclease-free water	To 20µL

Table 2.12: Components required for 20µL reaction volume ligation. Vector and insert volumes were calculated according to molecular weight and concentration.

2.2.10 LB agar plates

LB Ampicillin X-gal/IPTG plates were used for bacterial cultivation and blue/white screening. 40g/L LB Agar, Miller (Fisher) medium in dH₂O was mixed by swirling, and autoclaved to dissolve powder and sterilise. Liquid LB agar was allowed to cool until it could be held. 100µg/mL Ampicillin (from 100mg/mL stock), 1µL/mL X-Gal solution (20mg/mL stock) and 1µL/mL 100mM IPTG was added to the liquid LB agar then mixed well by swirling, and poured into Petri dishes. LB agar plates were allowed to set and stored at 4°C for up to 1 month.

2.2.11 DH5α competent cell transformation

LB plates were pre-warmed in a 37°C incubator. 50µL Mix & Go Cells (Zymo Research) in a 1.5mL tube were thawed on ice for 2 minutes. 1-5µL plasmid DNA/ligation product was added to the thawed cells and mixed by gently tapping the tube. The cells were incubated on ice for up to 10 minutes. 25-50µL

of the transformed cells were transferred to a pre-warmed LB ampicillin plate and spread with a sterile glass spreader. The plate was incubated overnight at 37°C.

Colonies were selected from the plate and transferred to LB ampicillin medium in a 50mL Falcon tube. LB cultures were incubated overnight at 37°C in a shaking incubator.

2.2.12 Plasmid extraction and purification

Plasmid extractions from bacterial cultures grown overnight in 1-5mL LB ampicillin medium were performed using the QIAprep Spin Miniprep Kit (Qiagen). Plasmid extractions from 100-250mL LB cultures were performed using the EndoFree Plasmid Maxi Kit (Qiagen), according to the manufacturer's protocols.

2.2.12.1 Miniprep of 1-5mL culture

Bacteria grown overnight in a 1-5mL culture were harvested by centrifuging at 6500 x *g* for 3 minutes. The supernatant was removed and the bacterial pellet resuspended in 250µL Buffer P1 and transferred to a microcentrifuge tube. 250µL Buffer P2 was added and the sample mixed by inverting the tube 4-6 times. Lysis was allowed to proceed for up to 5 minutes. 350µL Buffer N3 was added and the sample mixed by inverting the tube 4-6 times.

The sample was centrifuged at 15000 x *g* for 10 minutes. 800µL supernatant was removed to the QIAprep 2.0 spin column with 2mL collection tube and the sample centrifuged for 1 minute. Flow through was discarded and the collection tube replaced. 500µL Buffer PB was added to the column and the sample centrifuged for 1 minute. Flow through was discarded and the collection tube replaced. The sample was washed by applying 750µL Buffer PE to the QIAprep 2.0 column and centrifuged for 1 minute. Flow through was discarded and the collection tube replaced. The sample was centrifuged for a further minute to remove residual wash buffer.

The QIAprep 2.0 column was placed in a clean 1.5mL Eppendorf tube. 50µL sterile water, preheated to 50°C, was added to the centre of the column membrane and centrifuged for 1 minute to elute the purified DNA.

2.2.12.2 Maxiprep of 100-250mL culture

Bacteria grown overnight in a 100-250mL culture were harvested by centrifuging at 6000 x *g* for 15 minutes, at 4°C. The supernatant was removed and the bacterial pellet resuspended in 10mL Buffer P1. 10mL Buffer P2 was added and the sample mixed by inverting 4-6 times. Lysis was allowed to proceed for up to 5 minutes.

10mL Buffer P3 was added and the sample mixed by inverting 4-6 times. The lysate was poured into a QIAfilter Cartridge and incubated at room temperature for 10 minutes. The plunger was inserted into the QIAfilter Cartridge and the lysate filtered into a 50mL tube. 2.5mL Buffer ER was added to the filtered lysate, mixed by inverting 10 times and incubated on ice for 30 minutes.

A Qiagen-tip 500 was equilibrated by applying 10mL Buffer QBT and allowing the buffer to filter through by gravity flow. The filtered lysate was applied to the equilibrated Qiagen-tip 500 and allowed to filter through by gravity flow. The Qiagen-tip was washed with 30mL Buffer QC, twice.

15mL Buffer QN was added to elute the DNA into a clean endotoxin-free tube. 10.5mL isopropanol was added and the sample mixed to precipitate the DNA. The sample was centrifuged at 15000 x *g* for 30 minutes, at 4°C. The supernatant was discarded. The DNA pellet was washed with 5mL 70% ethanol and centrifuged at 10000 x *g* for 10 minutes. The supernatant was discarded.

The pellet was dried for 10 minutes and the DNA resuspended in 1mL TE buffer or dH₂O.

2.2.13 Glycerol stocks for long-term storage of transformants

Glycerol stocks were maintained for long term storage of constructs transformed into DH5α. Glycerol stocks were generated from bacterial cultures grown overnight in LB ampicillin medium. 0.5mL LB culture was added to a 1.5mL Eppendorf on ice. 0.5mL glycerol was added to the LB culture and mixed by slowly pipetting 2-3 times. Samples were immediately frozen in liquid nitrogen and stored at -80°C.

2.3 CRISPR Cas9 Gene Editing

2.3.1 Development of a *D. pseudoobscura cas9* integrated line

Prior to this project, *D. pseudoobscura cas9* lines were not available. Constructs were designed for insertion of the *cas9* gene into the *D. pseudoobscura* genome under a germline-specific promotor. Two constructs were designed: the first for assembly of the Cas9 cassette, the second for transformation of Cas9 and a cyan eye-marker into embryos. The second construct was injected into embryos to generate *D. pseudoobscura* lines expressing the Cas9 enzyme (see Chapter 3). The AmCyan eye-marker was used to identify and select transgenic *D. pseudoobscura* in F1 progeny of injected survivors.

2.3.2 Embryo DNA staining to establish timing of developmental stages

DNA staining was carried out on embryos aged 0-4 hours to identify the period of time in which blastoderm formation occurs, and therefore predict the time after egg deposition (AED) within which injection must occur. Cranberry juice agar plates containing embryos were collected from a laying pot

at hourly intervals and incubated at 25°C for 0-4 hours. Subsequent steps were performed separately for each plate.

Agar plates were rinsed with dH₂O to remove embryos onto mesh. Embryos were washed in 50% bleach for 4 minutes. The bleach solution was removed from the embryos through the mesh. Embryos were washed with dH₂O for five minutes. Embryos were removed from the mesh into a 10mL glass vial. Embryos were fixed in 5mL heptane and 2mL 4% paraformaldehyde (PFA) in PBS for 2 hours. The PFA fix (bottom phase) was removed. 2mL methanol was added to the vial and shaken for 2 minutes to remove the vitelline membrane. De-vitellinised embryos were removed to a 1.5mL Eppendorf and washed with methanol.

Embryos were rehydrated by washing three times in PBS + 0.1% Tween (PBST). The PBST was removed and replaced with a staining solution of 1µg/mL Hoechst 33258 in PBST. Embryos were stained for 15 minutes. The staining solution was removed and the embryos washed in PBS for 10 minutes.

The PBS was removed and mounting media (85% glycerol, 2.5% propyl gallate) was added. Embryos in mounting media were transferred to glass slides and covered with a siliconised coverslip. Slides were imaged on an Olympus BX50 fluorescence microscope with Hamamatsu ORCA-05G camera attachment and HCLImage software (v.2,2,6,4 Hamamatsu Corp. 2011).

2.3.3 Generating a pBluescript construct for *cas9* cassette assembly

A *cas9* cassette was designed for assembly of the germline specific *nanos (nos)* promotor and UTR sequences, and *cas9* gene sequence, into a pBluescript KS+ backbone (Ren *et al.* 2013). pBluescript (KS+) is a high copy number plasmid with a small backbone, and therefore ideal for assembly of components and amplification, prior to assembly of the *nos-cas9* cassette into the larger piggyBac construct to be used for insertion into the *D. pseudoobscura* germline. pBluescript (KS+) and *cas9* sequences were downloaded from Addgene (Addgene plasmid # 49330). The *D. pseudoobscura nos* 5' and 3' UTR sequences were downloaded from FlyBase (Thurmond *et al.* 2019, *D. pseudoobscura* release 3.04, May 2018). Assembly of component sequences was modelled in SnapGene (4.1.9 and 3.0.3, GSL Biotech) (Figure 2.1).

pBluescript (KS+) (gift from Sonia Lopez de Quinto) was digested with SacII and NotI (Table 2.10). Digested plasmid was run on a 1% agarose gel. The 2952bp band was excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

D. pseudoobscura nos promotor + 5' UTR and *nos* 3' UTR sequences were synthesised based on FlyBase reference sequences by Integrated DNA Technologies (IDT). *nos* 3' UTR was synthesised in two parts and joined by PCR amplification (see Table 2.13 for *nos* UTR primer sequences).

cas9 with nucleoplasmin NLS and 3xFLAG was obtained from the pAc-sgRNA-Cas9 construct (gift from Ji-Long Liu, Addgene plasmid #49330). The *cas9* sequence with flanking nuclear localisation signal and 3xFLAG tag was obtained from purified pAc-sgRNA-Cas9 plasmid by restriction digest using EcoRI and HindIII enzymes. The restriction digest reaction mix was run on a 1% agarose gel, and the 4284bp band corresponding to the *cas9* sequence excised and purified by gel extraction with the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

nos promotor + 5' UTR and 3' UTR synthesised DNA was amplified by PCR with primers including restriction sites (Table 2.13) and purified by gel extraction (QIAquick Gel Extraction Kit, Qiagen). Purified products were ligated into pGEM[®]-T Easy (Promega) according to the manufacturer's protocol. Colonies were grown at 37°C overnight on LB agar (100µg/mL ampicillin, 40µg/mL Xgal, 0.1mM IPTG). White colonies were selected and inoculated in LB with 100µg/mL ampicillin. *nos* promotor + 5' UTR and *nos* 3' UTR plasmids were extracted with the QIAprep Spin Miniprep Kit (Qiagen). pGEM[®]-T Easy-*nos* promotor + 5' UTR and pGEM[®]-T Easy-*nos* 3' UTR glycerol stocks were generated for long-term storage.

pGEM[®]-T Easy-*nos* promotor + 5' UTR was digested with SacII and EcoRI (NEB). pGEM[®]-T Easy-*nos* 3' UTR was digested with HindIII and KpnI (NEB) (Table 2.10). Restriction digest products were separated on a 1% agarose gel, and the 692bp (5' UTR) and 1029bp (3' UTR) bands purified by gel extraction (QIAquick Gel Extraction Ki, Qiagen).

nos promotor + 5' UTR, *cas9* and *nos* 3' UTR were ligated in pBluescript (KS+) by T4 DNA ligase (NEB), according to the manufacturer's protocol (Table 2.14). 3µL of pBluescript-*nos*-Cas9 ligation mixture was used to transform 50µL Mix & Go Competent Cells DH5α (Zymo Research). Transformed cells were incubated at 37°C overnight on LB amp/X-gal/IPTG agar. White colonies were selected and incubated overnight at 37°C in LB amp broth. pBluescript-*nos*-Cas9 plasmid was purified with the QIAprep Spin Miniprep Kit (Qiagen) and constructs sequenced to confirm correct assembly (Eurofins).

Product	Forward Primer	Reverse Primer
<i>nos</i> Promotor + 5' UTR + (3')	TCAGTGATGAAGGCCGGCAAT	gaattcATGACAAAGTCCAAGTTT
EcoRI	ATCG	TTACCGCAATTTGCGC
<i>nos</i> 3' UTR + (5') HindIII	aagctttAAAAAGTCTAGCCGAAG	TGAAGTTGAAGTAAACATATGT
	AACCCACACGG	A

Table 2.13: Primers for amplification of *nos* promotor + 5' UTR and 3' UTR fragments from synthesised DNA extended to include restriction sites EcoRI on the 3' end of *nos* 5' UTR and HindIII on the 5' end of *nos* 3' UTR, enabling ligation to *cas9* digested with EcoRI and HindIII.

Component	Amount Required for 20µL Reaction
10X T4 DNA Ligase Buffer	2µL
T4 DNA Ligase	1µL
pBluescript (KS+)	50ng
<i>nos</i> Promotor + 5' UTR	11.83ng
<i>cas9</i>	69.32ng
<i>nos</i> 3' UTR	16.90ng
dH ₂ O	To 20µL

Table 2.14: Ligation reaction with T4 DNA Ligase combining *nos* 5' UTR, *cas9* and *nos* 3' UTR into pBluescript (KS+).

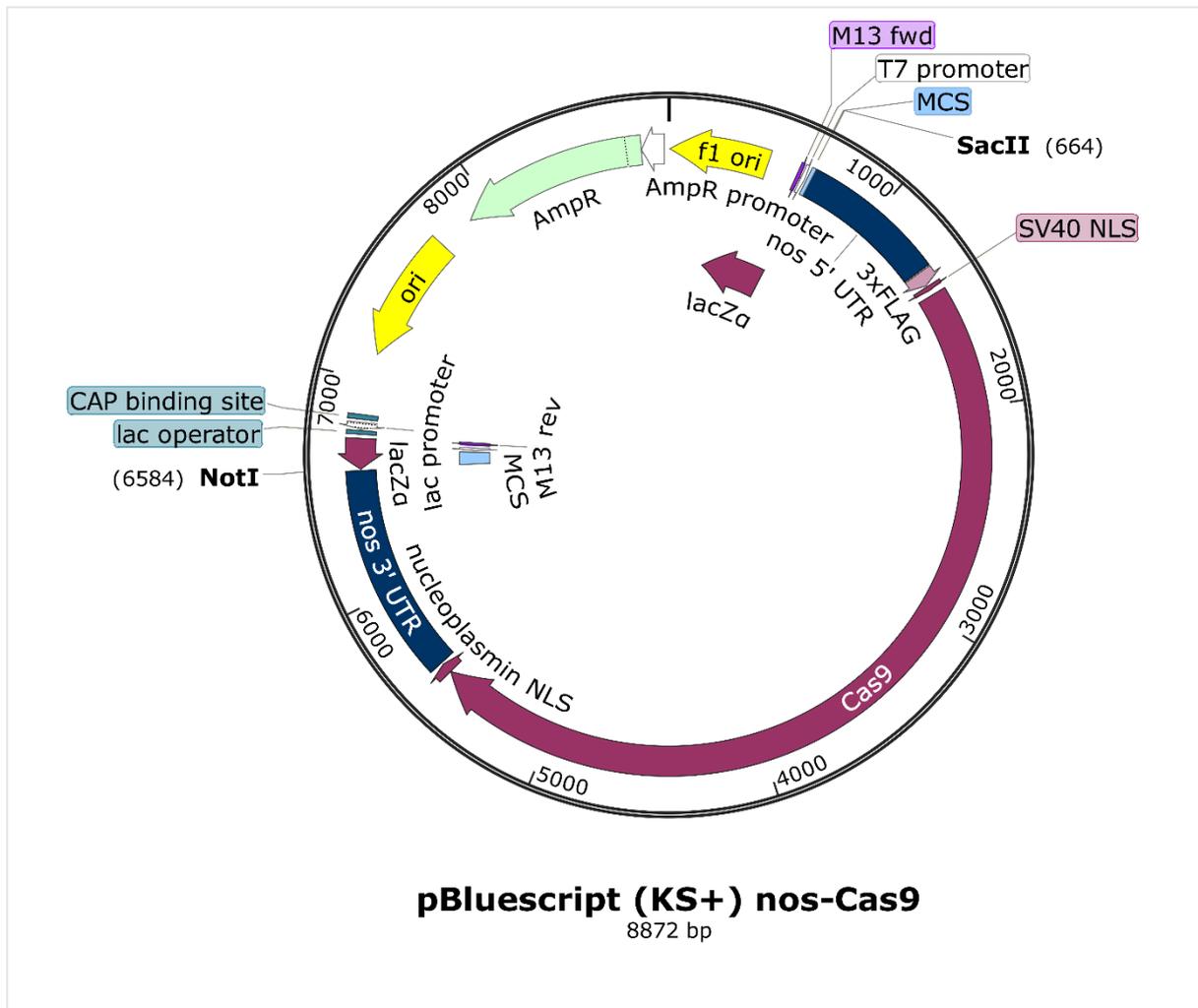


Figure 2.1: pBluescript (KS+) construct design for the assembly of a *nos*-Cas9 cassette into pBluescript prior to ligation into piggyBac. Construct map generated in SnapGene (3.0.3).

2.3.4 piggyBac-3xP3-AmCyan-nos-Cas9 construct assembly

piggyBac-3xP3-AmCyan-sv40-FRTlox-AeHex-ttAV2-K10 (gift from David Navarro) was digested with SacII and NotI to release the piggyBac-3xP3-AmCyan backbone (Figure 2.2; A). The 6437bp backbone fragment was excised from a 1% agarose gel and DNA purified by gel extraction (QIAquick Gel Extraction Kit, Qiagen).

The *nos promotor-5' UTR-cas9-nos 3' UTR* cassette was removed from pBluescript by restriction digest with SacII and NotI (Table 2.10) and the 5931bp fragment purified by gel extraction (QIAquick Gel Extraction Kit, Qiagen).

piggyBac-3xP3-AmCyan and *nos-cas9* fragments were ligated with T4 DNA ligase (NEB) according to the manufacturer's protocol (Figure 2.2). 3µL ligation mixture was used to transform 50µL Mix & Go DH5α competent cells. Transformed cells were grown overnight at 37°C on LB amp/X-gal/IPTG agar plates, then selected and grown in LB ampicillin broth overnight at 37°C. Glycerol stocks were generated from each culture. piggyBac-3xP3-AmCyan-nos-Cas9 (Figure 2.2; B) was purified with the Qiagen Plasmid Midi Kit (Qiagen). To verify the constructs contained the *nos-cas9* fragment, a sample of the plasmid was digested with SacII and NotI (Table 2.10), and run on a 1% agarose gel. Presence of bands at 6437bp and 5931bp confirmed piggyBac-3xP3-AmCyan backbone and *nos-cas9* fragments respectively. Constructs were sequenced to confirm presence of the insert.

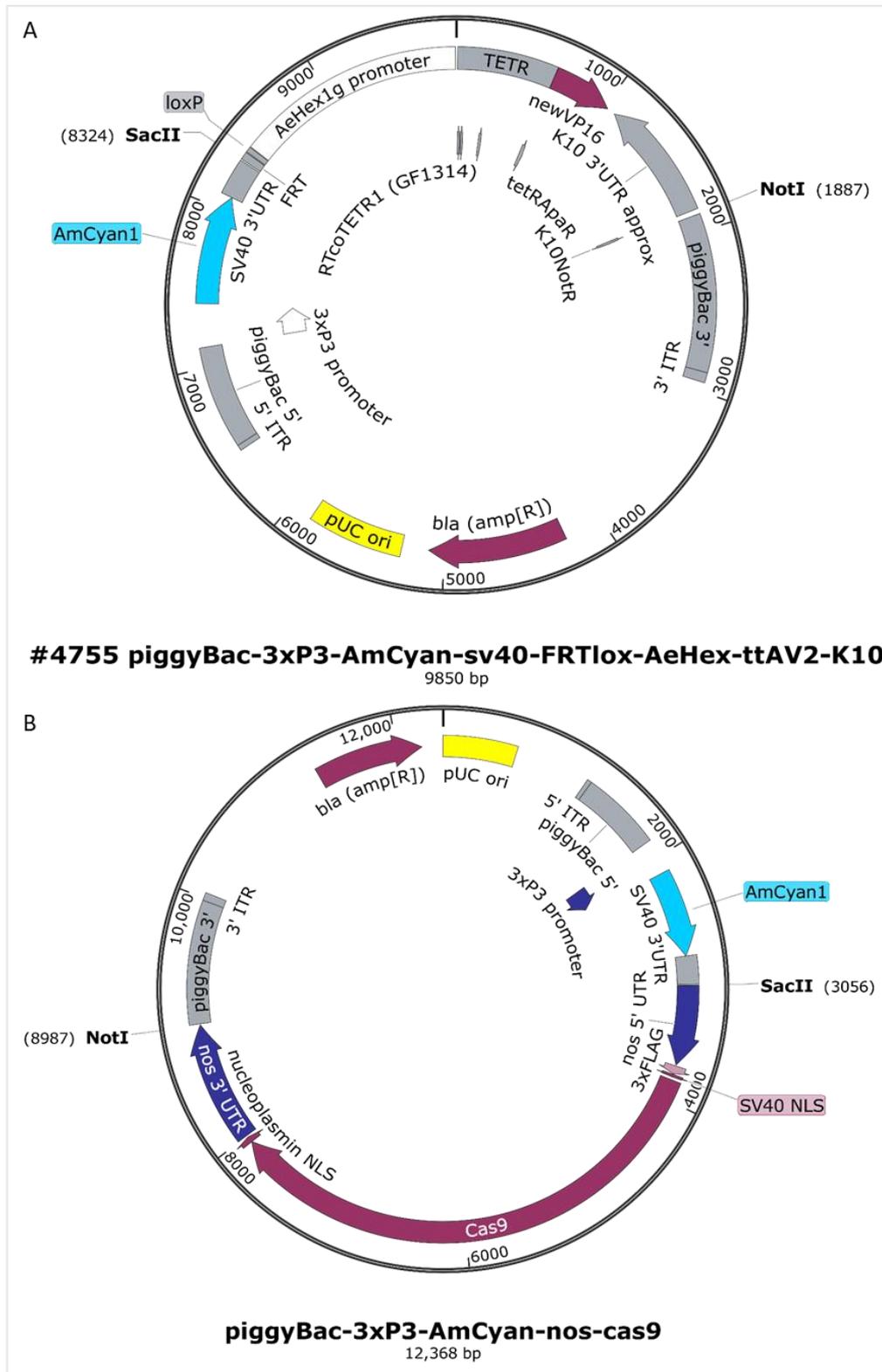


Figure 2.2: (A) *piggyBac-3xP3-AmCyan-sv40-FRTlox-AeHex-ttAV2-K10* plasmid (from David Navarro) used to retrieve the *piggyBac-3xP3-AmCyan* backbone for ligation with *nos-cas9*. (B) *piggyBac-3xP3-AmCyan-nos-Cas9* construct design for the integration of *cas9* into the *D. pseudoobscura* genome under the *nos* germline promoter with 3xP3-AmCyan fluorescent eye marker. The *piggyBac-3xP3-AmCyan* backbone and *nos-cas9* cassette were combined by ligation at the *SacII* and *NotI* restriction sites.

2.3.5 piggyBac-3xP3-AmCyan-nos-Cas9 Plasmid Injection in *D. pseudoobscura*

Injection solution was prepared from purified 500ng/ μ L piggyBac-3xP3-AmCyan-nos-Cas9 and 133ng/ μ L piggyBac Helper transposase expression plasmid in dH₂O (Handler *et al.* 1998).

Adult flies >7 days old were collected and transferred to a laying pot. An agar plate with yeast paste was attached to the bottom of the laying pot. Laying pots were maintained at 25°C.

Embryos were collected from laying pots at hourly intervals. Embryos were removed from the plate with a brush and placed on a microscope slide with a strip of double-sided Scotch tape, between two pieces of copper wire. A second slide with a strip of tape was placed on top and then removed, opening the chorion. The de-chorionated embryos were transferred with forceps to a line of glue (Scotch tape glue dissolved in heptane) on a glass coverslip. Embryos were allowed to dehydrate for approximately 20 minutes, then covered with halocarbon oil 700 (Sigma-Aldrich).

Injections were carried out under a Nikon Eclipse Ti-S with the Eppendorf FemtoJet micro-injection system. Pre-blastoderm embryos were injected at the posterior with the piggyBac-nos-Cas9-3xP3-AmCyan/piggyBac Helper injection solution. Coverslips of injected embryos were removed to agar plates with yeast and incubated at 21°C in high humidity for 36 hours. Larvae were collected from the plates and removed to a food vial. Injected larvae were incubated at 25°C until eclosion.

2.3.6 Transgenic adult collection and crosses

All crosses were carried out at 25°C.

Injected adult flies (hereafter referred to as F0) were collected as virgins and crossed with WT *D. pseudoobscura*. Male F0 flies were crossed with four WT females. Female F0 flies were crossed with two WT males. Transgenic offspring (F1) were identified by the presence of cyan fluorescent eyes (3xP3-AmCyan eye marker). Non-transgenic offspring were discarded. Transgenic F1 were used to set up individual crosses with WT to obtain a population of transgenic heterozygotes (F2). Transgenic F2 (homozygote and heterozygote) were collected as virgins and crossed with transgenic F2 from the same line (1 male, 1 female per cross). Transgenic F3 were collected and crossed. There were four potential combinations in F3 crosses: heterozygote male x heterozygote female, heterozygote male x homozygote female, homozygote male x heterozygote female, homozygote male x homozygote female. Any populations with homozygote non-transgenic offspring (i.e. no eye marker) were discarded. Over subsequent generations, populations producing homozygote offspring only were maintained.

2.3.7 Inverse PCR and sequencing to determine piggyBac-nos-Cas9 insertion sites

To identify the point of piggyBac-nos-Cas9 insertion in each transgenic line, an inverse PCR protocol was used (Labbe *et al.* 2010). This method relied on extraction of genomic DNA, genomic DNA digestion with three restriction enzymes and circularisation of the digested fragments. The genomic DNA was digested such that fragments of the transgene insert and genomic DNA were combined to form a plasmid. PCR was then used to amplify the ‘unknown’ fragment of DNA by use of primers within the ‘known’ section of the digested construct.

From each transgenic line, a single virgin F2 individual was collected. Genomic DNA was extracted using the single fly DNA extraction method (see 2.2.3.2). The DNA concentration of each sample was determined by NanoDrop spectrophotometry. 2.5µg genomic DNA was digested first with Taqα1 (NEB). Digested DNA was purified with the QIAquick PCR Purification kit (Qiagen). Taqα1 digested DNA was then digested a second time with MspI and HaeIII (NEB). The digested DNA was run on an agarose gel and purified by gel extraction using the QIAquick® Gel Extraction Kit. 15ng of purified DNA was re-circularised by T4 DNA Ligase, according to the manufacturer’s protocol.

The re-circularised DNA was amplified by PCR. Details of primers for inverse PCR can be found in Table 2.15. PCR products were purified with the QIAquick PCR Purification kit (Qiagen).

Purified samples were sequenced by Eurofins using the primers in Table 2.15.

Inverse PCR Primer	Forward Primer	Reverse Primer
piggyBac 5’ insertion	TGACACTTACCGCATTGACA	TCTTGACCTTGCCACAGAGG
piggyBac 3’ insertion	GTCAGTCAGAAACAACCTTGGC	CCTCGATATACAGACCGATAAAACACATG

Table 2.15: Primers for inverse PCR and product sequencing to identify the location of piggyBac insertion sites. Primer sequences were retrieved from (Labbe *et al.* 2010).

2.3.8 cas9 expression in *D. pseudoobscura* ovaries

To demonstrate the expression of *cas9* in transgenic lines, reverse-transcription PCR was performed on ovaries of *D. pseudoobscura* transgenic for the piggyBac-nos-Cas9 gene insert. Primers for the amplification of *cas9* were synthesised by IDT.

Ovaries of 10 female *D. pseudoobscura* from each of 27Axii, 29Ciii, 29Dviii, 29Fiii and 49Bi transgenic lines were dissected into testis buffer. Ovaries were pooled into the lid of a 1.5mL Eppendorf tube. RNA was extracted from ovaries using Trizol, according to the manufacturer’s protocol (see 2.2.2.1). cDNA was synthesised using the SuperScript III First Strand Synthesis System (Invitrogen) (see 2.2.2.2).

PCR was performed on the ovary cDNA with two sets of primers (Table 2.16). PCR product was run on an agarose gel and imaged.

Primer Pair	Forward	Reverse	Expected Product Length (bp)
Cas9 1	athtaggtgacactatagaaCAGCTGGTGCA GACCTACAA	taacctcactaaaggTCATCCGCTC GATGAAGCTC	930
Cas9 2	athtaggtgacactatagaaCAAGCTGATCC GGGAAGTG	taacctcactaaaggCTGTCTGCAC CTCGGTCTTT	468

Table 2.16: Primers for ovary cDNA PCR of cas9 expressed in transgenic *D. pseudoobscura* lines. cas9 sequence is shown in uppercase, T3 and SP6 partial sequences are lowercase.

2.3.9 Testing *D. pseudoobscura* integrated Cas9 by targeting marker genes *white* and *yellow*

Cas9 function in *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 lines was tested by targeting two marker genes, *white* (GA27183) and *yellow* (GA15665) with guide RNA.

Guide RNA sequences were identified using CRISPOR (Concordet and Haeussler 2018). A further three gRNA sequences were provided by courtesy of Nitin Phadnis (Dean Castillo, pers. comm. 2021).

gRNA oligos were designed for ligation into the pCFD3-U6:3-gRNA expression vector, for gRNA expression under the U6 promotor. Sense and anti-sense oligos were synthesised with extensions for ligation at the BbsI site in the pCFD3 vector (Port *et al.* 2014). Sense oligos were extended with the 5'-GTCG-3' sequence. Anti-sense oligos were extended with the 5'-AAAC-3' sequence.

Gene	gRNA	gRNA Oligo Sequence Sense	gRNA Oligo Sequence Anti-Sense
White	PLR013*	gtcgCAGGAGTTGTTGATACGCGG	aaacCCGCGTATCAACAACCTCTG
	PLR014*	gtcgTTGTTTAAGTGTTCCGCGAG	aaacCTGCCGAACACTTAAACAAC
	PLR015*	gtcgCCAAGAACTACGGCACCTTA	aaacTAGGGTGCCGTAGTTCTTGG
	White gRNA2	gtcgTGTCGACTAATTACAATC	aaacGATTGTAATTAGTCGGACA
Yellow	Yellow gRNA1	gtcgAAGTCGCCGCGCAGGGGGTTC	aaacGACCCCTGCGCGGCGACTT
	Yellow gRNA2	gtcgTACGTAAATGCCCGACAGTC	aaacGACTGTCTGGGCATTTACGTA

Table 2.17: Guide RNA oligo sequences for targeting *white* and *yellow* marker genes in *D. pseudoobscura*. * PLR013-15 gRNA designs courtesy of Nitin Phadnis (Dean Castillo, pers. comm).

2.3.9.1 gRNA expression construct assembly

The annealing reaction was set up on ice to anneal the gRNA oligos prior to ligation into the pCFD3 expression construct (Table 2.18). The annealing reactions were transferred to a thermocycler and the annealing reaction program initiated (Table 2.19).

pCFD3-dU6:3-gRNA was digested with BbsI and purified with the QIAquick PCR Purification kit (Qiagen). The pCFD3 ligation mix was set up on ice (Table 2.20) (Port *et al.* 2014). Ligations were incubated at room temperature for 30 minutes.

The pCFD3-U6:3-gRNA ligated constructs were transformed into Mix & Go DH5 α competent cells (Zymo Research) and grown at 37°C overnight on pre-warmed LB ampicillin plates. Colonies were inoculated into 5mL LB ampicillin medium and grown at 37°C overnight. Plasmid DNA was purified by miniprep (Qiagen). Correct ligation of gRNA oligos in pCFD3-U6:3-gRNA was confirmed by sequencing (Eurofins). Once sequences were confirmed, gRNA constructs were re-transformed into Mix & Go DH5 α (Zymo Research) and grown overnight. Colonies were inoculated into 100mL LB ampicillin and grown overnight at 37°C. Plasmid DNA was purified by maxiprep (Qiagen).

Component	Volume (μ L)
Sense Oligo (100μM)	1
Anti-Sense Oligo (100μM)	1
10X T4 Ligation Buffer (NEB)	1
dH₂O	6.5
T4 PNK (NEB)	0.5

Table 2.18: Reagents required for annealing reaction mix, to anneal gRNA sense and anti-sense oligos prior to ligation into pCFD3 expression vectors (Port *et al.* 2014).

Step	Temperature (°C)	Time
1	37	30min
2	95	5min
3	Ramp down to 25°C	5°C/min

Table 2.19: gRNA oligo annealing thermocycler program.

Component	Volume (μ L)
BbsI digested pCFD3	X (50ng)
Annealed Oligos Diluted 1:200	1
10X T4 Ligation Buffer (NEB)	1.5
dH₂O	To total volume 15 μ L

T4 DNA Ligase (NEB)	1
----------------------------	---

Table 2.20: Reagents required for ligation of annealed gRNA oligos into pCFD3-U6:3 expression vector.

2.3.9.2 pCFD3-U6:3-gRNA expression vector injection into *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9

>100 adult *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 from each of five lines, 0-3 days post-eclosion, were transferred to laying pots and maintained at 25°C for three days prior to embryo collection for injection. Individual laying pots were set up for each line tested. Embryos were collected and injected as previously described (see 2.3.5). A total of 5 lines were tested (Table 2.21). Injected embryos were maintained in a humid chamber at 21°C for 36 hours. Surviving larvae were collected and transferred to food vials. Larvae were maintained at 21°C until eclosion.

Surviving F0 adults were collected within 24 hours of eclosion. Adult F0 flies were crossed with *D. pseudoobscura* SLOB3 wild-type. Male F0 survivors were crossed with 5-6 virgin females 0-3 days post-eclosion. Female F0 survivors were crossed with 2-3 males 0-3 days post-eclosion. Crosses were maintained at 21°C. F0 adults were collected prior to eclosion of F1 and stored at -20°C prior to DNA extraction.

white and *yellow* are located on the X chromosome. Therefore, F0 males can only produce heterozygous transgenic F1 daughters, and WT F1 sons. F0 females can produce transgenic F1 sons with a single transgenic X, heterozygous transgenic F1 daughters, WT F1 sons and homozygous WT F1 daughters (Table 2.22). X'Y males may have a visible *white* or *yellow* mutant phenotype.

F1 adults were collected within 24 hours of eclosion. F1 sons of F0 females were screened for *white* or *yellow* mutant phenotypes. F1 sons of F0 males were discarded. F1 daughters of both F0 females and F0 males were crossed with *D. pseudoobscura* SLOB3 wild-type. Crosses were maintained at 21°C.

F2 adults were collected within 24 hours of eclosion. F2 males were screened for *white* or *yellow* mutant phenotypes.

<i>D. pseudoobscura</i> 3xP3-AmCyan-nos-Cas9 Line	
ID	gRNAs Tested
27Axii	W2 Y1 PLR013, PLR014, PLR015
29Ciii	W2
29Dviii	Y1

	Y2 PLR013, PLR014, PLR015
29Fiii	W2 Y2 PLR013, PLR014, PLR015
49Bi	PLR013, PLR014, PLR015

Table 2.21: *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 lines tested by injection of gRNA expression constructs targeting white and yellow markers. White 2 and Yellow 2 gRNA constructs injected as a single construct. PLR013, 14 and 15 injected as mix.

F0 Female x WT		F0 Male x WT Female			
Male	X'	X	X'	Y	
X	X'X	XX	X	X'X	XY
Y	X'Y	XY	X	X'X	XY

Table 2.22: F1 offspring genotypes from F0 transgenic crosses for white or yellow markers. X'X genotype flies are not expected to have visible phenotype. X'Y may have visible phenotype where CRISPR mutagenesis has resulted in non-homologous end-joining and disruption to marker gene transcription or translation.

2.3.9.3 PCR and high resolution melt analysis (HRMA)

To assess whether mutagenesis had taken place, F0 adults were collected and DNA extracted by the single fly DNA extraction method. Primers were designed to amplify the region surrounding the gRNA target sites. A second set of primers was designed to amplify a smaller region of 100-500bp surrounding the target site for high resolution melt analysis to identify potential mutations (Bassett *et al.* 2013; Bassett and Liu 2014; Housden *et al.* 2014; Housden and Perrimon 2016).

The gene region surrounding the gRNA target sites was amplified by PCR using the Q5 High-Fidelity Master Mix (NEB) (see 2.2.4.3). PCR product was used as DNA template for high resolution melt analysis. LightCycler 480 High Resolution Melting Master (HRM) (Roche) with the LightCycler 96 Instrument (Roche) was used to perform HRMA, according to the manufacturer's protocol.

The HRM master mix was prepared on ice (Table 2.24). 15µL HRM master mix was transferred to each well of a LightCycler 480 96-well plate (Roche). 5µL adjusted concentration DNA template was added to each well, and pipetted 3 times to mix. A *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 control was included in each plate for the subsequent analysis. A negative control was also included in each plate. Plates were sealed with LightCycler 480 Sealing Foil (Roche).

Plates were centrifuged for 2 minutes at 1500g, then transferred to the LightCycler 96 Instrument. The HRMA program was initiated (Table 2.25).

Once the HRMA program was complete, the raw data was transferred from the LightCycler 96 for melt curve analysis in the LightCycler 96 Application Software (v.1.1.0.1320. Roche Diagnostics International Ltd. 2011).

Gene	Primers	Forward Primer	Reverse Primer
White	White Exon 4	TCAGCCTGTTCCAAGCAGTT	TAGCGGTTTCGATACCCCTCAT
	White Exon 4 HRMA	AAATGAGCGTGCTCCCATCC	CGCCTCTCTTCTTACCGTT
	White PLR013-14 HRMA	TCCCGTAGCCACATTTTCCG	GCCCAAGACAAGAAAAACCCC
	White PLR015 HRMA	ACTCATCCCTCCTCGTGCT	TCCGATTGACCAACTGTGCGC
Yellow	Yellow Exon 2	GATGACGGCGTGGAGCTATT	CGAGTATTTGTGGTGTGGCCT
	Yellow Exon 2 HRMA	CATCTCTTCCCCACCCTCCA	GAGTCACCCTACACCCAGTTG

Table 2.23: Primer sequences for PCR and melt analysis of white and yellow gRNA target sites.

High Resolution Melting Master Mix	
Component	Volume (μ L)
Master Mix 2X	10
Primer Mix (forward and reverse primer)	1
MgCl₂	1
dH₂O	3

Table 2.24: Reagents required for high resolution melting master mix (Roche) for HRMA of white and yellow gRNA target sites.

Step	Temperature ($^{\circ}$ C)	Time (s)	Ramp ($^{\circ}$ C/s)	Acquisition Mode
Pre-Incubation	95	600	4.4	-
3-Step Amplification;	95	10	4.4	
	55	15	2.2	
	72	20	4.4	
45 Cycles	95	60	4.4	
	40	60	2.2	
	65	1	2.2	
	97	1	-	

Table 2.25: LightCycler 96 program for high resolution melt analysis of white and yellow gRNA target sites.

2.4 *D. pseudoobscura* Spermatocyte Cyst RNA Sequencing

2.4.1 Spermatocyte cyst dissection

Testis dissections were carried out as described in 2.2.1. The testes were dissected from a WT *D. pseudoobscura* into a drop of testis buffer. One testis was removed to a second drop of testis buffer on a siliconised slide and sliced open using a tungsten needle. The contents were emptied into the testis buffer and the testis sheath discarded. A further drop of testis buffer was added and the contents mixed by pipetting.

The slide was transferred to a Nikon Eclipse Ti-S microscope with digital camera attachment (Hamamatsu ORCA-05G). A spermatocyte cyst was identified visually and removed by capillary action using a glass capillary tube, into a 0.5mL microcentrifuge tube. This was repeated a total of ten times to collect a total of ten spermatocyte cysts from a single testis. Samples were flash frozen in liquid nitrogen then removed and stored at -80°C until library preparation. Images of cysts were taken with HClmage software (v.2,2,6,4 Hamamatsu Corp. 2011).

This was repeated a total of three times to give a total of 30 spermatocyte cysts from three males.

Ten further males were dissected in the same manner. From each sample, one short spermatid cyst and one long spermatid cyst were removed by capillary action using a glass capillary tube into a 0.5mL microcentrifuge tube and flash frozen in liquid nitrogen. A total of 20 spermatid cysts (10 short, 10 long) were dissected from 10 males.

2.4.2 Library preparation

Library preparation took place in two phases. An initial batch of ten spermatocyte cysts from a single male were dissected and libraries prepared by Helen White-Cooper. In the second phase, twenty spermatocyte cysts and twenty spermatid cysts were dissected and libraries prepared by Fiona Messer. Library preparation, sequencing and analysis will be described separately.

2.4.3 RNA-Seq 1: Initial sequencing of ten spermatocyte cysts

The SMART-seq Ultra Low Input Kit (Clontech, Takara Bio USA) was used for synthesis, amplification and purification of cDNA from primary spermatocyte RNA according to the manufacturer's protocols (Takara Bio USA 2018). The cDNA product was stored at -20°C between amplification and library preparation. Library preparation was carried out according to the manufacturer's protocol with the Nextera XT DNA Library Preparation Kit (Illumina). Libraries were stored at -20°C prior to quality checks and sequencing.

The libraries were sequenced on the Illumina HiSeq System (Illumina) to a depth of approximately 20 million 75 base-pair paired-end reads per sample. Sample processing and sequencing were performed by the Glasgow Polyomics Service.

2.4.4 RNA-Seq 2: further spermatocyte cyst sequencing and spermatid cyst sequencing

RNA libraries were prepared from single spermatocyte and spermatid cysts with the QIAseq FX Single Cell RNA Library Kit (Qiagen). RNA was extracted from the cell samples and amplified by following the manufacturer's protocol for 'Amplification of Poly A+ mRNA from Single Cells' (Qiagen). The cDNA product was stored at -20°C between amplification and library preparation. Library preparation was carried out according to the manufacturer's protocol 'Enzymatic Fragmentation and Library Preparation Using QIAseq FX Single Cell Amplified cDNA' (Qiagen). Libraries were stored at -20°C prior to quality checks and sequencing.

Sample quality was assessed at two points; post-amplification and post-library preparation. The presence of cDNA was assessed post-amplification using a high-sensitivity Qubit assay (ThermoFisher Scientific) with a 1:100 dilution of the original sample, following the manufacturer's protocol. All spermatocyte and spermatid samples showed presence of DNA. Library quality was assessed by DNA TapeStation (Agilent) capillary electrophoresis.

The libraries were sequenced on the Illumina HiSeq System (Illumina) to a depth of approximately 13 million 75 base-pair paired end reads per sample.

2.4.5 Bioinformatic analysis

Sequence data was stored in the .fastq format. Paired data files were uploaded and the subsequent analysis pipeline carried out on the Galaxy web server (Afgan *et al.* 2016). Sequence quality of the files was assessed using the FastQC tool (Andrews 2010). The Trimmomatic tool (v0.36.3) (Bolger *et al.* 2014) was used to identify, filter and trim poor quality sequences. Sequences were filtered to have a minimum length of 36 base pairs, with average quality Phred score of 20 across 4 base pairs. Sequences were trimmed by three base pairs leading and trailing. Nextera adaptor sequences were trimmed.

Sequence alignment was performed by HISAT2 (v2.0.5.2) (Kim *et al.* 2015). Transcript sequences were aligned to the *D. pseudoobscura* reference genome retrieved from FlyBase (v3.04, October 2017) (Gramates *et al.* 2017). The aligned .bam files were filtered using SAMtools (v1.1.2) (Li *et al.* 2009) to remove unmapped and duplicate reads and sorted by chromosomal co-ordinate. The htseq-count tool

(v0.6.1) (Anders *et al.* 2015) was used to perform counts on aligned and annotated data. The annotation file was retrieved from FlyBase (v3.04) (Gramates *et al.* 2017).

2.4.6 Statistical analysis: clustering and differential gene expression analysis

All statistical analysis was carried out in R Studio (R Core Team 2017).

Hierarchical cluster analysis (HCA) was carried out using the pvclust package (Suzuki and Shimodaira 2015) to explore the data and identify clustering of the samples into groups according to gene expression. Euclidian distances between pairwise observations were calculated by the Ward method (bootstrap=1000). Raw count data was normalised by the trimmed mean of M-values (TMM) (Robinson *et al.* 2010), and HCA applied to the normalised counts. TMM normalisation was used as this is the method also applied to raw count data in edgeR (Robinson *et al.* 2010; McCarthy *et al.* 2012).

Differential gene expression (DGE) analysis requires two or more defined groups between which gene expression levels are analysed to determine which genes show significant differences in expression level. Samples were assigned to groups based on the clusters indicated by HCA. Counts per million (CPM) were calculated for each gene, low count sequences (<7) were filtered, and the data normalised by TMM. Differential gene expression (DGE) analysis was carried out using the edgeR package (Robinson *et al.* 2010; McCarthy *et al.* 2012).

2.4.7 Identification of target genes for further investigation

To identify differentially expressed genes with a potential role in heteromorphic sperm development, lists of genes with certain known functions, either in *D. melanogaster* or *D. pseudoobscura*, were compared to lists of differentially expressed genes according to the RNA-seq data. Gene function and orthology data for further analysis of function of differentially expression genes was retrieved from FlyBase using the 'Query' function (Gramates *et al.* 2017; Thurmond *et al.* 2019). Genes with roles in germline stem cell maintenance and development, spermatogenesis, spermiogenesis and transcription were chosen for further investigation.

2.5 *In Situ* Hybridisation of *D. pseudoobscura* Testes

In situ hybridisation staining was used to further investigate expression patterns of genes identified as differentially expressed between eusperm and parasperm spermatocyte cysts by RNA-seq. The method used is described by Morris *et al.* (2009).

2.5.1 RNA probe design for *in situ* hybridisation

RNA probes were designed for each of the genes in Table 2.26. Sequences were downloaded from Flybase (Thurmond *et al.* 2019). Primers were designed using the NCBI Primer Blast tool to be

separated by 400-1000bp and the PCR products – the probes – modelled in Snapgene (v3.0.3). Probes extending over multiple exons were preferentially selected where applicable. All forward primers included a partial SP6 promoter sequence at the 3' end and reverse primers included a partial T3 promoter sequence at the 5' end. Primers were synthesised by IDT.

Gene Number (Name)	Gene Function	Forward Primer Sequence	Reverse Primer Sequence	Estimated PCR Product Size
GA19370 (<i>piwi</i>)	Transcriptional and post-transcriptional silencing	atntaggtgacctatagaaTGCCGATTGGACA AGATGCT	taaccctcactaaagggCTAACGCCGTCACGATA GA	700
GA18272 (<i>Nup154</i>)	Nuclear pore complex	atntaggtgacctatagaaCAAAATGACTGCT ACGCCCCG	taaccctcactaaagggTCTCTCCAAATGTGACG CCC	518
GA18732 (<i>dad</i>)	Germline stem cell population maintenance	atntaggtgacctatagaaTGC GACTGAAAG ATGCAGATTC	taaccctcactaaagggCTTATCGGCAGCTCGGT AT	934
GA21874 (<i>chic</i>)	Male germline stem cell maintenance	atntaggtgacctatagaaGGACGTGGAACG TGAAACT	taaccctcactaaagggCTGGTCGAAGCCGCTA ATCA	451
GA18558 (<i>twe</i>)	Meiotic cell cycle regulation	atntaggtgacctatagaaAGCACGCTTTTCA CGCATTT	taaccctcactaaagggTGACGTCTTGTGGCTAA	639
GA19239 (<i>l(3)72Ab</i>)	U5 snRNP component, spliceosome assembly	atntaggtgacctatagaaTCACAGACTTTGG CAGCGAT	taaccctcactaaagggCAAAGGGTACGGCCTT GAGT	831
GA10082 (<i>wa-cup</i>)	Spermatid elongation	atntaggtgacctatagaaCAGGCCAAGCCC AGTATCAA	taaccctcactaaagggATGGGGCAGTCCAAAG AGTG	765
GA18412 (<i>bol</i>)	Translational regulator	atntaggtgacctatagaaTTGTGGATCGTGC TGGTGT	taaccctcactaaagggAATCTTCGACCCGCAGT TGT	552

GA20060 (Dic61B)	Axonemal dynein complex, axoneme assembly	atntaggtgacctatagaaCTGTTTCTCGCCG TTTCTGC	taaccctactaaagggTGCAGGATGTTGCGTCT GAT	729
GA27927 (tbrd-2)	Control of gene activity in germ cells, interacts with tTAFs	atntaggtgacctatagaaGAGATCCACGAT GTCGGTCC	taaccctactaaagggTCCCATGGCTGTAGACC TCA	431
GA18735 (kmg)	Repression of somatic transcription in germ cell lineages	atntaggtgacctatagaaGCGACTTCCACTC CCCTTAC	taaccctactaaagggCTCGAAGGGCTTCCGAT TCA	647
GA11638 (wuc)	tMAC component, transcription factor	atntaggtgacctatagaaAACGTGCGACAA AAAGCTACC	taaccctactaaagggATTTGCCGTCCCGTTCA CT	472
GA12326 (comr)	tMAC component, transcription factor	atntaggtgacctatagaaATACTACCAAGC AGGTGCC	taaccctactaaagggCTTGGGGCAGCTTGTTT TCC	531
GA12700 (tomb)	tMAC component, transcription factor	atntaggtgacctatagaaTCCACCATCAAGG GATGTGC	taaccctactaaagggCGAGGAGACGATAACT G	555
GA13507 (mip40A)	tMAC component, transcription factor	atntaggtgacctatagaaGAACCCATACCG GAGCTGAG	taaccctactaaagggTTCCTGTGGAAGTGG AGCC	407
GA18051 (Caf1A)	tMAC component, transcription factor	atntaggtgacctatagaaGAGGTGAACCGT GCTCGTTA	taaccctactaaagggCTCTGCTCCTCGCCTATT T	690

GA21345 Both (achi/vis)	tMAC component, transcription factor	atttaggtgacctatagaaCTCAGGGATACG AGTCGTGC	taaccctcactaaagggCTTGTCCGGGATCGTTG TGA	447
GA21345 RAC (achi/vis)	tMAC component, transcription factor	atttaggtgacctatagaaGTCTCCACTCGTC ACTGCAA	taaccctcactaaagggAGCAAAGCCCTCCTCTT GTC	806
GA21345 RB (achi/vis)	tMAC component, transcription factor	atttaggtgacctatagaaCGTTGTCTACAGA TCCGAGGG	taaccctcactaaagggAGCAAAGCCCTCCTCTT GTC	493
GA23669 (mip40B)	tMAC component, transcription factor	atttaggtgacctatagaaACAAGAAGTCCT GTCGCCTG	taaccctcactaaagggTCCAGAACTCGCAGGG TTTC	562
GA26389 (Caf1B)	tMAC component, transcription factor	atttaggtgacctatagaaCACACGACAAAC AACGACCC	taaccctcactaaagggCCGTTGCGTATATCCCA GA	573
GA28313 (aly)	tMAC component, transcription factor	atttaggtgacctatagaaGTCGTTCTGTGTC AGATGGA	taaccctcactaaagggAAGTGGACGAAAGAGG CCAG	415
GA14055 (dila)	Sperm axoneme assembly	atttaggtgacctatagaaGGTTCCCCAAAAG TCCACCA	taaccctcactaaagggAGGTCGTTGGGAAGAC GTTC	888
GA26457 (asl)	Axoneme assembly, centriole duplication	atttaggtgacctatagaaTGAAGAGTTTCAC TCCCCGC	taaccctcactaaagggACTCTGATATCGCTGGT GCG	434
GA27003 (mil)	Nucleosome assembly factor, cytoskeleton-based morphogenesis	atttaggtgacctatagaaTTTTTGTCCGGCAT GGACGC	taaccctcactaaagggTGTCCGGTTGACAGTTG AGG	514
GA25911 (Mst36Fb)	Unknown function	atttaggtgacctatagaaGCCCCGGGAGAT AAACAACA	taaccctcactaaagggATCCCTTTCCTCTGCTC G	791

GA12730 (U2A)	U2 snRNP component, spliceosome assembly	atntagtgacactatagaaCTCTGGACCAGTT CGACACC	taaccctactaaagggCGCTGCATATCCTCCGG ATT	514
GA25581 (Sf3b5)	Histone modification and U2 snRNP component	atntagtgacactatagaaTTTTTCCGAAATG GGCGAACG	taaccctactaaagggTTAGTCATCCAGCTTCT CGGG	258
GA21384 (Prp8)	Spliceosome assembly, U5 snRNP component	atntagtgacactatagaaCAGACGCGCATC AAGATTGG	taaccctactaaagggCATAGACATTGGCGCG GTTG	760
GA23025	Unknown function, possibly nucleosome assembly	atntagtgacactatagaaACGAGTGTTGTAA ACGGATCA	taaccctactaaagggCTATAGGGCAGAGTTG GCGG	735
GA14905 (Mondo)	Carbohydrate metabolism, sugar dependent transcription	atntagtgacactatagaaGTACCCCGTTCC ACAACAT	taaccctactaaagggTTGTCCTTCCAACGGAG TCG	702
GA12828 (Rcd-1)	CCR4-NOT complex, negative regulation of translation	atntagtgacactatagaaCCGTCTATAACTC CGCCCAC	taaccctactaaagggAGACGTAGGTAGCAAC GCA	495

GA19264 RD (woc RD)	Chromatin binding, transcription regulation	atntaggtgacctatagaaAGACCCAAGTGA TTTCATGTCG	taaccctactaaagggAGCGCATTTCCACATCG TC	805
GA19264 RE (woc RE)	Chromatin binding, transcription regulation	atntaggtgacctatagaaAGATTCGCCTCAC ACGACAG	taaccctactaaagggCGTTGACATCATCATT GGAGG	524
GA10314	Unknown function, possibly dendrite morphogenesis	atntaggtgacctatagaaAGAAACAGCTGA TGCACGGA	taaccctactaaagggCAAATATCCTTGGGCAG CG	603
GA28347 (exu2)	mRNA localisation	atntaggtgacctatagaaTGTCCTGTCGTA AAGTACC	taaccctactaaagggTGTCAGTGTCTGTTTCT G	523
GA21903 (Pbp45)	snRNA activating protein complex	atntaggtgacctatagaaGCAACTATATCCC GCTGCCT	taaccctactaaagggCCTGTTATAGGCCGTCT CC	446
GA18636 (croc)	Transcription factor, cell differentiation	atntaggtgacctatagaaATGCACACTCTGT TCAGCGA	taaccctactaaagggCGTTTGATGGCCTCCTC CTT	542
GA17585 (stc)	Transcription factor, NFX1 family	atntaggtgacctatagaaACCATTCCCTGCT CCCAG	taaccctactaaagggTTCTCCCGTTCATCGT TGG	658
GA18699 (Trs20)	Vesicle transport via TRAPP complexes	atntaggtgacctatagaaTCACACTGGCTCC AATGAACA	taaccctactaaagggATTGGCGACTTGATGAC CGT	431

Table 2.26: Genes investigated by *in situ* hybridisation. Candidate genes were identified as differentially expressed in RNA-seq of spermatocyte and spermatid cysts. Primers were designed for the amplification of candidate genes based on FlyBase sequences (Thurmond et al. 2019). Gene names and functions were inferred from orthology with *D. melanogaster*. SP6 and T3 partial primer sequences are lowercase.

2.5.2 RNA probe synthesis

Testes were dissected from 20 adult *D. pseudoobscura* males <2 days post-eclosion into a drop of testis buffer and placed into the lid of a 1.5mL Eppendorf tube. RNA was extracted from testes using the RNAqueous-Micro Total RNA Isolation kit (Invitrogen), according to the manufacturer's protocol. cDNA was synthesised by reverse transcription from testis RNA, with the SuperScript™ II Reverse Transcriptase kit (Invitrogen) and Oligo(dT) primer, according to the manufacturer's protocol.

Probe templates were synthesised by PCR of cDNA with the primers described in Table 2.26 using *Taq* 2X Master Mix (NEB). The PCR products were re-amplified with T3 and SP6 primers using *Taq* 2X Master Mix (NEB). Product yield was estimated by gel electrophoresis against a molecular weight standard DNA ladder (NEB). Probe template PCR product with T3 and SP6 extensions was purified using the QIAquick PCR Purification Kit (Qiagen), to be used for RNA probe synthesis.

Probe synthesis mix was prepared on ice (Table 2.27). 200ng probe template was added to the reaction mix and the volume adjusted to 20µL. The probe synthesis reaction mix was incubated at 37°C for 2 hours.

40µL carbonate buffer (60mM Na₂CO₃, 40mM NaHCO₃ to pH 10.2) was added to the probe synthesis reaction mix to hydrolyse the RNA probe. The hydrolysis reaction was incubated in a 60°C heat block for 15 minutes per 500bp of probe. 60µL of hydrolysis-neutralisation buffer (200mM sodium acetate, 1% acetic acid) was added to stop the hydrolysis reaction. To precipitate, three volumes of ice-cold 100% ethanol was added to the hydrolysed probe and inverted to mix, then incubated at -80°C for 30 minutes. The precipitated probe mixture was centrifuged for 15 minutes at 15000 x g, at 4°C.

The RNA pellet was washed in cold 70% ethanol. The ethanol was removed and the pellet air-dried for 10 minutes.

The RNA pellet was resuspended in 100µL sterile water and the purified probe stored at -80°C.

Probe Synthesis Mix Component	Volume (µL)
10x DIG RNA Labelling Mix (Roche)	2
10x Transcription Buffer (Roche)	2
T3 RNA Polymerase (NEB)	1
Probe Template DNA	Volume required for 200ng
dH ₂ O	To 20µL

Table 2.27: Reagent required for RNA probe synthesis mix for *in situ* hybridisation.

2.5.3 Dissection and preparation of *Drosophila* testes

For each *in situ* stain, 20 testes were dissected from WT adult male *D. pseudoobscura* <2 days post-eclosion. Testes were transferred to a drop of testis buffer in the lid of a 1.5mL Eppendorf tube.

600µL of 4% paraformaldehyde (PFA) fixing solution (4% PFA w/v in 100mM HEPES pH6.9, 2mM MgSO₄, 1mM EGTA) was added to each tube. The tubes were inverted several times and the samples incubated for 20 minutes at room temperature. Samples were washed twice in 500µL PBST for 5 minutes. PBST was removed and replaced with 500µL 50mg/mL proteinase K (Roche) in PBST for 5 minutes. Proteinase K solution was removed and 500µL 2mg/mL glycine added, and incubated for two minutes to stop the digestion reaction.

The samples were washed three times for 5 minutes in PBST, re-fixed in 4% PFA for 20 minutes, washed three times for ten minutes in PBST.

The PBST was removed and replaced with 1mL 1:1 PBST:hybridisation buffer (50% formamide, 5x SSC, 100µL/mL denatured sonicated salmon sperm DNA, 50µg/mL heparin, 0.1% Tween 20, to pH 4.5 with citric acid) for 10 minutes. The 1:1 PBST:HB was replaced with 1mL HB.

Samples were stored at -20°C in hybridisation buffer for up to two weeks.

2.5.4 *In situ* hybridisation to target mRNA transcripts in *D. pseudoobscura* testes

Testes were transferred to nylon mesh baskets in a 24-well tissue culture plate with each corresponding to a single probe. Samples were pre-hybridised in 1mL HB for 1 hour in a 65°C water bath.

2µL of RNA probe was diluted in 300µL HB. Diluted RNA probe was denatured at 80°C for 10 minutes and immediately transferred to ice for 2 minutes. Testes were hybridised with the denatured RNA-probes by incubating at 65°C in a water bath overnight.

The hybridisation solution was removed from the samples and the samples washed in HB for 10 minutes. Samples were washed a total of six times over the course of 3 hours. Samples were washed in decreasing concentrations of HB in PBST (4:1, 3:2, 2:3 and 1:4 HB:PBST) and two times in PBST, for 15 minutes per wash. Samples were incubated at 4°C overnight with alkaline phosphatase-conjugated anti-DIG antibody in PBST (1:2000). Anti-DIG antibody was pre-adsorbed by incubating with fixed *Drosophila* embryos (see 2.3.2) for at least 3 hours at 4°C.

The antibody solution was removed from the sample and the samples washed 4 x 20 minutes in PBST. Samples were washed in high pH buffer (100mM NaCl, 100mM Tris pH9.5, 50mM MgCl₂, 0.1% Tween) for 5 minutes, three times.

Samples were incubated at room temperature in NBT/X-Phosphate solution (4.5µL NBT (Roche) and 3.5µL X-Phosphate (Roche) per 1mL in PBST) for 10 minutes to 4 hours, as necessary for staining to develop. Staining solution was removed and replaced with 500µL PBST was added to the sample to stop the staining reaction.

Stained samples were washed in PBST for 5 minutes, three times. Samples were then dehydrated through an ethanol series by washing for ten minutes each in 30, 50, 70, 90, 100 and 100% ethanol. Samples were transferred from baskets to glass staining blocks in 100% ethanol. Samples were incubated in 1:1 ethanol:methyl salicylate for 15 minutes and 100% methyl salicylate for 1-15 minutes, to remove background staining. Background staining shows as pink/purple, true staining is light to dark blue.

The methyl salicylate was removed and replaced with GMM (1.6g/mL Canada Balsam in methyl salicylate). Testes were transferred in 100-200µL GMM from the staining block to a glass coverslip. The coverslip was picked up with a microscope slide, and the mounting medium was allowed to set before imaging.

Stained testes were examined on an Olympus Bx50 (Olympus Europa, Germany) at 20x magnification using differential interference contrast (DIC) microscopy. Images were taken with a JVC KY-F75U 3-CCD digital camera attachment and KYLink software (JVC Kenwood Corp., USA).

2.6 Immunofluorescence and Staining to Characterise *D. pseudoobscura* Testis Structure

2.6.1 Immunofluorescence for comparison of testis germinal proliferation centre structure of *D. pseudoobscura* and *D. melanogaster*

Immunofluorescence experiments were carried out using the method described in Jiang and White-Cooper (2003) and White-Cooper *et al.* (2000). All steps were carried out at room temperature unless otherwise specified.

Testes were dissected from 0-2 post-eclosion males into a drop of testis buffer. Testes were transferred to a fresh drop of testis buffer on a poly-L-lysine coated microscope slide. Approximately five pairs of testes were transferred to each slide. Testes were cut at the basal end with a tungsten needle, then squashed under a siliconised coverslip and immediately frozen in liquid nitrogen. The coverslip was removed with a scalpel and the slide transferred to methanol pre-chilled to -20°C. Testes were fixed in the ice-cold methanol for five minutes. Methanol was drained from the slide surface and

testes covered in acetone for five minutes. The acetone was drained from the slide and the testes covered in PBS + 1% Triton X-100 for 10 minutes.

Testes were permeabilised by two 15 minute incubations in PBS-DOC (PBS + 0.3% Triton X-100 + 0.3% sodium deoxycholate). The slide was transferred to PBST-FCS (PBS + 0.1% Triton X-100 + 5% foetal calf serum) for blocking for 30 minutes.

Primary antibodies were diluted in PBST-FCS according to Table 2.28. Testes were incubated in the primary antibody solution in a humid chamber overnight at 4°C.

The slide was washed six times in PBST for 10 minutes per wash. Secondary antibodies were diluted 1:500 in PBST-FCS (Table 2.29). RNase A was added to the secondary antibody solution to a final concentration of 0.5mg/mL. Testes were incubated in the secondary antibody solution for two hours at room temperature. The slide was washed six times in PBST for 10 minutes per wash. To counterstain DNA, testes were incubated for 10 minutes in 1µg/mL Hoechst 33258 in PBST. The Hoechst solution was drained and the slide washed for 10 minutes in PBS.

Testes were mounted under a siliconised coverslip in mounting medium (85% glycerol + 2.5% n-propyl gallate) and the coverslip sealed with clear nail polish. Testes were examined using an Olympus Bx50 (Olympus Europa, Germany) microscope and imaged with the Hamamatsu ORCA-05G camera attachment and HImage software (v.2,2,6,4 Hamamatsu Corp. 2011). Individual images were taken for each fluorescence channel. Images were processed to combine multiple fluorescence channels in Photoshop (v.23.1.0, Adobe 2022).

Antigen	Antibody	Host Species	Dilution	Supplier
vasa	Anti-vasa (Pflanz <i>et al.</i> 2015)	Rabbit	1:2000	Gift from R. Pflanz
escargot	Anti-esg (CL3573)	Mouse	1:100	Abcam (cat: ab213596)
DE-cadherin	Anti-DE cadherin (DCAD2)	Rat	1:10	DSHB (ID: AB_528120)
N-cadherin	Anti-cadN (DN-Ex #8)	Rat	1:20	DSHB (ID: AB_528121)
fasciclin III	Anti-fasIII (7G10)	Mouse	1:20	DSHB (ID: AB_528238)
armadillo	Anti-arm (N2 7A1)	Mouse	1:20	DSHB (ID: AB_528089)

hts	Anti-hts (1B1)	Mouse	1:20	DSHB (ID: AB_528070)
discs large	Anti-dlg (Papagiannouli and Mechler 2009)	Mouse	1:500	Gift from F. Papagiannouli

Table 2.28: Primary antibodies used in immunohistochemistry of squashed testes, primary antibody dilutions and host species. Anti-*esg* dilution according to Liu et al. (2019). Anti-*dlg* dilution according to Papagiannouli and Mechler (2009).

Secondary Antibody	Fluorescence	Dilution	Supplier
Goat anti-rabbit	Alexa-555	1:500	Invitrogen
Goat anti-mouse	Alexa-488		
Goat anti-mouse	Alexa-555		
Goat anti-rat	Alexa-488		

Table 2.29: Secondary antibodies and dilutions for immunohistochemistry. Alexa-488 conjugated anti-mouse was used as secondary antibody with anti-rabbit secondary antibody. Alexa-555 conjugated anti-mouse was used as secondary antibody with anti-rat secondary antibody. Where only anti-mouse secondary antibody was required, Alexa-555 conjugated anti-mouse was used.

2.6.2 Dissection and DNA staining of whole testis and seminal vesicle

Whole testes were dissected and fixed as previously described for immunohistochemistry and immunofluorescence, leaving seminal vesicle and ejaculatory duct attached. Testes and seminal vesicles were permeabilised by two 15 minute incubations in PBS-DOC, then washed in PBST for 10 minutes. PBST was drained from the slide and testes incubated in 1µg/mL Hoechst 33258 in PBST for 10 minutes. The Hoechst solution was drained and the slide washed for 10 minutes in PBS.

Testes and seminal vesicles were mounted under a siliconised coverslip in mounting medium (85% glycerol + 2.5% n-propyl gallate) and the coverslip sealed with clear nail polish. Testes and seminal vesicles were examined using an Olympus Bx50 (Olympus Europa, Germany) microscope and imaged with the Hamamatsu ORCA-05G camera attachment and HClmage software (v.2,2,6,4 Hamamatsu Corp. 2011). Images were processed in Photoshop (v.23.1.0, Adobe 2022).

2.6.3 Dissection and DNA staining of mature *D. pseudoobscura* sperm

Mature sperm were dissected from the seminal vesicle and stained for DNA for reference images of the three sperm morphs; eusperm, parasperm 1 and parasperm 2.

Testes and seminal vesicles from five males, 0-2 days post-eclosion, were dissected as previously described. Seminal vesicles were separated from the testes, accessory glands and ejaculatory duct with micro-scissors. The seminal vesicles were transferred to a fresh drop of testis buffer on a poly-l-lysine coated microscope slide. The contents emptied from the seminal vesicle into the buffer, and the drop covered with a siliconised coverslip. The slide was transferred to liquid nitrogen and the coverslip immediately removed. The slide was transferred to ice-cold methanol (-20°C) for five minutes, then to acetone for a further five minutes. Hoechst staining was then carried out as previously described for whole testes.

2.7 Generating Tagged Proteins of Differentially Expressed Genes to Investigate Localisation and Function in *D. pseudoobscura* Spermatogenesis

2.7.1 Kmg protein sequence analysis

Kmg protein sequences were downloaded from FlyBase (Gramates *et al.* 2017; Larkin *et al.* 2021). *D. melanogaster* and *D. pseudoobscura* Kmg protein sequences were aligned by Protein BLAST (Altschul *et al.* 1997; Altschul *et al.* 2005). Alignments were performed using BLOSUM62 scoring. Gap costs were set at “Existence” = 11, and “Extension” = 1. Alignments were performed across whole sequences. Alignments were visualised by the Multiple Sequence Alignment Viewer (v1.21.0, NCBI). *D. melanogaster* and *D. pseudoobscura* Kmg protein domains were predicted by analysis of the amino acid sequences by InterPro (Blum *et al.* 2021). *D. melanogaster* Kmg protein domains are also described by Kim *et al.* (2017). *D. melanogaster* and *D. pseudoobscura* Kmg protein sequences are available in Appendix 1.

2.7.2 Protein-GFP fusion for genome insertion with the piggyBac transposon system

GFP fusion constructs were designed for *kmg* using the piggyBac insertion vector (gift from David Navarro). The piggyBac vector contained a 3xP3-AmCyan eye marker to allow selection of transgenics. Insertion of the piggyBac-AmCyan-gene-GFP construct within the genome should result in expression of the tagged protein, showing protein localisation within the testes.

2.7.2.1 piggyBac Kmg N- and C-terminal GFP Tag Construct Designs and Assembly

Two versions of the Kmg-GFP fusion construct were designed; piggyBac-3xP3-AmCyan-Kmg-GFP(N) and piggyBac-3xP3-AmCyan-Kmg-GFP(C), for N- and C-terminal GFP tagged Kmg protein, respectively (Figure 2.3). To design Kmg-GFP fusion constructs, piggyBac, *kmg* and GFP sequences were required.

D. pseudoobscura kmg sequence was downloaded from Flybase (Gramates *et al.* 2017; Thurmond *et al.* 2019; Larkin *et al.* 2021). GFP sequence can be found in Parker *et al.* (2001).

Primers were designed to amplify the required regions of the *kmg* gene. For the C-terminal GFP tag, primers were designed for amplification of the promoter, 5' UTR and gene fragment, and for amplification of the 3' UTR fragment. Primers included restriction sites for ligation with GFP and piggyBac. For the N-terminal GFP tag, primers were designed for amplification of the promoter and 5' UTR fragment, and for amplification of the gene and 3' UTR fragment. Primers were designed to amplify GFP from genomic DNA; a start codon was included in the primer for the GFP N-terminal tag. Primer sequences were designed to add restriction sites to the target sequence, for ligation into the piggyBac construct. Primers were designed such that the GFP and *kmg* sequences were in-frame after ligation of the fragments.

kmg fragments were amplified from synthesised gene blocks (IDT). GFP fragments were amplified from genomic DNA extracted with the general DNA extraction method as previously described, from *D. melanogaster w;UAS-eGFP-TEVt-RFP-M4/CyO* (gift from Jianqiao Jiang). Fragments were amplified with GoTaq G2 Master Mix (Promega) and PCR products run on a 1.1% agarose gel to confirm fragment size. Products were purified by gel extraction with the QIAquick Gel Extraction Kit (Qiagen). PCR products were cloned into pGEM[®]-T Easy (Promega) by TA cloning with T4 DNA Ligase (Figure 2.4). Sequences were confirmed by sequencing (Eurofins).

Kmg and GFP fragments were digested from pGEM[®]-T Easy with AvrII and NotI (Table 2.31). Digested fragments were purified by gel extraction with the QIAquick Gel Extraction Kit. Purified fragments were quantified using the Nanodrop ND1000 (Thermo Scientific).

Fragments were sequentially ligated into piggyBac. Ligations were carried out with T4 DNA Ligase (NEB). The GFP fragment was ligated into piggyBac first, using AvrII and NotI restriction sites. The piggyBac-GFP ligation was transformed, grown and purified as previously described (see 2.2.11 and 2.2.12).

Purified plasmid samples were sequenced to confirm the presence of the GFP sequence in piggyBac (Table 2.32).

For the C-terminal tag, the *kmg* 3' UTR was then ligated in to the piggyBac GFP backbone. piggyBac was digested with NotI to produce a linear fragment. *kmg* 3' UTR was digested out of pGEM[®]-T Easy with NotI and purified by gel extraction with the QIAquick Gel Extraction Kit. *kmg* 3' UTR was ligated into piggyBac GFP with T4 DNA Ligase (NEB). Ligation product was transformed as previously described. Presence of the *kmg* 3' UTR insert was assessed by digesting ~300ng purified plasmid

product with NotI. The digest was run on a 1.1% agarose gel. The presence of a 442bp band indicated the successful ligation of the *kmg* 3' UTR insert (Table 2.30). The plasmid was sequenced to confirm the correct orientation of the *kmg* 3' UTR insert (Eurofins).

piggyBac GFP *kmg* 3' UTR was digested with AvrII and purified with the QIAquick PCR Purification Kit (Qiagen). *kmg* promoter + 5' UTR + coding sequence was digested from pGEM[®]-T Easy with AvrII and purified by gel extraction. piggyBac GFP *kmg* 3' UTR and *kmg* promoter + 5' UTR + coding sequence were ligated with T4 DNA ligase (NEB), and transformed as previously described. Purified plasmid was digested with AvrII to confirm the presence of the insert (Table 2.30). To confirm the orientation of the insert, purified plasmid samples containing the insert were sequenced (Eurofins).

The N-terminal construct was assembled with a similar sequential method. Once the presence of N-terminal GFP in piggyBac was confirmed, the AvrII site was used to ligate the *kmg* promoter + 5' UTR fragment. Once presence of the fragment and correct orientation were confirmed, the NotI site was used to ligate the *kmg* coding sequence + 3' UTR fragment.

For both C- and N-terminal constructs, the smaller of the two *kmg* fragments were ligated into the construct after GFP, as it was easier to distinguish between religated piggyBac GFP and constructs with the small insert when the overall construct size was lower.

Completed constructs were sequenced to confirm correct assembly. Constructs were retransformed into Mix & Go DH5 α (Zymo Research), grown in 100mL LB ampicillin media at 37°C overnight. The plasmid DNA was purified by maxiprep (Qiagen).

Injections were carried out as previously described (see 2.3.5).

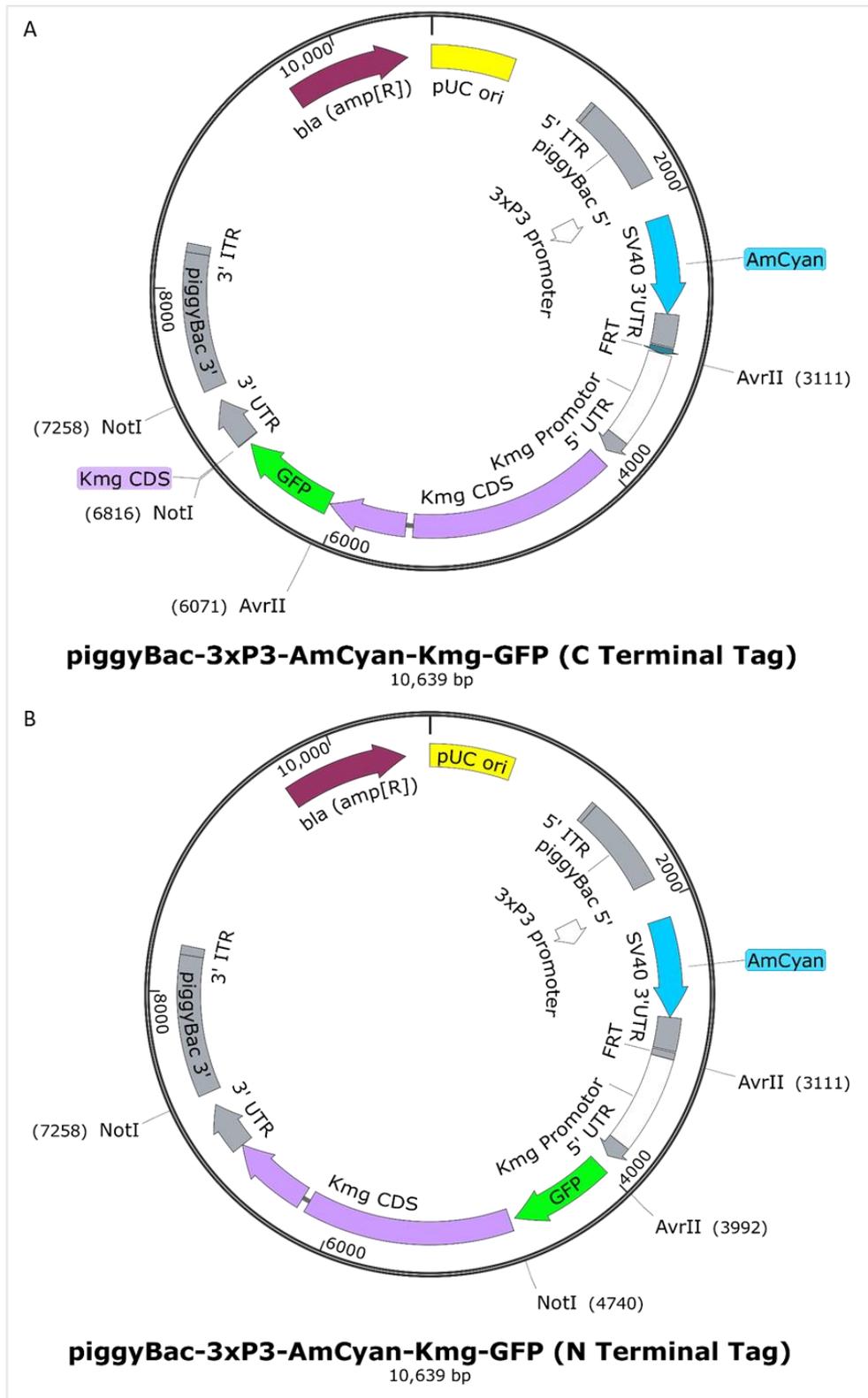


Figure 2.3: Construct designs for piggyBac mediated insertion of GFP tagged Kmg in *D. pseudoobscura*. piggyBac insertion vector contains 3xP3-AmCyan eye marker for selection of transgenic F1 adults. GFP tag in frame with kmg CDS. Constructs designed in SnapGene (v3.0.3. GSL Biotech LLC. 2011). (A) piggyBac vector, AmCyan eye marker, kmg promotor, UTR and coding sequences with integrated C-terminal GFP tag. (B) piggyBac vector, AmCyan eye marker, kmg promotor, UTR and coding sequences with integrated N-terminal GFP tag.

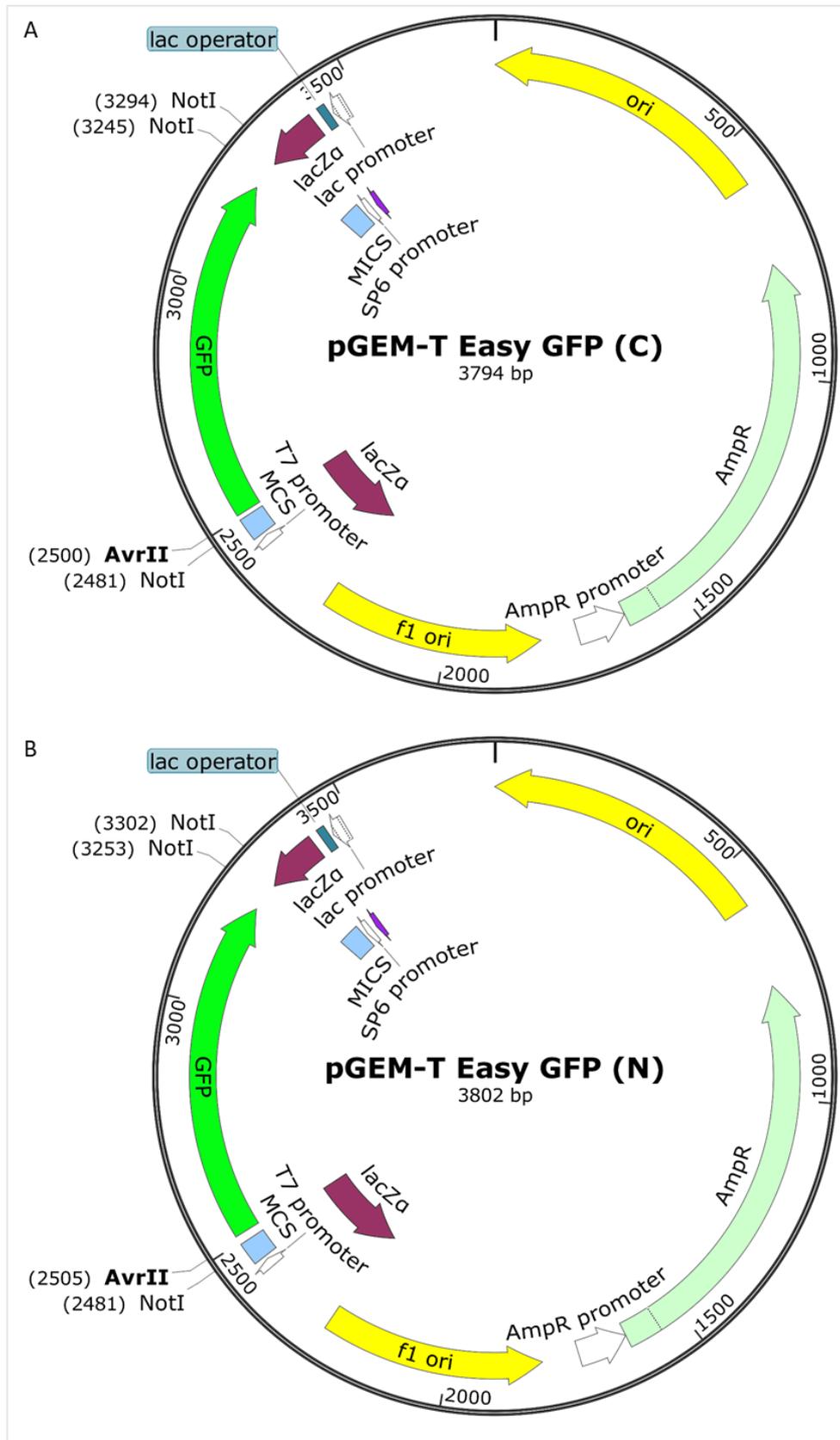


Figure 2.4: GFP PCR product in pGEM[®]-T Easy (Promega) for construction of piggyBac GFP tag constructs. AvrII restriction site from forward primers. NotI restriction site already present in GFP sequence. (A) GFP for C-terminal tags. (B) GFP for N-terminal tags, with start codon.

piggyBac-Kmg-GFP			
Fusion Construct Component	Forward Primer	Reverse Primer	Predicted Product Size (bp)
piggyBac	-	-	6492
kmg Promotor + 5' UTR	cctaggCTTTTGCAGGA GGTTTCGCC	cctaggTTGCTTAAAAGA AATGGAGATTACATGA AC	881
kmg Promotor + 5' UTR + Coding Sequence	cctaggCTTTTGCAGGA GGTTTCGCC	cctaggATATTGTTTCATC ACCGGCGGG	2960
kmg Coding Sequence + 3' UTR	gcggccgcGGATTTACCA TCAATGGTAGAAAGTA CTGGC	gcggccgcGCCGCTAGTC GTCTTAATATG	2518
kmg 3' UTR	gcggccgcTTAGAACCGG ACTCTGACTGAAAAGG G	gcggccgcGCCGCTAGTC GTCTTAATATG	442
GFP N-tag	cctaggATGAGTAAAGG AGAAGA ACTTTTCACT GGA	GCTCTCCCATATGGTC GACCTG	PCR product = 780 Digest product = 748
GFP C-tag	cctaggAGTAAAGGAGA AGA ACTTTTCACTGG	GCTCTCCCATATGGTC GACCTG	PCR product = 777 Digest product = 745
piggyBac kmg GFP C-Terminal Tag Construct Total	-	-	10639
piggyBac kmg GFP N-Terminal Tag Construct Total	-	-	10639

Table 2.30: Primer sequences for PCR of kmg and GFP fragments for GFP fusion constructs assembled in piggyBac. Restriction site extensions shown in lowercase, target sequences shown in uppercase. Predicted fragment sizes are given.

piggyBac-Kmg-GFP Fusion		
Construct Component	5' Restriction Site	3' Restriction Site
piggyBac	AvrII	NotI
kmg Promotor + 5' UTR	AvrII	AvrII

<i>kmg</i> Promotor + 5' UTR + Coding Sequence	AvrII	AvrII
<i>kmg</i> Coding Sequence + 3' UTR	NotI	NotI
<i>kmg</i> 3' UTR	NotI	NotI
GFP N-tag	AvrII	NotI
GFP C-tag	AvrII	NotI

Table 2.31: Restriction sites for ligation of *kmg* fragments and GFP sequence into piggyBac vector.

Primer Name	Primer Sequence
piggyBac Sequencing Forward	GCCAGTTCACACCTCCTAC
GFP Sequencing Reverse	ACGTGTCTTGTAGTTCCCGTC
<i>kmg</i> Sequencing 1	CTCACAGGAAGTCGCATGGT
<i>kmg</i> Sequencing 2	GTACCAGTGCCAGAAGTGCT
<i>kmg</i> Sequencing 3 Reverse	GTCACGATTCGCCTTCAAGC

Table 2.32: Primer sequences for sequencing gene fragments and purified ligation products.

2.7.2.2 Collection of F0 injection survivors

Injection survivors were collected within 24 hours of eclosion. Adult F0 flies were crossed with *D. pseudoobscura* SLOB3 wild-type. Male F0 survivors were crossed with 5-6 virgin females 0-3 days post-eclosion. Female F0 survivors were crossed with 2-3 males 0-3 days post-eclosion. Crosses were maintained at 21°C.

2.7.2.3 Screening for transgenic F1 from F0 survivor crosses

F1 offspring from F0 injection survivor crosses were collected with 24 hours of eclosion. F1 adults were screened for the AmCyan fluorescent eye-marker, indicating successful piggyBac mediated transgenesis. Transgenic F1 adults were crossed with *D. pseudoobscura* SLOB3 wild-type. Crosses were maintained at 25°C. Transgenic F1 adults were collected prior to eclosion of the F2 for sequencing, and stored at -20°C prior to DNA extraction.

2.7.2.4 F2 transgenic collection and strategy for generating stable Kmg-GFP fusion lines

F2 offspring from F1 transgenic crosses were collected within 24 hours of eclosion. F2 adults were screened for the AmCyan fluorescent eye-marker. F2 transgenic males were crossed with virgin F2 transgenic females to generate homozygous F3 transgenic *D. pseudoobscura*. F3 transgenic *D. pseudoobscura* were crossed, and homozygous lines maintained.

3 Developing Integrated *cas9* Lines for CRISPR Cas9 Gene Editing in *D. pseudoobscura*

Drosophila pseudoobscura, as a non-model *Drosophila* species, presents particular challenges in experimental design and use as an experimental model system. In particular, the lack of genetic tools limits the level of investigation possible into gene function in this species. In attempting to gain further insights into the genetic control of sperm morph development in *D. pseudoobscura*, it was important to expand the range of genetic tools available. The development of the CRISPR Cas9 gene editing system has resulted in a powerful tool for gene editing in non-model species (Sun *et al.* 2017). In order to allow subsequent studies of genes of interest, I designed an approach to adapt this system to *D. pseudoobscura*, based on the methods developed in *D. melanogaster*.

Multiple methods for use of the CRISPR/Cas9 gene editing system in *D. melanogaster* have been described. An early method relied on injection of plasmids containing *cas9* under the non-tissue specific *hsp70* promoter, with a second sgRNA vector containing the guide RNA backbone and target specific sequences under the U6 promoter (Gratz *et al.* 2013). Another method used direct injection of *in vitro* transcribed *cas9* mRNA and sgRNA, which was highly efficient for generation of mosaic adults in the F0 generation (Bassett *et al.* 2013). An alternative approach relied on genomic integration of the *cas9* sequence with the germline-specific promoters and regulatory elements of *vasa* and *nanos* for germline expression and localisation to the germline, therefore increasing the likelihood that mutagenesis will take place in the germline and can be transmitted to progeny (Kondo and Ueda 2013; Ren *et al.* 2013; Ren *et al.* 2014; Sebo *et al.* 2014). Direct comparisons of efficiency between *hsp70* and *vasa* promoter and regulatory elements for controlling *cas9* expression found that germline-specific expression under the *vasa* promoter and UTRs resulted in a higher percentage of mutations transmitted to progeny (Gratz *et al.* 2014). Since the approach of an integrated *cas9* with germline-specific expression was previously demonstrated to be highly efficient, I selected this method to adapt for use in *D. pseudoobscura*.

To generate *D. pseudoobscura* lines expressing Cas9 in the germline, I designed a construct based on the piggyBac transposon vector, which has been successfully used for transgenesis in other non-model *Drosophila* and insect species (Handler *et al.* 1998; Horn and Wimmer 2000; Labbe *et al.* 2010; Schetelig and Handler 2013; Tanaka *et al.* 2016; Kalajdzic and Schetelig 2017; Kudo *et al.* 2018). The *cas9* sequence was flanked by the *D. pseudoobscura* GA19020 *nanos* (*nos*) promoter, 5' UTR and 3' UTR control regions, for germline expression and RNA localisation. This was based on the method

described for *D. melanogaster* (Kondo and Ueda 2013; Gratz *et al.* 2015). The construct was assembled and injected into *D. pseudoobscura* embryos. Transgenic F1 progeny of injected F0 were collected and homozygote lines established from their progeny.

The broad aim of generating integrated Cas9 lines was to use this system for genome editing for further investigation of genes of interest potentially implicated in development of the multiple sperm morphs in *D. pseudoobscura*. Validation of the system was first attempted by targeting two X-linked genes, *white* and *yellow*, as these should allow for the rapid identification of successful mutants, and thus selection of the best lines for future use. In addition, *white* and *yellow* mutant backgrounds would be of use for future gene editing experiments. Unfortunately, this validation step revealed that none of the lines generated *white* or *yellow* mutants after injection of the sgRNA constructs. As a result, Cas9 lines could not be used for subsequent gene editing experiments.

3.1 Analysis of *D. pseudoobscura* Embryo Development

Injection of *D. pseudoobscura* embryos, either with a piggyBac transposon vector, or subsequently with guide RNA expression vectors, must take place prior to formation of the pole cells – the germline precursors. To establish the timing of pole cell formation, embryos were collected at hourly intervals and aged for 0-5 hours at 25°C, fixed, and stained for DNA with Hoechst 33258. Embryo stages were determined with reference to *D. melanogaster* (Foe *et al.* 1993).

A total of 46 embryos were fixed, stained with Hoechst 33258, and imaged. Examples of embryos fixed at each time point after egg deposition (AED) are shown in Figure 3.1.

Embryos 0-1 hours AED contained up to eight nuclei (Figure 3.1; A-D). Nuclei were concentrated in the interior of the embryo. The polar body was visible in some embryos. The embryo at these stages is syncytial; nuclei are in a common cytoplasm, and mitosis occurs without the process of cell division. Cellularisation does not take place until the pole cells bud off from the embryo posterior, after which the syncytial blastoderm will undergo cellularisation by invagination of the plasma membrane.

1-2 hours AED embryos had undergone further mitotic divisions as indicated by the increased number of syncytial nuclei (Figure 3.1; E-H). Nuclei were still concentrated in the interior of the embryo, however nuclear migration to the embryo periphery was occurring in some later stage embryos (Figure 3.1; H). Condensed DNA is not clear in Figure 3.1; F, suggesting this embryo may be in interphase, so DNA is more condensed and harder to detect.

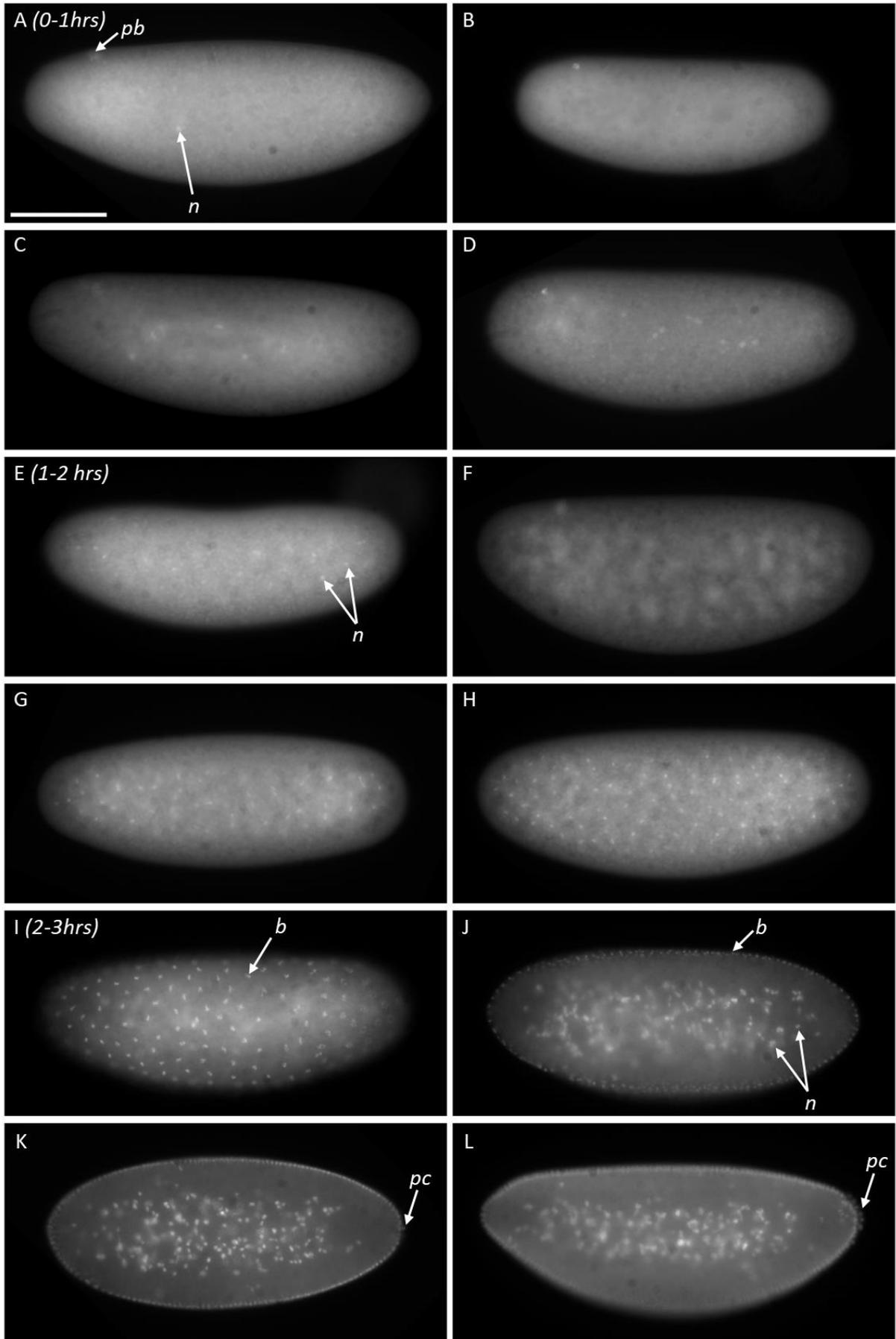
At 2-3 hours AED, nuclear migration to form the syncytial blastoderm had occurred, with interior nuclei forming the yolk (Figure 3.1; I-L). Figure 3.1; J shows an embryo immediately prior to pole cell

formation. Cellularisation of the pole cells at the posterior was observed in some embryos 2-3 hours AED (Figure 3.1; K and L).

3-4 hours AED embryos showed substantial variation in the precise stage of development, although pole cell formation had occurred in all embryos by this time point (Figure 3.1; M-P). Some embryos were undergoing the process of blastoderm cellularisation (Figure 3.1; M). Some embryos were undergoing formation of the cephalic and posterior transverse furrows (Figure 3.1; N and O). Some embryos were observed to have completed formation of the cephalic and posterior transverse furrows, in addition to invagination of the ventral furrow, containing the pole cells (Figure 3.1; P).

At 4-5 hours AED, embryos had completed formation of the cephalic and posterior transverse furrows, and invagination of the ventral furrow, and germ band migration was observed (Figure 3.1; Q-S).

Formation of the pole cells was observed between 2-3 hours AED. Injection of *D. pseudoobscura* embryos was therefore performed 1-2 hours AED. Embryos with clear blastoderm cellularisation were not injected.



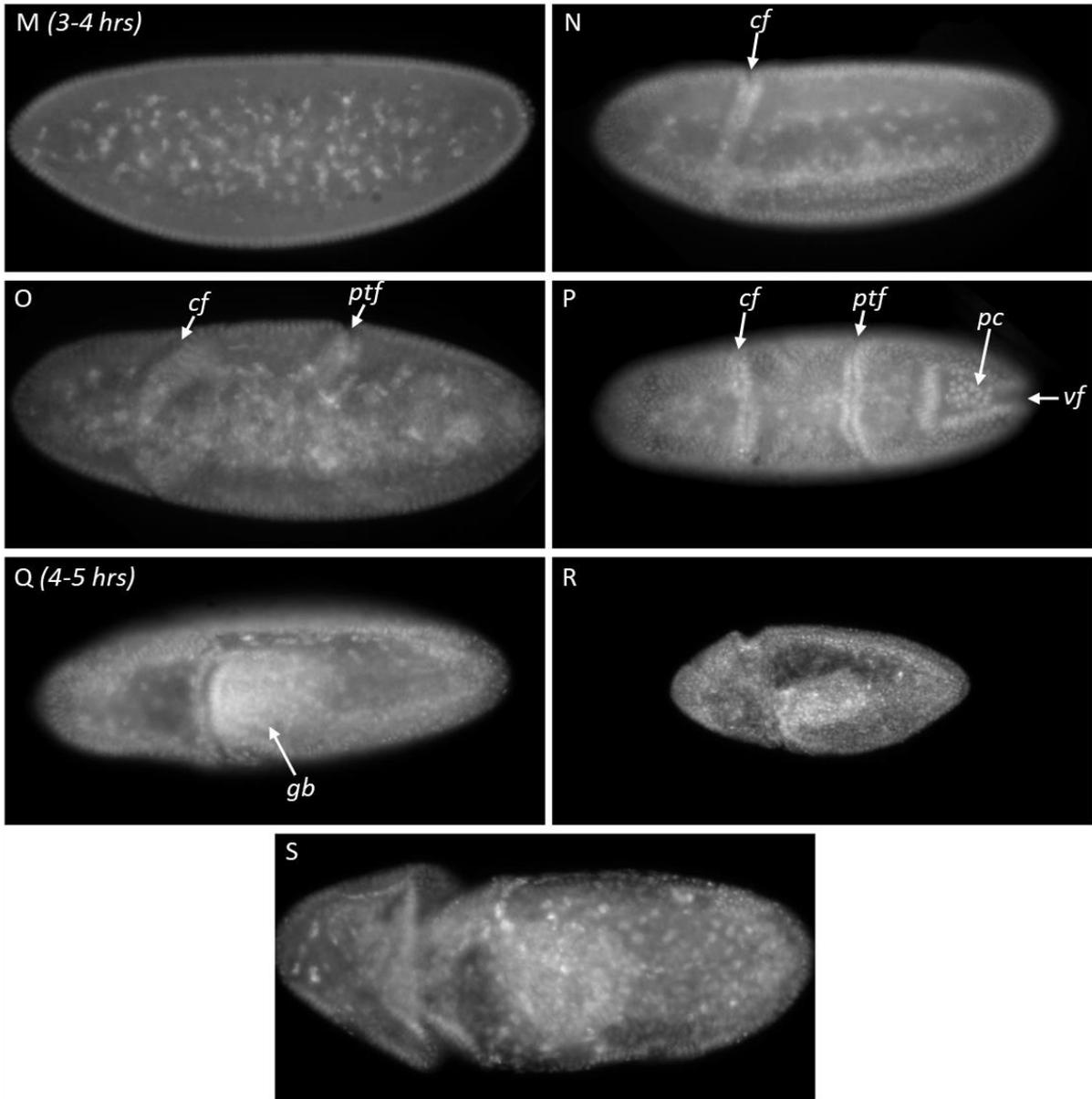


Figure 3.1: *D. pseudoobscura* SLOB3 wild type embryos stained for DNA (Hoechst 33258). Embryos were fixed at regular time points to establish timing of development stages. *pb* = polar body, *n* = nucleus/nuclei, *b* = blastoderm, *pc* = pole cells, *cf* = cephalic furrow, *ptf* = posterior transverse furrow, *vf* = ventral furrow, *gb* = germ band. Scale = 100 μ m. A-D: 0-1 hours after egg deposition (AED). Polar body visible. Up to 8 nuclei visible in embryo interior. E-H: 1-2 hours AED. More nuclei present, nuclei continue to divide within the interior of the embryo. Nuclei begin migration outwards to periphery. I-L: 2-3 hours AED. Nuclei have migrated outwards to form syncytial blastoderm. Internal nuclei form yolk. Formation of pole cells at posterior. M-P: 3-4 hours AED. Cellularisation of blastoderm. Cephalic furrow and posterior transverse furrow form. Invagination of the ventral furrow internalises pole cells. Q-S: 4-5 hours AED. Anterior migration of the germ band (Foe et al. 1993).

3.2 Generating *D. pseudoobscura* Lines Expressing *cas9* Under the *nanos* Promotor

3.2.1 Cas9 construct for genome integration and expression: assembly and verification

Insertion of the *cas9* gene into *D. pseudoobscura* for gene editing required assembly of a construct capable of insertion into the *D. pseudoobscura* genome. The piggyBac transposon system was chosen as it can be used for insertion into insect genomes, with no reliance on other genetic tools. To ensure expression and localisation of Cas9 in the *D. pseudoobscura* germline, the gene sequence was flanked by the 430bp *nos* promotor and 221bp 5' UTR, and the 227bp 3' UTR and 702bp downstream region. *nos* drives transcription in the female germline and localises to the poleplasm, so transcripts are incorporated in the pole cells (Verrotti and Wharton 2000; Forrest and Gavis 2003; Forrest *et al.* 2004).

The *nos-cas9* cassette was assembled first in the pBluescript KS+ vector, as this was a smaller vector with more restriction sites available, therefore allowing for easier cloning. Sequencing of the pBluescript-*nos*-Cas9 construct showed the presence of *nos* promotor and 5' UTR, *cas9* and *nos* 3' UTR in the correct assembly and orientation (sequences available in Appendix 2).

The *nos-cas9* cassette was digested out of the pBluescript backbone and ligated into the piggyBac-3xP3-AmCyan vector. piggyBac-3xP3-AmCyan-*nos*-Cas9 assembly was confirmed by restriction digest. Two bands were present on an agarose gel after restriction digest; indicating that the piggyBac-3xP3 backbone and *nos-cas9* insert were present.

3.2.2 Embryo injection with piggyBac-*nos*-Cas9 construct

The piggyBac-3xP3-AmCyan-*nos*-Cas9 construct was injected into embryos 1-2 hours AED, with the piggyBac helper plasmid.

A total of 68 injected adults (F0) were collected and crossed individually (30 females, 38 males). Females were crossed with two wild type males. Males were crossed with four wild-type females. From F0 adult crosses, three females and five males produced no progeny.

3.2.3 Screening and collection of transgenic F1

F1 progeny of F0 injection survivors and wild type crosses were screened for the AmCyan fluorescent eye marker, indicating insertion of the piggyBac-3xP3-AmCyan-*nos*-Cas9 construct.

Transgenic F1 were collected from five of the 60 successful F0 crosses. A total of 10 transgenic F1 adults were collected (Table 3.1). Transgenic F1 showed strong fluorescence in the ocelli, with weaker

fluorescence in the compound eye (Figure 3.2). The 3xP3-AmCyan eye marker was weaker in *D. pseudoobscura* compared to *D. melanogaster*. This may have been due to the presence of red eye pigmentation in *D. pseudoobscura*, whereas previous applications of the piggyBac vector backbone in our lab had utilised a *white*-background *D. melanogaster* line. The lack of eye pigmentation in these *D. melanogaster* lines may result in less interference with the fluorescence emission. It is also possible that the 3xP3 eye promoter results in a more localised expression in *D. pseudoobscura*, compared to that observed in *D. melanogaster*.

<i>D. pseudoobscura</i> Mosaic F0 Parent ID	3xP3-AmCyan-nos-cas9 F1 Offspring ID	Male/Female
27	27A	Male
29	29A	Female
	29B	Female
	29C	Male
	29D	Male
	29E	Male
	29F	Male
49	49A	Female
	49B	Female
55	55A	Male
61	61A	Female

Table 3.1: *D. pseudoobscura* piggyBac 3xP3-AmCyan-nos-cas9 F0 injection survivors collected which subsequently produced transgenic F1 progeny.

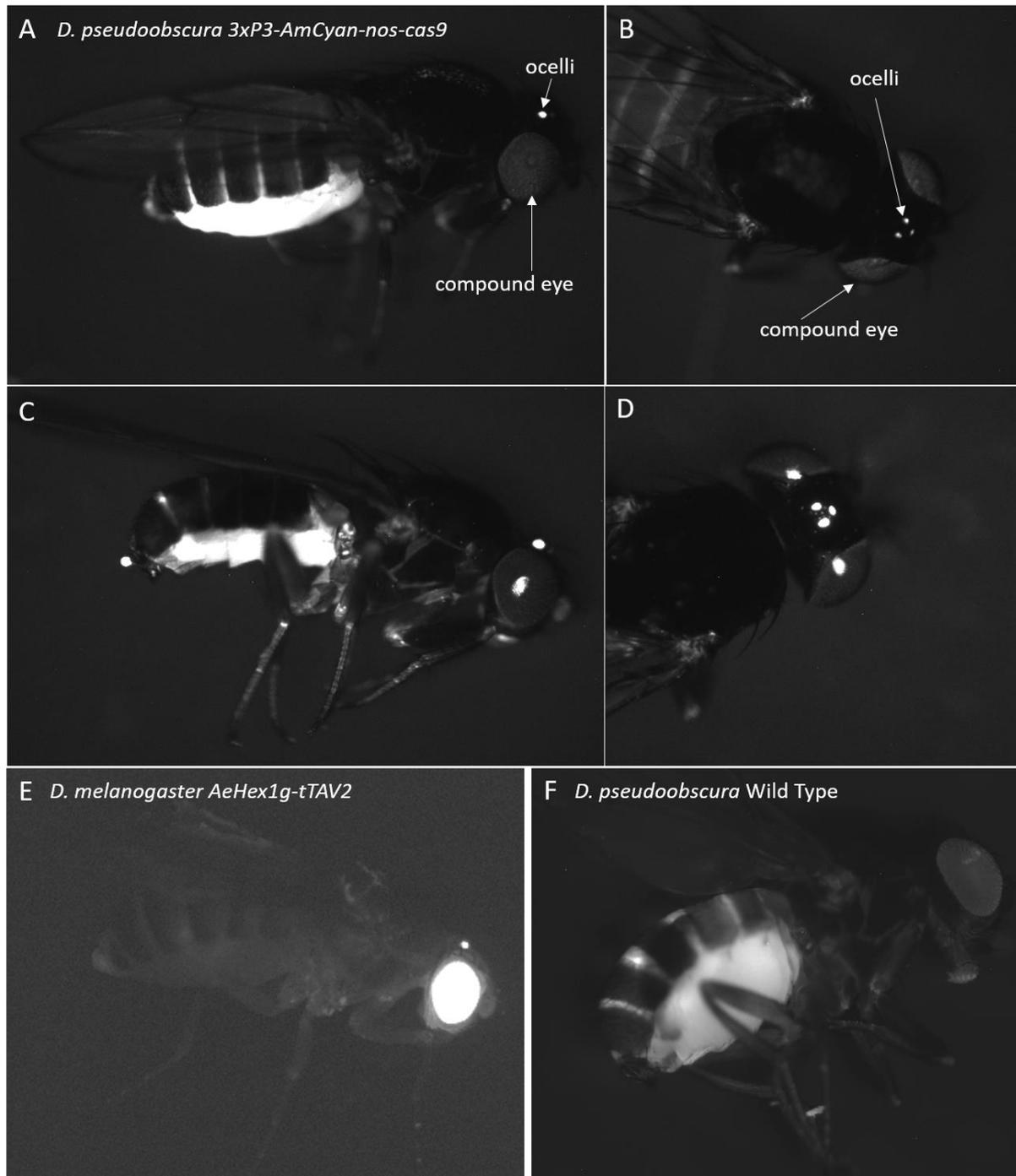


Figure 3.2: *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 and *D. melanogaster* AeHex1g-tTAV2 exhibiting cyan fluorescence from AmCyan eye marker. A-D: *D. pseudoobscura* F1 transgenic offspring from *D. pseudoobscura* piggyBac-3xP3-AmCyan-nos-Cas9 F0 injection survivors crossed with wild type. Ocelli and compound eye exhibiting fluorescence from AmCyan eye marker. Fluorescence also present in abdomen. E: *D. melanogaster* AeHex1g-tTAV2 exhibiting fluorescence from the same 3xP3-AmCyan eye marker. Compound eye has greater fluorescence intensity in comparison to *D. pseudoobscura*. F: *D. pseudoobscura* SLOB3 wild type. Compound eye does not exhibit cyan fluorescence. No fluorescence in abdomen.

3.2.4 F2 and F3 crosses generated stable *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 lines

Transgenic F1 individuals were crossed with WT. F2 offspring were screened and transgenic F2 collected within 18h of eclosion. Transgenic F2 were backcrossed to another transgenic F2 from the same F1 line (Figure 3.3).

F3 flies were screened for the AmCyan eye marker. Comparisons between transgenic F3 from some lines showed apparently discrete variation in eye marker fluorescence intensity. These individuals were assumed to be homozygote for the transgene and selectively chosen for F3 crosses.

Where possible, homozygote F3 males and females were crossed to generate an F4 population homozygous for the piggyBac transgene insert. Where there was no obvious difference, transgenic individuals were crossed and selected again at the F4 stage. F4 were then selected for crossing based on the intensity of the fluorescent eye marker. Where there was no obvious variation between heterozygotes and homozygotes in subsequent generations, it was assumed that the homozygote condition was lethal and the stock maintained as heterozygous.

Some F2 crosses resulted in all-transgenic females, and males either heterozygous or not transgenic. In these cases, homozygote female F2 flies were crossed with heterozygote male F2 flies, generating all transgenic F3 populations, with females homozygous for the piggyBac gene insert and males heterozygous for the gene insert. These populations could then be maintained as transgenic without selection. It was assumed in these cases that the gene insert was on the X chromosome.

The results of F2 and F3 crosses can be found in Table 3.2.

F1 male 27A produced only transgenic F2 females, and approximately equal proportions of transgenic and non-transgenic F2 males. 27A likely had two insertions, one autosomal and one X-linked.

F1 male 29D produced transgenic females, and non-transgenic males, suggesting that the transgene insertion was on the X-chromosome.

F1 males 29F and 55A also produced non-Mendelian F2 progeny ratios, however it was not clear where the transgene insertion was.

F1 Transgenic ID	F/M	Transgenic F2 Females	Transgenic F2 Males	No Transgene F2 Females	No Transgene F2 Males	Chi Square P Value
29B	F	18	32	25	34	0.12

27A	M	57	35	0	26	3.16^{E-12}
29C	F	25	19	13	28	0.10
29D	M	57	0	0	45	1.43^{E-22}
29E	M	26	16	20	24	0.43
29F	M	11	14	33	40	1.79^{E-05}
49A	F	22	21	23	9	0.08
49B	F	16	11	15	17	0.70
55A	M	23	17	35	35	0.03
65A	F	34	26	34	25	0.49

Table 3.2: Progeny from F1 transgenic crosses. Progeny were screened for 3xP3-AmCyan eye marker. Chi-square test for fit to null hypothesis of Mendelian 1:1:1:1 ratio.

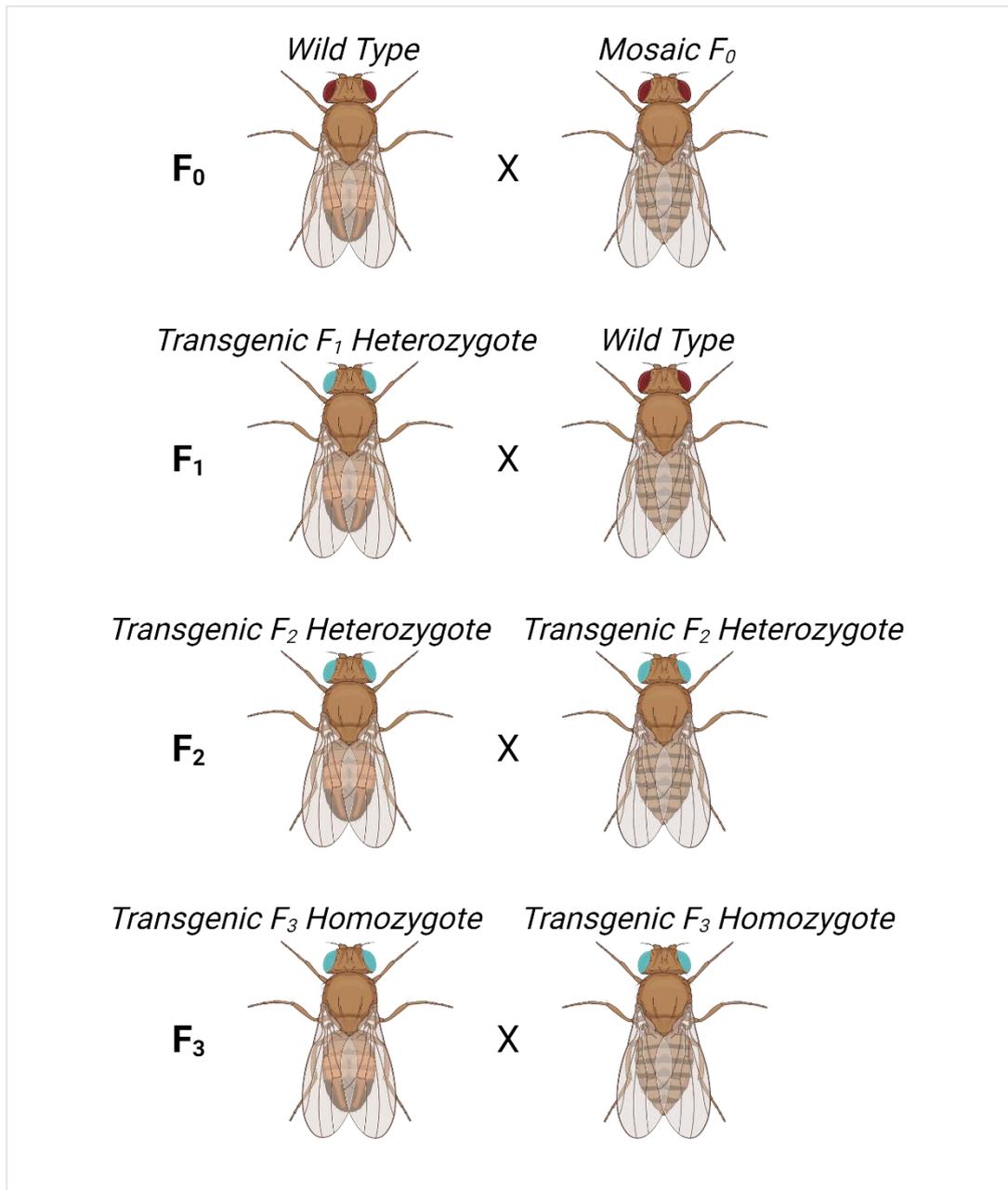


Figure 3.3: Crosses to generate stable transgenic *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 lines. Cyan eye indicates transgenic.

3.2.5 Locations of piggyBac insertion in F2 transgenic *D. pseudoobscura*

DNA was extracted from F2 transgenic *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 to determine the location of piggyBac gene inserts in the genome. An inverse PCR method was used to amplify the regions upstream and downstream of the insert. PCR products were cloned into pGEM-T Easy and the constructs sequenced with T7 and SP6 primers.

Full sequences of inverse PCR products in pGEM-T Easy are available in Appendix 2. Inverse PCR BLAST hits are summarised in Table 3.3.

3.2.5.1 F1 transgenic male 27A

Line 27A had multiple insertions, according to F2 transgenic offspring ratios. One insertion location was identified by inverse PCR, in the gene region of *GA24860* on chromosome 3. The second insertion site on the X chromosome was not identified.

3.2.5.2 F1 progeny of transgenic F0 29

Five F1 transgenic progeny were produced by the F0 injection survivor ID 29. F2 transgenic offspring ratios of F1 male 29D indicated an insertion site on the X chromosome. F2 transgenic progeny of F1 male 29E indicated an autosomal insertion. It was assumed therefore that there were multiple insertion events in injection survivor ID 29, and as such results of F1 29B-F inverse PCR BLAST searches returned multiple hit sites.

F1 transgenic 29B inverse PCR identified multiple potential insertion sites on chromosome 2 and 3. Two potential insertion sites were identified in F1 transgenic 29C; *GA13327* on the X chromosome, and *GA26607* on chromosome 2. Insertion sites could not be identified in F1 transgenics 29D or 29E. Inverse PCR identified an insertion site in 29F in the gene region of *GA28527*, *alpha-tocopherol transfer protein-like*.

F1 transgenic male 29D F2 females were transgenic, and males non-transgenic, indicating that the transgene was present on the X chromosome. Inverse PCR did not identify the insertion site. A region of the *D. melanogaster* X chromosome was identified by BLAST, which may be orthologous to the insert site in *D. pseudoobscura*.

F1 transgenic male 29F progeny ratios were significantly different from Mendelian ratios (χ^2 Test, $p = 1.79 \times 10^{-5}$). F2 transgenic and non-transgenic males and females were present in the F2 generation, although fewer transgenic males and females were produced than expected. Inverse PCR indicated the transgene insert was present on the X chromosome, counter to the presence of transgenic males and non-transgenic females in the F2 generation. It is possible that the insert had disrupted a meiotic drive element which then decreased transmission to the F2 generation, or that there were multiple insertions with at least one having an effect on viability.

3.2.5.3 F1 progeny of transgenic F0 49

F1 transgenic female 49A progeny ratios were not significantly different from Mendelian ratios (χ^2 Test, $p = 0.08$), although the number of non-transgenic males appeared low. It was not possible to identify the location of the insert from inverse PCR data.

F1 transgenic female 49B produced approximately equal numbers of transgenic and non-transgenic females and males. Female F1 progeny ratios did not distinguish between autosomal or X chromosome insertion. Inverse PCR did indicate a possible insertion site on the X chromosome.

3.2.5.4 F1 transgenic male 55A

F1 transgenic male 55A progeny ratios were non-Mendelian (χ^2 Test, $p = 0.03$), with a lower than expected proportion of transgenic males and females. The presence of both transgenic and non-transgenic males and females in the F2 generation indicated that the insert was autosomal. The data indicate that the transgene insert reduced viability.

Line	Inverse PCR BLAST Hits	Chromosome	Location
27A	GA24860	3	NC_046680.1 (13939328..13940795)
29B	GA14970	2	NC_046679.1 (31251209..31255172, complement)
	GA19318	2	NC_046679.1 (31244748..31250248, complement)
	GA24746	3	NC_046680.1 (9501773..9506273)
	GA26139	2	NC_046679.1 (31219687..31225814, complement)
	LOC117183732 GA Unavailable	2	NC_046679.1 (31226894..31233010, complement)
	GA14972	2	NC_046679.1 (31234280..31239234, complement)
	<i>D. melanogaster</i> Chromosome 2R		22749452-22749503
	29C	GA13327	X

	GA26607	2	NC_046679.1 (11241531..11243014, complement)
	<i>D. melanogaster</i> Chromosome 2L		6820311-6820413
29D	<i>D. melanogaster</i> Chromosome X		18985845-18985885
29E	-	-	-
29F	GA28527	X	NC_046683.1 (58802634..58803801)
49A	-	-	-
49B	LOC117184636 GA Unavailable	X	NC_046683.1 (33555890..33557554, complement)
55A	-	-	-
61A	-	-	-

Table 3.3: *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 transgenic F1 inverse PCR sequence BLAST hits. BLAST hits indicated possible transgene insertion sites. Chromosome and gene locations of insert sites are given.

3.3 Validation of CRISPR Cas9 Gene Editing in *D. pseudoobscura*

3.3.1 Confirmation of *cas9* expression in *D. pseudoobscura* ovaries

Integration of the piggyBac-3xP3-AmCyan-nos-Cas9 construct in the genome was determined by expression of the 3xP3-AmCyan eye marker. Expression of *nos-cas9* mRNA required further confirmation by PCR of cDNA from transgenic lines. RNA was extracted from ovaries of transgenic *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9. Two sets of primers were used to amplify two regions of the *cas9* mRNA, with expected PCR product sizes of 930 and 468 base pairs. Agarose gels of PCR products from transgenic *D. pseudoobscura* cDNA are shown in Figure 3.4.

PCR of cDNA from transgenic lines indicated that the *cas9* RNA was expressed in ovaries. There is substantial variation in the intensity of bands indicating variable concentrations of PCR product. This suggests that *cas9* expression is variable between the transgenic lines tested here, although some of the observed variation may have resulted from variability in cDNA template concentration, since a positive endogenous gene PCR was not conducted on these samples.

The *nos* promoter and UTRs flanking the *cas9* sequence should result in germline specific expression and localisation to the poleplasm. PCR of ovary cDNA suggests that the *cas9* is expressed in the female germline, as predicted.

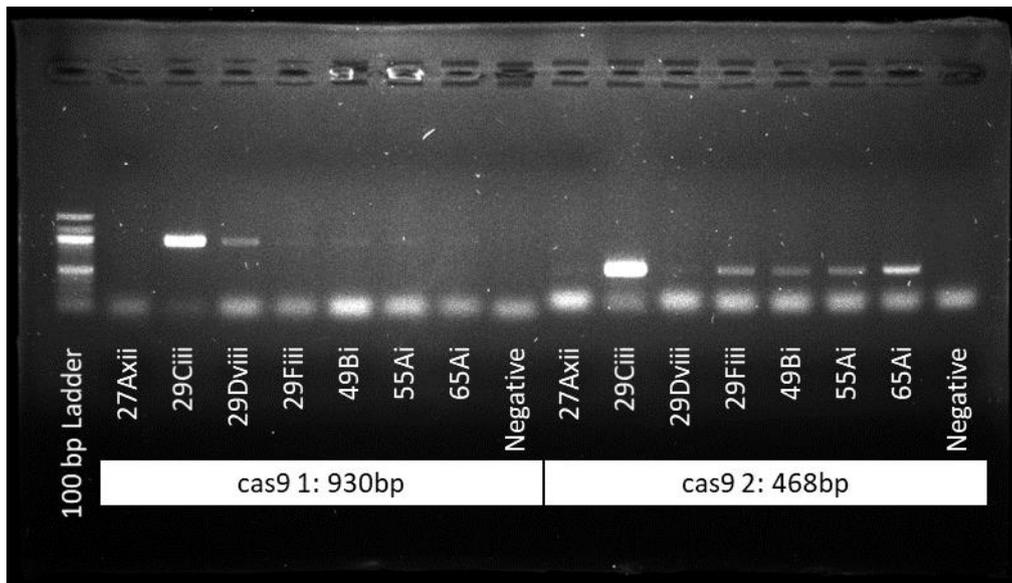


Figure 3.4: Reverse transcription PCR products of *cas9* mRNA expressed in ovaries of transgenic *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 shown on agarose gel. Expected product sizes of *cas9* mRNA are 930 and 468 base pairs.

3.3.2 Assessment of Cas9 function in *D. pseudoobscura* lines expressing endogenous *cas9* by targeting *white* and *yellow* marker genes for NHEJ mutation

Expression of the *cas9* gene in transgenic *D. pseudoobscura* had been confirmed by reverse transcription PCR, but correct translation and functioning of the Cas9 protein required further validation. Injection of a guide RNA expression vector targeting marker genes into *D. pseudoobscura* 3xP3-AmCyan-nos-Cas9 embryos was used to establish whether the Cas9 enzyme was functional. Two marker genes were selected, *GA27183* and *GA15665*, orthologous for *D. melanogaster white* and *yellow*, respectively. Both *white* and *yellow* were previously used to test mutation efficiency in the development of CRISPR gene editing in *D. melanogaster*, as the mutant phenotypes are visible in the F1 generation, and so can be easily screened (Bassett *et al.* 2013; Kondo and Ueda 2013; Ren *et al.* 2013; Port *et al.* 2014; Ren *et al.* 2014). Mutation of *white* in *D. pseudoobscura* was expected to result in a white eye phenotype, and *yellow* a paler yellow body phenotype.

A total of six different guide RNA expression vectors targeting different sites in the ORF of either *white* or *yellow* were injected into a total of four lines. Yellow 2 was injected into two lines, 29Dviii and 29Fiii. Yellow 1 was injected into two lines, 29Dviii and 27Axii. White 2 was injected into two lines, 29Fiii and 27Axii. PLR013-15, targeting three sites in the *white* ORF, was injected as a mix into four lines, 29Dviii, 29Fiii, 27Axii and 49Bi. 40 surviving F0 adults injected with gRNA expression vectors targeting *yellow*, and 28 surviving F0 adults injected with gRNA expression vectors targeting *white*,

were collected and crossed with *D. pseudoobscura* SLOB3 wild type. 13 *yellow* injected F0 and 2 *white* injected F0 produced no F1 progeny. Embryo survival after injection was very low (approximately 1%); as such, the number of surviving adults collected was lower than would have been ideal. Comparable studies of *D. melanogaster* have achieved survival rates through to the adult stage of at approximately 10% (Bassett *et al.* 2013; Gratz *et al.* 2014).

Both the GA27183 *white* and GA15665 *yellow* genes are located on the X chromosome, and therefore F1 male progeny of F0 females were screened for *white* or *yellow* mutant phenotypes. F1 female progeny of both male and female F0 injection survivors were collected and crossed with *D. pseudoobscura* SLOB3 wild type. F2 male progeny were collected and screened for *white* or *yellow* mutant phenotypes. No F1 male progeny of F0 females, nor F2 male progeny of F1 females, were found to have a *white* or *yellow* mutant phenotype, indicating that Cas9-mediated mutation of the marker genes was unsuccessful. Previous studies targeting *yellow* and *white* in *D. melanogaster* systems have demonstrated high rates of successful mutagenesis and transmission of mutations to progeny, upwards of 50%, with some studies reporting mutagenesis upwards of 70% (Bassett *et al.* 2013; Gratz *et al.* 2014; Port *et al.* 2014). With high efficiencies reported by others, it was expected that mutants would be observed in the validation experiments presented here, particularly in the case of the PLR013-015 gRNAs, which have been used to generate *white* mutants in *D. pseudoobscura* previously (Phadnis Lab, referenced in Schroeder *et al.* 2020). This suggests that there was an issue with the *D. pseudoobscura* Cas9 lines which resulted in a lack of successful mutagenesis.

3.3.3 High resolution melt analysis of F0 *white* and *yellow* gRNA injection survivors

Mutagenesis of the *white* and *yellow* genes was also assessed by high resolution melt analysis (HRMA). DNA was extracted from *D. pseudoobscura* F0 injection survivors and *white* and *yellow* gene regions were amplified by PCR. The PCR product was used as the template for HRMA. Melt curves were analysed and compared to a non-injected Cas9 control. In principle, HRMA can reveal whether indels have been generated at the sgRNA site in the injected animals, which would be a mosaic for such events. The presence of a mix of products after the PCR reaction is detected by the altered pattern of melting of heteroduplex DNA (Bassett *et al.* 2013; Bassett and Liu 2014). Analysis showed melt curves were highly variable and potential mutants could not be identified by this method (Figure 3.5).

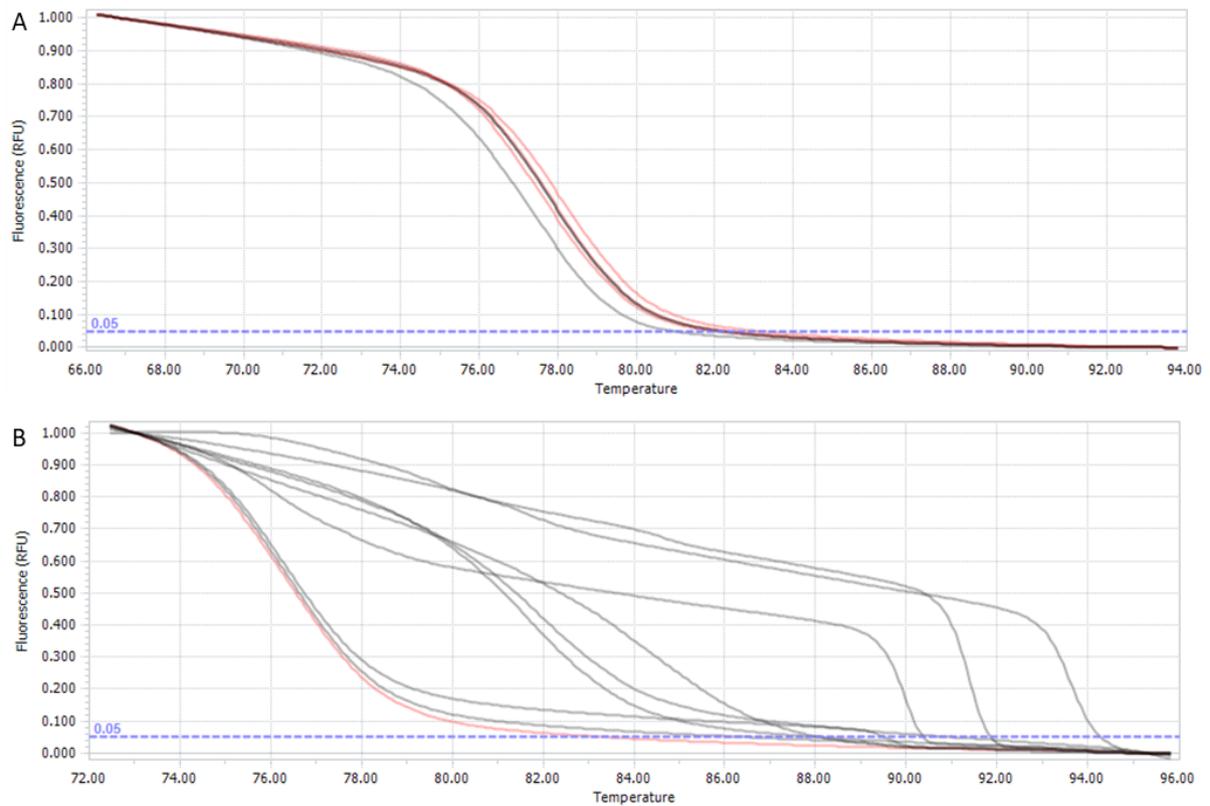


Figure 3.5: Examples of normalised melt curves from HRMA of *yellow* gene in *D. pseudoobscura* injected with *Yellow 1* gRNA. A: Black indicates 29Fiii female F0 adults (*Fi*, *Fiv*, *Fv*, *Fvi*), red indicates *D. pseudoobscura* non-injected controls (WT, 29Fiii Cas9 line, 29Dviii Cas9 line). B: Black indicates 29Dviii female and male F0 adults (*Fii-v*, *Mi-iv*, *Mvi*), red indicates *D. pseudoobscura* non-injected 29Dviii control.

gRNA	Line	F0 ID	Male/Female	F1 Females	F1 Males	Notes	F1 Females Crossed	F2 Transgenics	Notes
Yellow 2	29Dviii	Mi	M	92	93	-	Yes	No	-
Yellow 2	29Dviii	Mii	M	95	69	-	Yes	No	Three F1 crosses, no F2
Yellow 2	29Dviii	Miii	M	111	85	-	Yes	No	Two F1 crosses, no F2
Yellow 2	29Dviii	Miv	M	0	0	Eggs present, no larvae	-	-	-
Yellow 2	29Fiii	Mi	M	107	120	-	Yes	No	Two F1 crosses, no F2
Yellow 2	29Fiii	Mii	M	80	87	-	Yes	No	One F1 cross, no F2
Yellow 2	29Fiii	Miii	M	111	106	-	Yes	No	-
Yellow 2	29Fiii	Miv	M	96	68	-	Yes	No	-
Yellow 2	29Fiii	Mv	M	0	0	No eclosion	-	-	-
Yellow 2	29Fiii	Mvi	M	142	115	-	Yes	No	-
Yellow 2	29Fiii	Mvii	M	103	94	-	Yes	No	-
Yellow 2	29Fiii	Mviii	M	78	68	-	Yes	No	-
Yellow 2	29Fiii	Mix	M	53	49	-	Yes	No	One F1 cross, no F2
Yellow 2	29Fiii	Mx	M	90	99	-	Yes	No	One F1 cross, no F2
Yellow 2	29Fiii	Mxi	M	122	118	-	Yes	No	Two F1 crosses, no F2
Yellow 2	29Fiii	Mxii	M	94	103	-	Yes	No	Two F1 crosses, no F2
Yellow 2	29Dviii	Fi	F	0	0	No cross	-	-	-
Yellow 2	29Dviii	Fii	F	52	61	-	Yes	No	-

Yellow 2	29Dviii	Fiii	F	0	0	No cross	-	-	-
Yellow 2	29Dviii	Fiv	F	0	0	No eggs	-	-	-
Yellow 2	29Dviii	Fv	F	41	32	-	Yes	No	-
Yellow 2	29Fiii	Fi	F	0	0	No cross	-	-	-
Yellow 2	29Fiii	Fii	F	1	0	-	Yes	No	-
Yellow 2	29Fiii	Fiii	F	53	48	-	Yes	No	-
Yellow 2	29Fiii	Fiv	F	25	28	-	Yes	No	-
Yellow 2	29Fiii	Fv	F	0	0	No cross	-	-	-
Yellow 2	29Fiii	Fvi	F	47	39	-	Yes	No	-
Yellow 2	29Fiii	Fvii	F	0	0	No eggs	-	-	-
Yellow 2	29Fiii	Fviii	F	32	36	-	Yes	No	One F1 cross, no F2. All F1 Xs, low F2 numbers.
Yellow 2	29Fiii	Fix	F	0	0	No eggs	-	-	-
Yellow 2	29Dviii	Fvi	F	6	14	-	Yes	No	-
Yellow 2	29Dviii	Mv	M	0	0	No genitals, small testes	-	-	-
Yellow 2	29Fiii	Fx	F	0	0	No eclosion	-	-	-
Yellow 1	29Dviii	Fi	F	53	42	-	Yes	No	-
Yellow 1	29Dviii	Fii	F	50	47	-	Yes	No	One F1 cross, no F2
Yellow 1	29Dviii	Fiii	F	51	65	-	Yes	No	-
Yellow 1	27Axii	Fi	F	36	45	-	Yes	No	-

Yellow 1	27Axii	Mi	M	81	65	-	Yes	No	One F1 cross, no F2
Yellow 1	27Axii	Mii	M	0	0	No testes	-	-	-
Yellow 1	27Axii	Miii	M	0	0	No testes	-	-	-
White 2	29Fiii	Fi	F	10	14	Two F1 males, gonadless.	Yes	No	-
White 2	29Fiii	Mi	M	0	0	-	-	-	-
White 2	29Fiii	Fii	F	39	33	-	Yes	No	Two F2 crosses, no F2
White 2	29Fiii	Fiii	F	0	0	-	-	-	-
White 2	27Axii	Fi	F	48	45	-	Yes	No	-
White 2	27Axii	Fii	F	20	14	-	Yes	No	-
White 2	27Axii	Fiii	F	49	60	-	Yes	No	-
White 2	27Axii	Fiv	F	36	32	-	Yes	No	-
White 2	27Axii	Mi	M	59	78	-	Yes	No	One F1 cross, no F2
PLR013-15	49Bi	Fi	F	8	6	-	Yes	No	-
PLR013-15	27Axii	Fi	F	25	24	-	Yes	No	Three F1 crosses, no F2
PLR013-15	27Axii	Mi	M	6	8	-	Yes	No	One F1 cross, no F2
PLR013-15	27Axii	Mii	M	42	48	-	Yes	No	One F1 cross, no F2
PLR013-15	27Axii	Fii	F	4	6	-	Yes	No	-
PLR013-15	27Axii	Miii	M	64	56	-	Yes	No	-
PLR013-15	29Dviii	Mi	M	51	45	-	Yes	No	-
PLR013-15	29Dviii	Miii	M	58	56	-	Yes	No	-

PLR013-15	29Dviii	Miv	M	47	45	-	Yes	No	-
PLR013-15	29Dviii	Fii	F	9	12	-	Yes	No	-
PLR013-15	29Dviii	Mii	M	54	47	-	Yes	No	-
PLR013-15	29Dviii	Fiii	F	13	3	-	Yes	No	-
PLR013-15	27Axii	Fiii	F	12	13	-	Yes	No	-
PLR013-15	29Fiii	Fi	F	18	19	-	Yes	No	-
PLR013-15	29Fiii	Fii	F	20	16	-	Yes	No	-
PLR013-15	29Fiii	Miii	M	46	46	-	Yes	No	-
PLR013-15	29Fiii	Mii	M	41	35	-	Yes	No	-
PLR013-15	29Fiii	Fiii	F	6	1	-	Yes	No	-
PLR013-15	29Fiii	Mi	M	69	69	-	Yes	No	-

Table 3.4: F0 adult survivors of white and yellow gRNA injections. F0 adults crossed with D. pseudoobscura SLOB3 wild-type. F1 adults collected and males screened for transgenic phenotype (white or yellow). Female F1 adults collected and crossed with wild-type. F2 adults were collected and screened for transgenic phenotype.

4 Gene Expression Analysis of *D. pseudoobscura* Spermatogenesis

Drosophila spermatogenesis involves transcription and translational of thousands of genes, many of which are essential for sperm development. The primary spermatocytes are the most transcriptionally active cells in the adult fly, expressing genes required for spermatogenesis, meiosis and spermiogenesis (Li *et al.* 2022). Some transcription does also take place after meiosis, during the spermatid elongation stage, however it is far less than prior to meiosis (Olivieri and Olivieri 1965; Lindsley and Tokuyasu 1980; Fuller 1993; Li *et al.* 2022).

D. pseudoobscura spermatocyte cysts cannot be distinguished visually. Spermatids can be distinguished, based on length, although elongating euspermatids and elongated paraspermatids have a similar appearance. As the majority of transcription takes place prior to meiosis, there may be transcriptional variation between spermatocyte cysts in *D. pseudoobscura*, relating to the development of eusperm and parasperm morphs. Analysis of primary spermatocyte cyst gene expression by RNA-seq was hypothesised to show differential gene expression of genes required for sperm development. Candidate genes identified as differentially expressed could then be analysed further to examine their role in the development of the eusperm and parasperm morphs. Lack of transcriptional variation between spermatocyte cysts would indicate that the process of differential development between the sperm morphs takes place after the spermatocyte stage, for example during elongation.

Differential gene expression in *D. pseudoobscura* spermatocytes was investigated by RNA sequencing of individual spermatocyte cysts. Initially, ten spermatocyte cysts from a single male were sequenced. Subsequently, a further twenty spermatocyte cysts from two further males (ten per male) were sequenced.

Differential gene expression analysis of *D. pseudoobscura* spermatocyte cysts revealed transcriptional variation between spermatocyte cysts in genes expressed during spermatogenesis. Genes exhibiting differential expression in primary spermatocyte cysts may have roles in eusperm- or parasperm-specific development. Candidate genes identified by RNA-seq analysis were further investigated by RNA *in situ* hybridisation, to examine expression patterns in the testes, and to further narrow down candidate lists of genes which have a significant role in controlling differential development between the sperm morphs for further study.

Spermatocyte cyst identity was unknown prior to cluster analysis of RNA-seq data. Assignment to groups was based on cluster analysis, and an assumption of the number of possible groups based on eusperm and parasperm cyst morphs. Initial analysis was performed assuming two groups; eusperm and parasperm. Data was later reanalysed assuming three cyst groups; eusperm, parasperm 1, and parasperm 2.

4.1 RNA-Seq 1: Differential Gene Expression Analysis of *D. pseudoobscura* Spermatocyte Cysts (Clontech)

4.1.1 Sequence quality

Dissection, library preparation and sequencing was carried out by Helen White-Cooper. Analysis of the sequencing data was carried out as part of this project.

Ten spermatocyte cysts were dissected from a single male *D. pseudoobscura* SLOB3 wild type. RNA extraction and cDNA synthesis from individual spermatocytes was carried out using the SMART-seq Ultra Low Input Kit (Clontech) and libraries prepared using the Nextera CT Kit (Illumina). Libraries were then sequenced by HiSeq (Illumina). All ten cyst libraries were successfully sequenced to generate 75 base pair paired-end reads.

Raw sequence quality was high (mean Phred >20) with sequence lengths above 70 base-pairs. Sequences were trimmed to remove adaptor sequences and sequence regions of low quality (Phred <20), then filtered to remove reads less than 36 base pairs in length and of low quality (mean Phred <20 across 4bp). Paired-end read counts before and after trimming are summarised in Table 4.1.

Reads were aligned to the *D. pseudoobscura* reference genome version 3.04 (Gramates *et al.* 2017). The data were filtered to remove unmapped and duplicate reads. Aligned reads were counted to give a total count for each annotated gene per spermatocyte cyst. Unmapped reads were also counted. Count data are summarised in Table 4.2. A total of 15167 genes were represented in RNA sequencing data from ten primary spermatocyte cysts, of a total 16959 mapped genes in the *D. pseudoobscura* genome (Gramates *et al.* 2017).

Raw and normalised count data are available in Appendix 3.

Spermatocyte Cyst	Raw Data Paired-End	Filtered Paired-End	% Sequences Filtered
Sample ID	Reads	Reads	
A-I	24802522	15179051	38.8
A-II	26193771	17868611	31.8

A-III	24852745	17576436	29.3
A-IV	27715722	19686499	29.0
A-V	26978953	20107564	25.5
A-VI	26212608	18318641	30.1
A-VII	26207415	17707434	32.4
A-VIII	24854003	14596449	41.3
A-IX	20437753	11085563	45.8
A-X	26753897	17762476	33.6

Table 4.1: Paired-end read counts of raw *D. pseudoobscura* SLOB3 wild type single primary spermatocyte cyst RNA sequence data. Read counts are provided for data prior to and post trimming and filtering by Trimmomatic (Bolger et al. 2014). The percentage of sequences filtered was calculated.

Spermatocyte Cyst	Mapped Read Count	Unmapped Read Count	Genes Represented
Sample ID			
A-I	8018903	2024291	13035
A-II	10848369	2223182	11667
A-III	11821718	1898492	13004
A-IV	12229972	2596344	12805
A-V	13001775	2475162	13461
A-VI	12394516	1863215	13477
A-VII	10813885	2140710	12347
A-VIII	6836088	2069626	13482
A-IX	7039375	1265323	12681
A-X	11003019	2166960	12264
Total			15167

Table 4.2: Counts of mapped and unmapped reads from *D. pseudoobscura* SLOB3 wild type single spermatocyte cyst RNA sequencing. Reads were aligned to *D. pseudoobscura* reference genome. Unmapped reads include reads for which no feature is mapped, or for which alignment was ambiguous. Read counts performed by htseq (Anders et al. 2015). The total number of genes represented in each primary spermatocyte dataset is given.

4.1.2 Hierarchical cluster analysis of primary spermatocyte cyst transcription data

Count data were normalised by TMM. Cluster analysis of normalised gene counts was used to group cysts according to transcription profile. Hierarchical cluster analysis (HCA) was used to generate a cluster dendrogram, from which groups could be inferred.

Cluster dendrograms were variable dependent on the methods of calculating pairwise distances and clustering (Figure 4.1, Figure 4.2). HCA of Euclidean pairwise distances of normalised filtered data (TMM, CPM ≥ 2) showed cyst VIII separated from the remaining cysts. VII was also separated. The structure of the cluster containing cysts I-VI, IX and X was variable between clustering methods based on Euclidean pairwise distances (Figure 4.1; A-F).

HCA of the Maximum pairwise distances of normalised filtered data (TMM, CPM ≥ 2) showed a different clustering compared to Euclidean pairwise distances, with cysts I, II, VII and VIII in a single cluster, separate from the other cysts (Figure 4.1; G). Clustering based on Maximum pairwise distances was more robust when filtering of the normalised data was increased (CPM ≥ 3 , CPM ≥ 5), whereas filtering data had more of an impact on the clustering based on Euclidean pairwise distances (Figure 4.2). Interestingly, increased filtering of the data before HCA based on Euclidean pairwise distances resulted in dendrograms more similar to those observed with Maximum pairwise distances: I, II and VII formed a cluster (Figure 4.2; A and C), similar to the I, II, VII and VIII cluster (Figure 4.2; B, D and F).

As clustering was variable dependent on the methods applied, statistical analysis with the pvclust R package was used to examine the validity of the two most variable models: Ward D2 clustering based on Euclidean or Maximum pairwise distances (Suzuki and Shimodaira 2015).

Statistical analysis of the primary spermatocyte cyst clustering by HCA found that Ward D or Ward D2 clustering of Euclidean pairwise distances produced only one significant cluster, containing cysts III and IX (bootstrap = 10000, $p < 0.01$). Clustering by Ward D2 of Euclidean distances, calculated from data with higher filtering (CPM > 5) resulted in three significant clusters, containing six cysts. Group 1 contained III and IX (bootstrap 10000, $p < 0.01$), group 2 contained IV and X (bootstrap = 10000, $p = 0.03$), and group three contained I and II (bootstrap = 10000, $p = 0.02$).

By contrast, Ward D2 clustering of Maximum pairwise distances produced two significant clusters encompassing all of the cyst data (bootstrap = 10000, $p = 0.02$). Group 1 contained cysts III-VI, IX and X. Group 2 contained cysts I, II, VII and VIII.

Statistical analysis indicated that Ward D2 HCA based on Maximum pairwise distances was better able to sort the data, therefore this clustering was used in subsequent differential gene expression analysis (Figure 4.1; G, Figure 4.2; B).

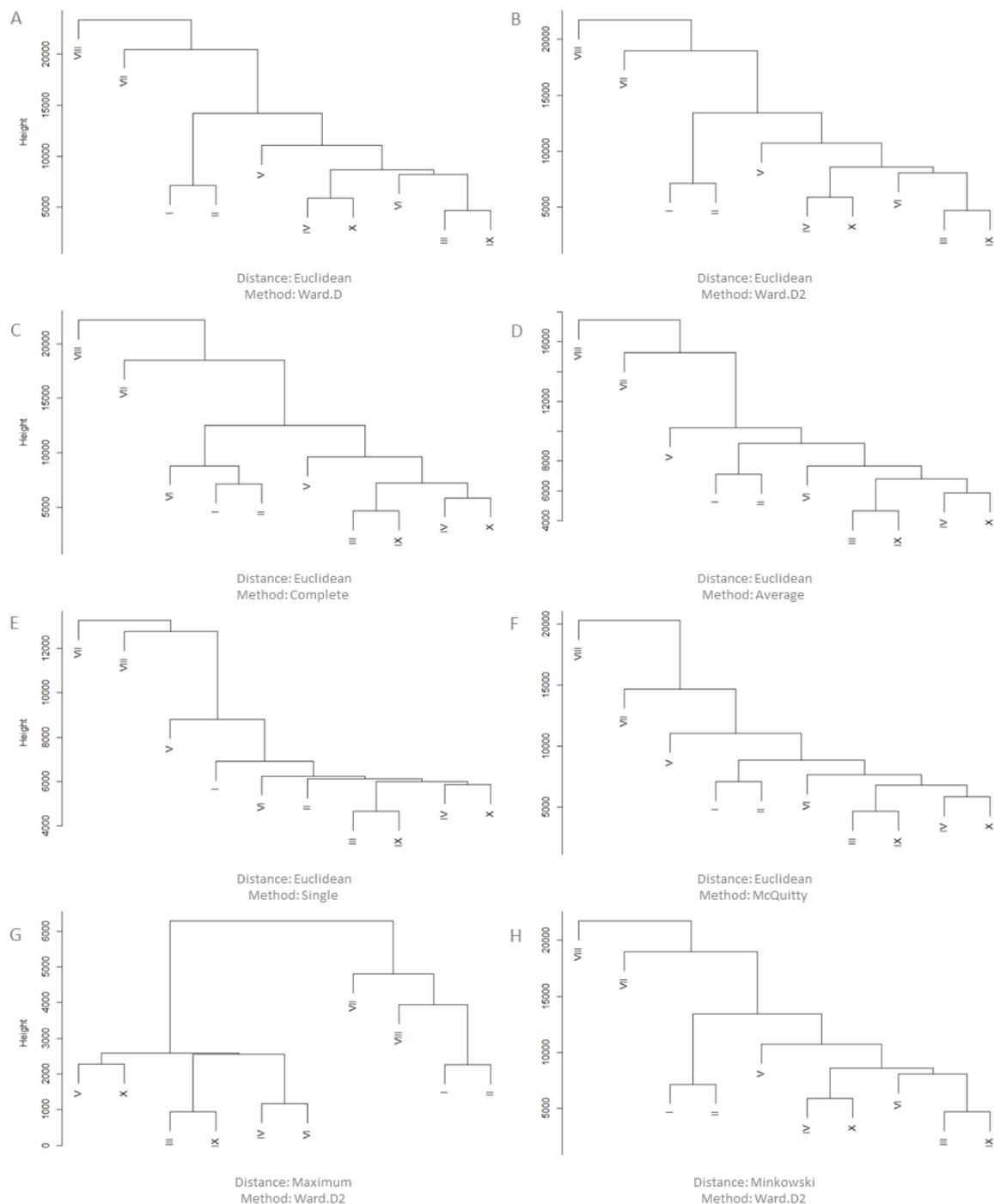


Figure 4.1: Hierarchical cluster analysis of *D. pseudoobscura* A primary spermatocyte cyst normalised transcription data. Normalisation by TMM. Filtered CPM ≥ 2 . A-F: Pairwise distance Euclidean. Clustering method, A: Ward D. B: Ward D2. C: Complete. D: Average. E: Single. F: McQuitty. G: Pairwise distance Maximum, clustering method Ward D2. H: Pairwise distance Minkowski, clustering method Ward D2.

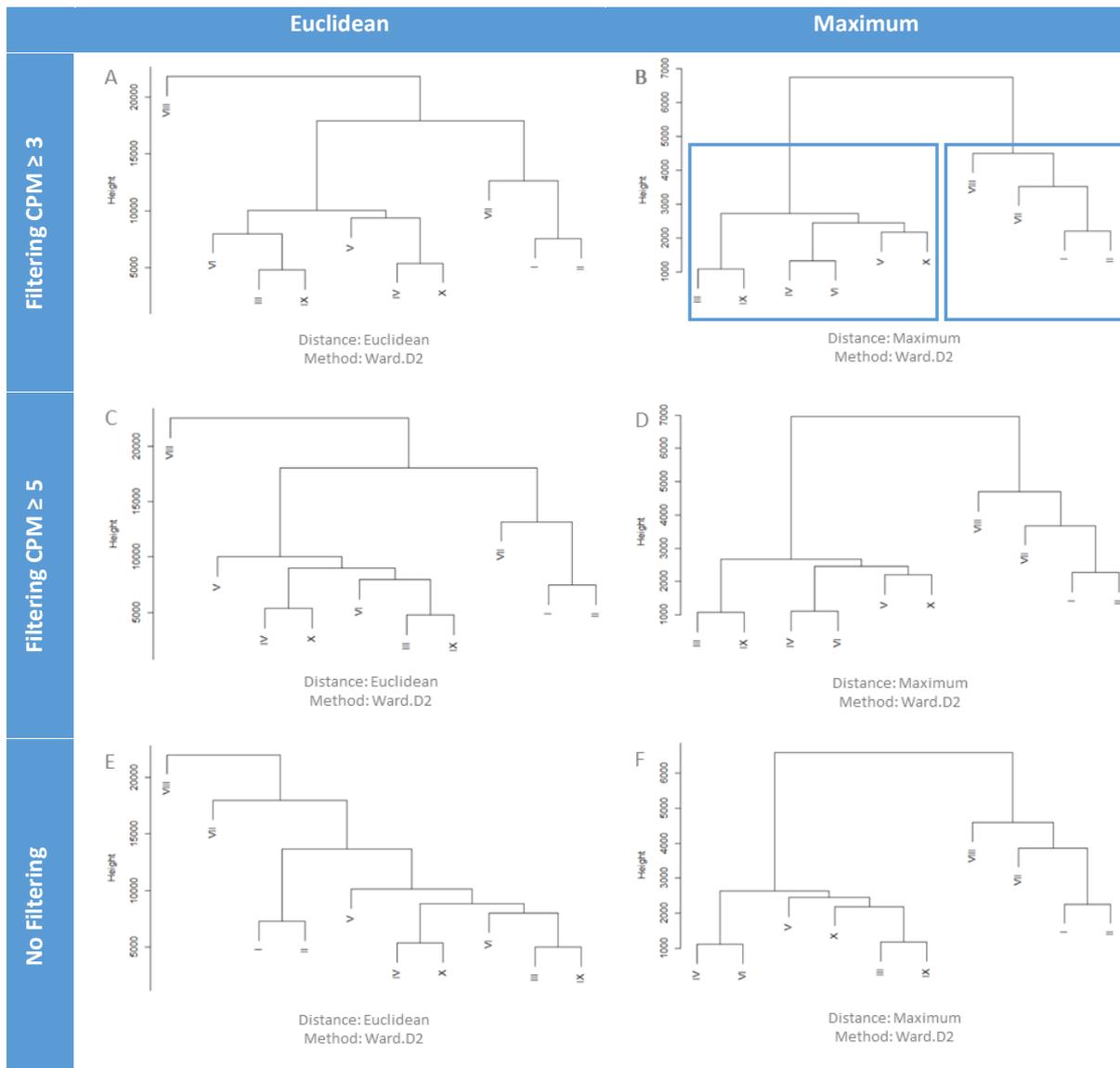


Figure 4.2: Hierarchical cluster analysis of *D. pseudoobscura* A primary spermatocyte cyst normalised transcription data. Normalisation by TMM. A, C, E: Pairwise distance Euclidean, clustering method Ward D2. A: Filtered CPM ≥ 3 . C: Filtered CPM ≥ 5 . E: No filtering. B, D, F: Pairwise distance Maximum, clustering method Ward D2. B: Filtered CPM ≥ 3 . Subsequent analysis based on clustering indicated by grey outline. D: Filtered CPM ≥ 5 . F: No filtering.

4.1.3 Differential gene expression analysis of spermatocyte cyst transcription data

DGE analysis was based on Ward D2 clustering of Maximum pairwise distances, which found two groups of cysts. Group 1 was defined as III-VI, IX and X. Group 2 was defined as I, II, VII and VIII (Figure 4.2; B).

1399 genes were found to be significantly differentially expressed by DGE analysis of two cyst groups ($p < 0.05$). The complete list of differentially expressed genes is available in Appendix 4. Differentially

expressed genes were examined by function of orthologous genes in *D. melanogaster*. Differentially expressed genes with orthologues known to function in spermatogenesis, spermiogenesis and transcription are summarised in Table 4.3. Genes with no known orthologue in *D. melanogaster* with testis-specific expression are also described in Table 4.3.

Group 2 could be further subdivided into Group 2.1, consisting of cyst VIII, and Group 2.2, consisting of cysts I, II and VII. 1528 genes were found to be significantly differentially expressed by DGE analysis of three cyst groups ($p < 0.05$). The complete list of differentially expressed genes is available in Appendix 4. Differentially expressed genes were examined by function of orthologous genes in *D. melanogaster*. Differentially expressed genes with functions in spermatogenesis, spermiogenesis and transcription are summarised in Table 4.4.

867 genes were found to be significantly differentially expressed by DGE analysis of both two and three cyst groups (Appendix 4).

<i>D. pseudoobscura</i> Gene Name	<i>D. melanogaster</i> Gene Name	Gene Function	p-Value	Highest Mean Normalised Count	Fold Change	
GA11638	wuc	tMAC	0.046419	Group 1	1.6	
GA20441	Trf	TFIID	0.019037	Group 1	1.6	
GA28988	Taf10		0.037246	Group 1	2.0	
GA19104	Hsf		0.001702	Group 2	1.8	
GA15016	REPTOR-BP	Transcription factor	0.011856	Group 1	2.3	
GA10314	CG10431		0.02294	Group 1	3.0	
GA18658	CG5098		0.025069	Group 2	1.5	
GA24806	CG3328		0.000972	Group 1	3.0	
GA11117	Abd-B		0.027575	Group 1	2.1	
GA26624	acj6		0.02259	Group 2	4.2	
GA10208	D19B		0.030317	Group 2	1.6	
GA28518	RAF2		0.014325	Group 2	1.9	
GA26380	CG32856		0.049712	Group 1	2.2	
GA11313	tgo		1.96E-06	Group 2	6.5	
GA19946	Cdc5		0.043617	Group 2	2.3	
GA11485	mei-P26		Spermatogenesis	0.003981	Group 2	1.7

<i>exu2</i> (GA28347)	<i>exu2</i>	Spermatogenesis, spermiogenesis	0.0002	Group 2	2.0
GA18272	<i>Nup154</i>		0.002036	Group 2	2.2
GA18412	<i>bol</i>		0.000313	Group 2	1.8
GA19706	<i>spag4</i>		0.019338	Group 2	1.6
GA25911	<i>Mst36Fb</i> <i>CG43339</i> <i>Mst36Fa</i>		0.018535	Group 2	1.7
GA27003	<i>mil</i>		0.049593	Group 2	1.4
GA16511	<i>Cul3</i>		0.027697	Group 2	1.5
GA17195	<i>pcm</i>		0.000315	Group 2	3.3
GA17729	<i>Lasp</i>		0.004368	Group 2	1.8
GA17771	<i>Grip84</i>		0.00898	Group 1	3.4
GA18038	<i>Syx5</i>		0.049448	Group 2	1.4
GA20593	<i>mtsh</i>		0.019871	Group 2	1.6
GA24628	<i>eIF4E3</i>		0.024788	Group 2	1.4
GA25980	<i>Cdlc2</i> <i>ctp</i>		4.76E-05	Group 2	2.3
GA28638	<i>Rcd7</i>		0.030932	Group 2	1.7
GA28949	<i>ACXE</i> <i>ACXA</i> <i>ACXB</i> <i>ACXC</i>		0.001854	Group 2	1.7
<i>Hsp83</i> (GA11622)	<i>Hsp83</i>	Spermiogenesis	0.002326	Group 2	1.7
GA21728	<i>betaTub85D</i>		0.034929	Group 2	1.4
GA10082	<i>wa-cup</i>	Post meiotic transcription, localised RNA	0.016534	Group 2	1.5
GA11961	<i>p-cup</i>		0.018359	Group 1	1.7
GA14633	<i>schuy</i>		0.023658	Group 2	1.7
GA11266	<i>m-cup</i>		0.029362	Group 2	1.4
GA19739	<i>ana1</i>	Axoneme assembly	0.002222	Group 2	1.9
GA14315	<i>Cep135</i>		0.016974	Group 2	1.7
GA27380	<i>Kap3</i>		0.039128	Group 2	2.7

GA12268	<i>Cby</i>		0.001521	Group 1	1.9
GA19437	<i>Fmr1</i>		0.01346	Group 2	1.8
GA20166	<i>SAK</i>		0.004547	Group 2	1.8
GA26457	<i>asl</i>	Axoneme assembly, cell cycle	0.000162	Group 2	2.1
GA22018	<i>tacc</i>	Cell cycle	0.002302	Group 2	1.7
GA11545	<i>polo</i>		0.038427	Group 2	1.7
GA21186	<i>Klp3A</i>		0.040311	Group 2	2.1
GA18583	<i>mmps</i>		0.007921	Group 2	1.6
GA23025	*	Testis-specific expression, no known <i>D. melanogaster</i> orthologue	0.04993	Group 2	1.5
GA28004	-		0.02254	Group 2	1.6
GA26318	-		0.005908	Group 2	1.7

Table 4.3: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq A based on hierarchical cluster analysis clustering into two cyst groups. Group 1: I, II, VII, VIII. Group 2: III-VI, IX X. *Orthologue of milkah (*mil*) previously supported, model updated and previous orthology no longer supported (Gramates et al. 2017; Larkin et al. 2021).

<i>D. pseudoobscura</i> Gene Name	<i>D. melanogaster</i> Gene Name	Gene Function	p-Value	Highest Mean Normalised Count	Fold Change
GA14575	<i>Taf1</i>	TFIID	0.040895	Group 2.1	3.2
GA16761	<i>Taf6</i>		0.035269	Group 1	1.4
GA25290	<i>can</i>	TFIID, spermatogenesis	0.020819	Group 2.1	2.6
GA14764	<i>Trf2</i>	TFIID, transcription factor	0.031868	Group 2.1	2.3
GA19104	<i>Hsf</i>	Transcription factor	0.000934	Group 1	1.7
GA18658	<i>CG5098</i>		0.027242	Group 1	1.5
GA24806	<i>CG3328</i>		0.000526	Group 2.2	2.1
GA11117	<i>Abd-B</i>		0.007338	Group 2.2	2.0

GA26624	<i>acj6</i>		0.008029	Group 1	2.5
GA28518	<i>RAF2</i>		0.000369	Group 1	1.6
GA11313	<i>tgo</i>		3.52E-05	Group 1	8.3
GA19946	<i>Cdc5</i>		0.013122	Group 1	1.9
GA30460	<i>Camta</i>		0.049984	Group 2.1	1.8
GA26453	<i>abd-A</i>		0.031071	Group 1	1.3
GA21024	<i>cg</i>		0.026698	Group 2.1	2.8
GA20511	<i>crp</i>		0.013991	Group 2.1	2.6
GA21689	<i>Hnf4</i>		0.030431	Group 2.1	2.2
GA20695	<i>Nulp1</i>		0.028754	Group 1	1.3
GA11041	<i>Pbp49</i>		0.005035	Group 2.1	2.4
GA11485	<i>mei-P26</i>		0.000255	Group 1	1.5
<i>Brca2</i> (GA15693)	<i>Brca2</i>	Spermatogenesis	0.049537	Group 2.1	1.3
<i>exu2</i> (GA28347)	<i>exu2</i>		0.000962	Group 1	2.2
GA18272	<i>Nup154</i>		0.009899	Group 1	2.7
GA18412	<i>bol</i>		0.00023	Group 1	1.8
GA17195	<i>pcm</i>		0.002573	Group 1	4.0
GA17729	<i>Lasp</i>		0.017523	Group 1	2.1
GA20593	<i>mtsh</i>		0.005673	Group 1	1.5
GA24628	<i>eIF4E3</i>		0.027783	Group 1	1.4
GA25980	<i>Cdlc2</i> <i>ctp</i>	Spermatogenesis, spermiogenesis	0.000308	Group 1	2.6
GA28004	-		0.038883	Group 1	1.6
GA28949	ACXE ACXA ACXB ACXC		0.002015	Group 1	1.7
GA10094	<i>tra2</i>		0.011277	Group 2.1	1.3
GA12841	<i>Mer</i>		0.040034	Group 2.1	1.7
GA15695	<i>bgcn</i>		0.021513	Group 2.1	2.7
<i>hyd</i> (GA27318)	-		0.004494	Group 2.1	1.7
GA23030	-		0.039393	Group 1	1.4

GA26552	<i>eIF4G2</i>		0.009	Group 1	1.3
<i>Hsp83</i> (GA11622)	<i>Hsp83</i>	Spermiogenesis	3.50E-05	Group 1	1.5
GA21728	<i>betaTub85D</i>		0.047308	Group 1	1.5
GA10082	<i>wa-cup</i>		0.027478	Group 1	1.6
GA11961	<i>p-cup</i>	Post meiotic transcription, localised RNA	0.020846	Group 2.1	1.3
GA11266	<i>m-cup</i>		0.009026	Group 1	1.3
GA15791	<i>hubl</i>		0.049183	Group 2.2	2.2
GA23820	<i>t-cup</i>		0.047622	Group 2.2	1.6
GA19739	<i>ana1</i>		0.00219	Group 1	1.8
GA14315	<i>Cep135</i>	Axoneme assembly	0.024196	Group 1	1.7
GA27380	<i>Kap3</i>		0.000503	Group 2.1	1.9
GA12268	<i>Cby</i>		0.007285	Group 2.1	1.8
GA19437	<i>Fmr1</i>		0.016563	Group 1	1.8
GA20166	<i>SAK</i>		0.004168	Group 1	1.7
GA26457	<i>asl</i>		Axoneme assembly, cell cycle	4.42E-06	Group 1
GA22018	<i>tacc</i>	Cell cycle	0.000765	Group 1	1.6
GA11545	<i>polo</i>		0.023988	Group 1	1.6
GA21186	<i>Klp3A</i>		0.011677	Group 1	1.7
GA18583	<i>mmps</i>		0.002972	Group 1	1.5
GA26318	-		0.01596	Group 1	1.8

Table 4.4: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq A based on hierarchical cluster analysis clustering into three cyst groups. Group 1: III-VI, IX, X. Group 2.1: VIII. Group 2.2: I, II, VII.

4.2 RNA-Seq 2: Differential Gene Expression Analysis of *D. pseudoobscura* Spermatocyte Cysts

4.2.1 Spermatocyte cyst dissection for cDNA library preparation

Ten spermatocyte cysts were dissected from each of two male *D. pseudoobscura* SLOB3 wild type. Cysts were imaged prior to collection and RNA extraction (Figure 4.3). Imaging of spermatocyte cysts prior to sequencing show that while all collected samples appeared to be primary spermatocyte cysts,

there was some size variation. This is likely due to variation in spermatocyte cyst age, as spermatocytes grow substantially during the 32-cell stage (Fuller 1993).

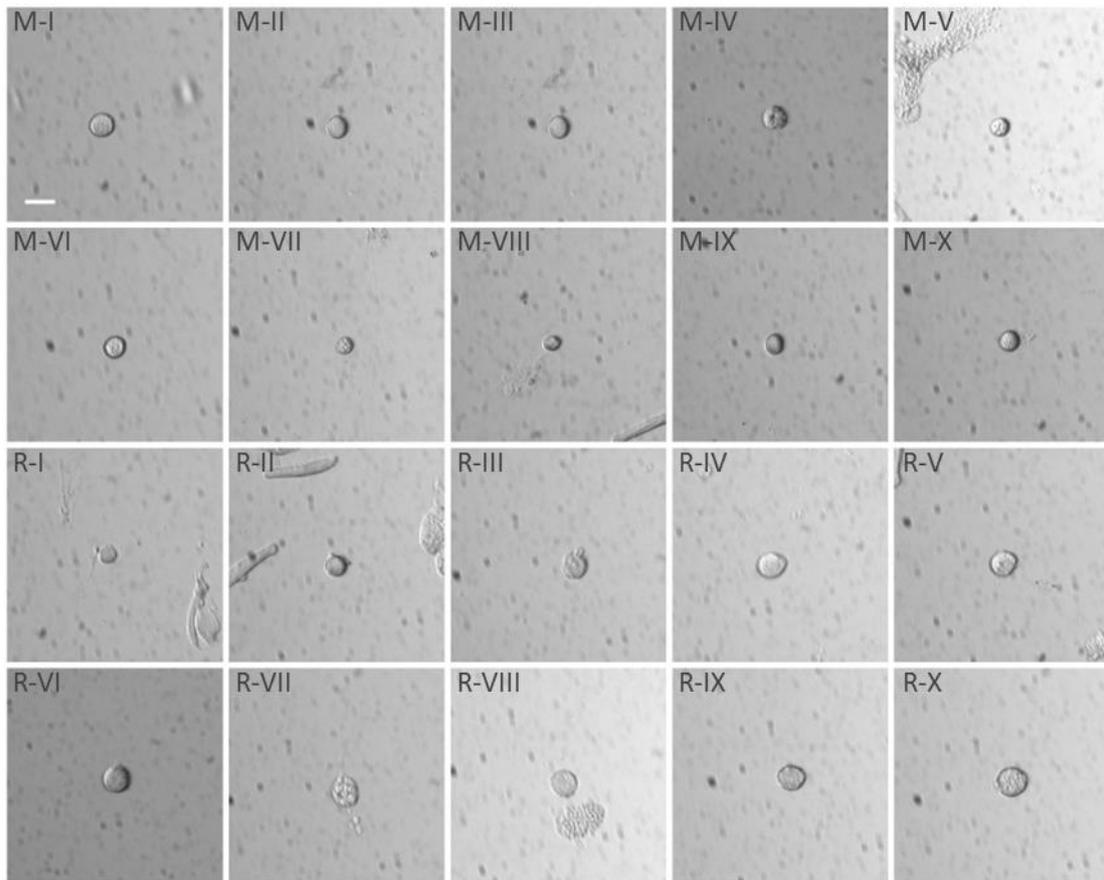


Figure 4.3: Primary spermatocyte cysts dissected from two *D. pseudoobscura* *SLOB3* wild-type males. Cysts were collected into individual tubes and RNA extracted for preparation of cDNA libraries. Scale = 20 μ m. Primary spermatocyte cysts were dissected from males M and R. Dust in optical system shows as background in cyst images. R-VIII: Only intact cyst was collected.

4.2.2 Sequence quality

RNA extraction, cDNA and library synthesis from individual spermatocytes was carried out using the QIAseq FX Single Cell RNA Library kit (Qiagen). Libraries were then sequenced by HiSeq (Illumina). All twenty cyst libraries were successfully sequenced to generate 75 base pair paired-end reads.

Raw sequence quality was high (mean Phred >20) with sequence lengths above 70 base-pairs. Sequences were trimmed to remove adaptor sequences and sequence regions of low quality (Phred <20), then filtered to remove reads less than 36 base pairs in length and of low quality (mean Phred <20 across 4bp). Paired-end read counts before and after trimming are summarised in Table 4.5.

Reads were aligned to the *D. pseudoobscura* reference genome version 3.04 (Gramates *et al.* 2017). The data were filtered to remove unmapped and duplicate reads. Aligned reads were counted to give a total count for each annotated gene per spermatocyte cyst. Unmapped reads were also counted. Count data are summarised in Table 4.6. A total of 14120 genes were represented in RNA sequencing data from the ten ‘male M’ primary spermatocyte cysts. 15624 genes were represented in RNA sequencing data from the ten ‘male R’ primary spermatocyte cysts.

There was a substantial decrease in the number of mapped reads compared to the number of paired-end reads after filtering. This may have been due to the removal of reads after alignment, where one read of the pairs was unmapped, where the read was an optical duplicate, or where the alignment of the read to the reference genome was below the designated quality cut-off (MAPQ = 10). This reduction in read counts and variability of read counts between cyst samples resulting from filtering may have impacted the subsequent cluster analysis.

Raw and normalised count data are available in Appendix 5.

Sample	Raw Data Paired-End Reads	Filtered Paired-End Reads	% Sequences Filtered
M-I	4251167	3743696	11.94
M-II	3947312	3562495	9.75
M-III	4077629	3593355	11.88
M-IV	4083993	3681126	9.86
M-V	4743126	4131090	12.90
M-VI	4714146	4327306	8.21
M-VII	4939250	4489359	9.11
M-VIII	4237165	3941442	6.98
M-IX	5302219	4838803	8.74
M-X	5452067	4927608	9.62
R-I	6136566	6057750	1.28
R-II	5692396	5201313	8.63
R-III	4933356	4683761	5.06
R-IV	4006873	3646782	8.99
R-V	4812480	4487290	6.76
R-VI	5327584	5004531	6.06
R-VII	5011866	4629391	7.63
R-VIII	5428161	4917172	9.41

R-IX	5684472	4989530	12.23
R-X	5217420	4488830	13.96

Table 4.5: Stage two RNA-seq analysis of primary spermatocyte cysts. Paired-end read counts of raw D. pseudoobscura SLOB3 wild type single primary spermatocyte cyst RNA sequence data of 20 spermatocyte cysts dissected from two males (M and R). Read counts are provided for data prior to and post trimming and filtering by Trimmomatic (Bolger et al. 2014). The percentage of sequences filtered was calculated.

Sample	Mapped Read Count	Unmapped Read Count	Genes Represented
M-I	948317	324033	9810
M-II	1142662	328460	9586
M-III	380009	261076	8249
M-IV	940849	350224	9674
M-V	810436	337909	9659
M-VI	634643	378779	7877
M-VII	498841	149025	6353
M-VIII	196747	105307	3407
M-IX	54705	28653	2308
M-X	345435	131211	4193
M Total			14120
R-I	23671	3644	1447
R-II	375854	100176	7456
R-III	1030788	901362	14755
R-IV	1063488	320358	9912
R-V	1189470	289828	8503
R-VI	542231	471238	9884
R-VII	650872	87552	7766
R-VIII	863518	193298	7201
R-IX	1386217	325745	8871
R-X	1323210	388655	9472
R Total			15624

Table 4.6: Counts of mapped and unmapped reads from D. pseudoobscura SLOB3 wild type single spermatocyte cyst RNA sequencing. Reads were aligned to D. pseudoobscura reference genome. Unmapped reads include reads for which no feature is mapped, or for which alignment was ambiguous. Read counts performed by htseq (Anders et al. 2015). The total number of genes represented in each primary spermatocyte dataset is given.

4.2.3 Hierarchical cluster analysis of primary spermatocyte cyst transcription data

Count data were normalised by TMM. Cluster analysis of normalised gene counts was used to group cysts according to transcription profile. Hierarchical cluster analysis (HCA) was used to generate a cluster dendrogram, from which groups could be inferred.

Multiple methods for clustering and calculating pairwise distances were applied to examine robustness of primary spermatocyte cyst clustering based on transcription. Figure 4.4 shows HCA dendrograms of *D. pseudoobscura* M primary spermatocyte cyst normalised RNA-seq data, with varied clustering methods. Cysts clustered into two main groups, consisting of I-VI, VIII and IX in group 1, and VII and X in group 2. Cysts within group 1 were more similar to each other than VII and X in group 2 were to each other. The sub clustering within group 1 was more variable between clustering method, although multiple clustering methods showed clustering of cysts VIII, II and VI and clustering of cysts I, III, IV, V and IX.

Methods for calculation of pairwise distances also did not modify the two main clusters (Figure 4.5). Euclidean and Minkowski pairwise distances resulted in identical dendrograms. There was some variation in group 1 sub-clustering with Maximum pairwise distances. Figure 4.5 D-F show dendrograms of Euclidean pairwise distances and the Ward D2 clustering method, with variation in filtering of normalised transcription data. Neither increasing nor removal of filtering resulted in changes to clustering by HCA.

Size of cyst did not appear to influence clustering, as cysts VII and X were consistently clustered by all clustering and pairwise distance methods, despite cysts VII and VIII appearing similar in size, and smaller in comparison to the other cysts dissected from M.

HCA dendrograms showed clustering of *D. pseudoobscura* R primary spermatocyte cyst normalised RNA-seq data, with varied clustering methods (Figure 4.6). Cysts cluster into two main groups, consisting of VII in group 1, the remaining cysts in group 2. Alternative methods of calculating pairwise distances show similar clustering of VII and clustering I-VI and VIII-X (Figure 4.7). Clustering within group 2 was more variable between clustering methods, dendrograms showed I and II clustered outside of III-VI and VIII-X, indicating that transcription was less similar in these cysts compared to the other cysts within group 2 (Figure 4.6).

Increasing filtering to $\text{CPM} \geq 3$ or removing filtering did not alter clustering, although increasing filtering to $\text{CPM} \geq 5$ altered the clustering in group 2 (Figure 4.7).

Primary spermatocyte cyst R VII (Figure 4.3; Q) was separately clustered by all HCA methods. Similar to clustering of M spermatocyte cysts, clustering was not based on size. Sub-clustering within group 2 may have been influenced by cyst growth, as cysts R I and II were clustered outside of the remaining group 2 cluster; both I and II appeared smaller in size compared to other spermatocyte cysts (Figure 4.3; K and L).

Combining M and R spermatocyte cyst data sets for hierarchical cluster analysis was used to examine whether clustering based on transcription profile was replicated across the two samples (Figure 4.8). Combined HCA showed clustering into two main groups; group 1 consisting of cysts MVII, MX and RVII, and group 2 consisting of MI-MVI, MVIII-MIX, RI-RVI and RVII-RX. Previous analysis of the datasets had clustered MVII and MX, and RVII in separate from the other cysts in each respective dataset (Figure 4.4, Figure 4.6). Clustering of these cysts when combining the datasets indicates that MVII, MX and RVII are more transcriptionally similar to each other than to other cysts from the same data set.

Within the Group 2 cluster, two sub-groups could be defined. Group 2.1 contained cysts MII, MVI and MVIII, and group 2.2 contained MI, MIII, MIV, MV, MIX, RI, RII, RIII, RIV, RV, RVI, RVIII, RIX and RX (Figure 4.8; C). Previous analysis had shown MII, MVI and MVIII clustered, which appeared to be maintained when the datasets were combined. Within group 2.2, clustering did not appear similar to that observed in separate analysis of the M and R cyst data (Figure 4.4, Figure 4.6).

Statistical analysis of Ward D2 and clustering of Euclidean and Maximum pairwise distances showed the Group 1 and Group 2 clusters were significant (bootstrap = 10000, Euclidean $p = 0.03$, Maximum $p = 0.03$).

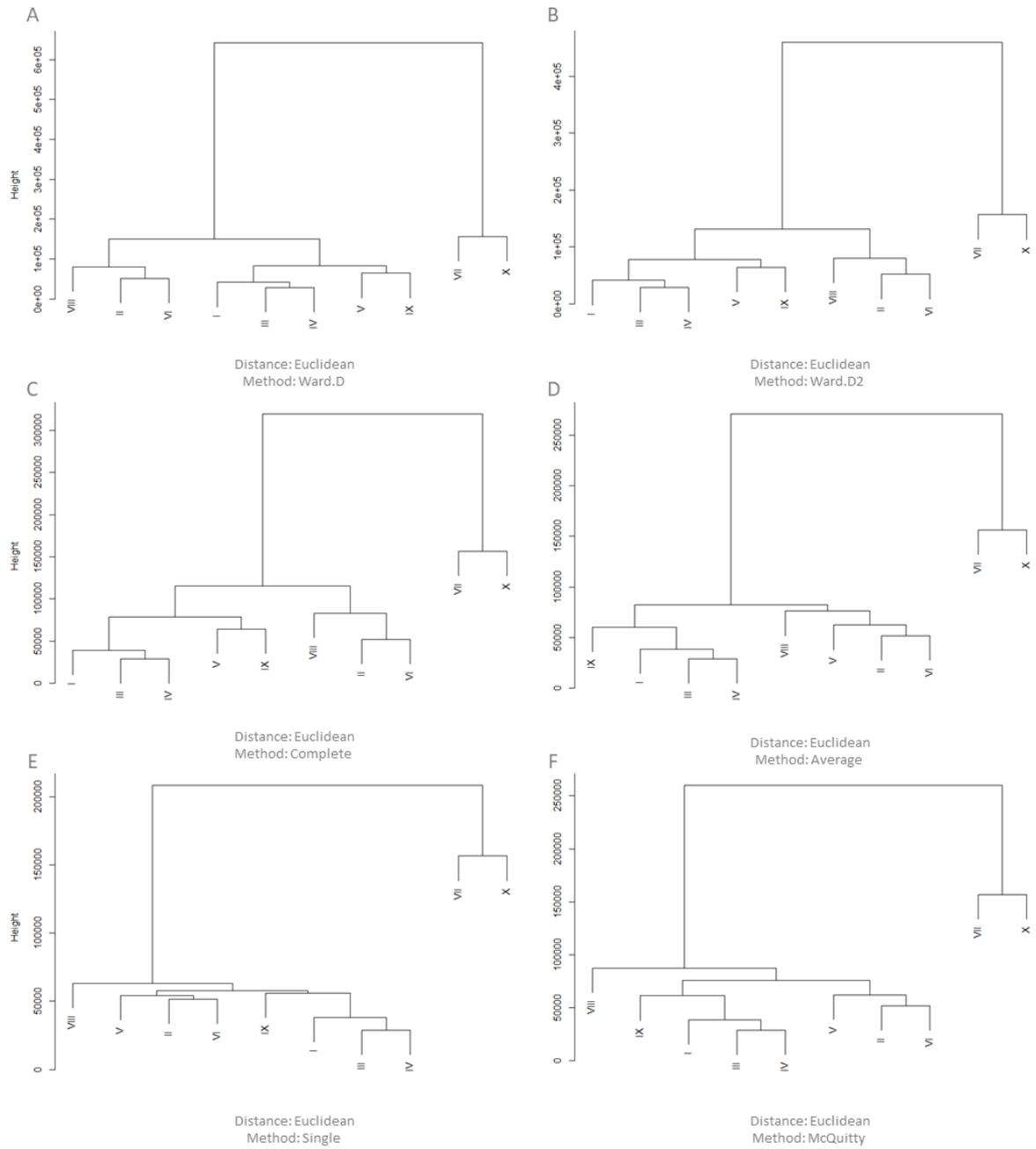


Figure 4.4: Hierarchical cluster analysis of *D. pseudoobscura* M primary spermatocyte cyst normalised transcription data. Normalisation by TMM. Filtered CPM ≥ 2 . Pairwise distance Euclidean. Clustering method, A: Ward D. B: Ward D2. C: Complete. D: Average. E: Single. F: McQuitty.

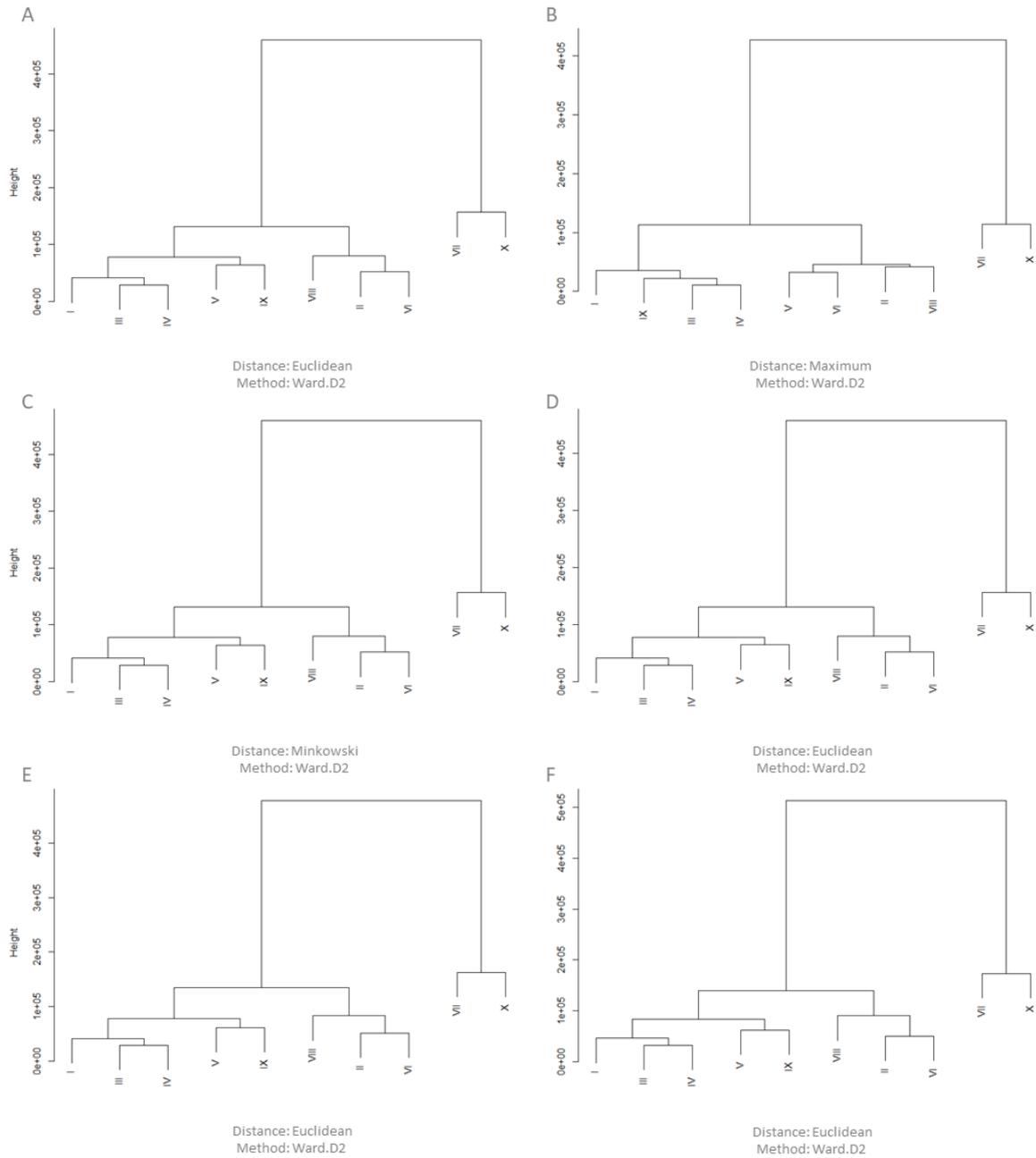


Figure 4.5: *D. pseudoobscura* M primary spermatocyte cyst normalised transcription data hierarchical cluster analysis. Normalisation by TMM. A-C: Filtered CPM ≥ 2 . A: Pairwise distance Euclidean, clustering method Ward D2. B: Pairwise distance Maximum, clustering method Ward D2. C: Pairwise distance Minkowski, clustering method Ward D2. D-F: Pairwise distance Euclidean, clustering method Ward D2. D: Filtered CPM ≥ 3 . E: Filtered CPM ≥ 5 . F: No filtering.

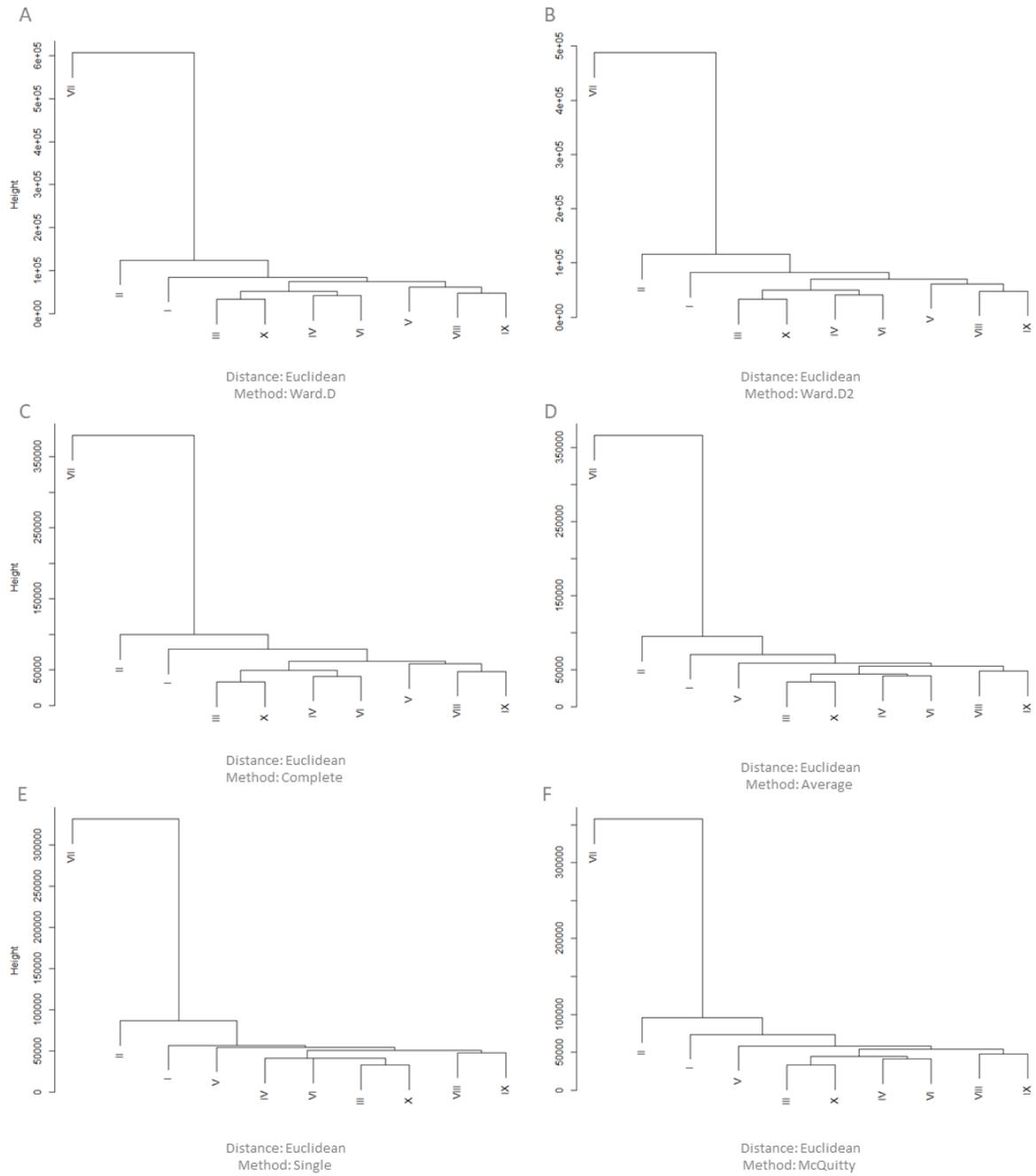


Figure 4.6: *D. pseudoobscura* R primary spermatocyte cyst normalised transcription data hierarchical cluster analysis. Normalisation by TMM. Filtered CPM ≥ 2 . Pairwise distance Euclidean. Clustering method, A: Ward D. B: Ward D2. C: Complete. D: Average. E: Single. F: McQuitty.

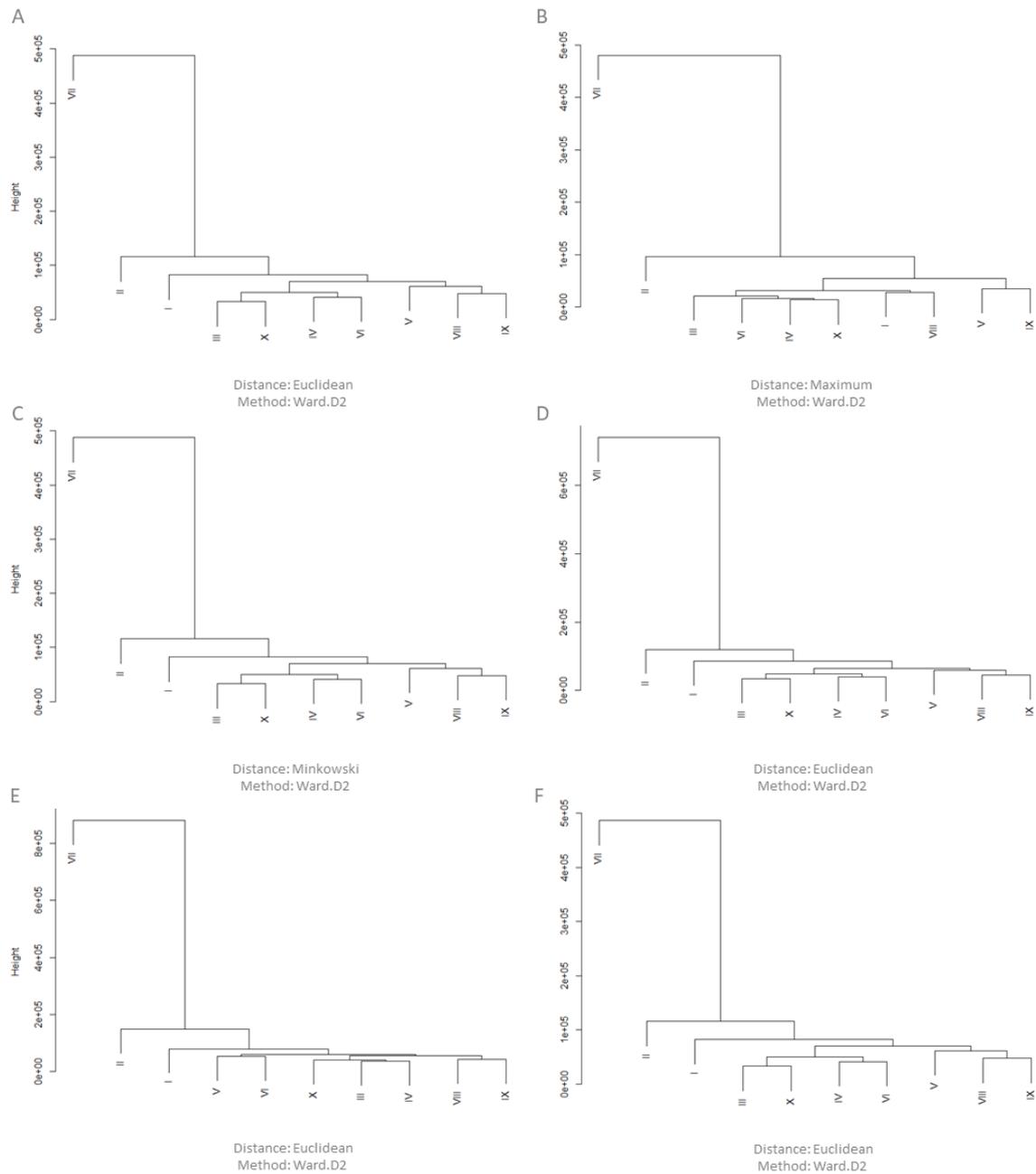


Figure 4.7: *D. pseudoobscura* R primary spermatocyte cyst normalised transcription data hierarchical cluster analysis. Normalisation by TMM. A-C: Filtered CPM ≥ 2 . A: Pairwise distance Euclidean, clustering method Ward D2. B: Pairwise distance Maximum, clustering method Ward D2. C: Pairwise distance Minkowski, clustering method Ward D2. D-F: Pairwise distance Euclidean, clustering method Ward D2. D: Filtered CPM ≥ 3 . E: Filtered CPM ≥ 5 . F: No filtering.

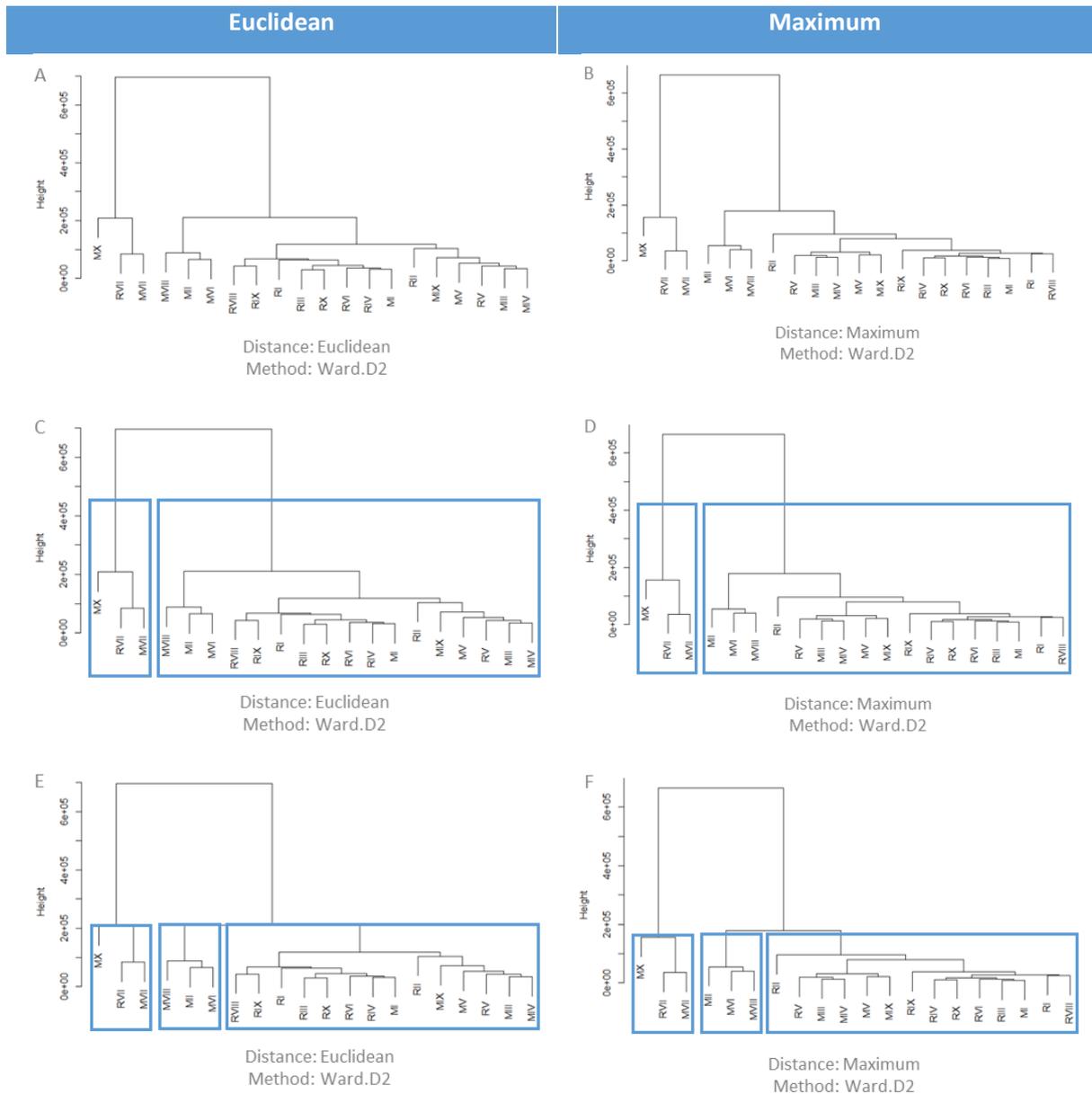


Figure 4.8: Hierarchical cluster analysis of combined M and R primary spermatocyte cyst normalised transcription data. Normalisation by TMM. Filtered CPM ≥ 2 . Clustering method Ward D2. A-C: Pairwise distances Euclidean. A: Cluster dendrogram of normalised M and R primary spermatocyte cyst RNA-seq data. B: Tree cut into two clusters. C: Tree cut into three clusters. D-F: Pairwise distances Maximum. D: Cluster dendrogram of normalised M and R primary spermatocyte cyst RNA-seq data. E: Tree cut into two clusters. F: Tree cut into three clusters.

4.2.4 Differential gene expression analysis of primary spermatocyte cysts

Hierarchical cluster analysis of primary spermatocytes was used to define groups for differential gene expression (DGE) analysis. The number of groups defined was dependent on the potential number of primary spermatocyte cyst morphs. DGE analysis was performed twice; based on two cyst groups, and based on three cyst groups (Alpern *et al.* 2019).

1613 genes were found to be significantly differentially expressed by DGE analysis of two cyst groups ($p < 0.05$). The complete list of differentially expressed genes is available in Appendix 6. Differentially expressed genes were examined by function of orthologous genes in *D. melanogaster*. Differentially expressed genes with functions in spermatogenesis, spermiogenesis and transcription are summarised in Table 4.7.

747 genes were found to be significantly differentially expressed by DGE analysis of three cyst groups ($p < 0.05$). The complete list of differentially expressed genes is available in Appendix 6. Differentially expressed genes with functions in spermatogenesis, spermiogenesis and transcription are summarised in Table 4.8.

433 genes were found to be significantly differentially expressed by DGE analysis of both two and three cyst groups (Appendix 6).

<i>D. pseudoobscura</i> Gene Name	<i>D. melanogaster</i> Gene Name	Gene Function	p-value	Highest Mean Normalised Count	Fold Change
GA13507	<i>mip40</i>	tMAC,	0.010108	Group 1	6.9
GA21108	<i>topi</i>	spermatogenesis, spermiogenesis	0.008682	Group 2	28.5
GA21613	<i>gas41</i>	TFIID, spermatogenesis, spermiogenesis	0.041388	Group 1	4.9
GA19104	<i>Hsf</i>	Transcription Factor	0.00932	Group 2	64.4
GA17585	<i>stc</i>		0.037311	Group 1	5.4
GA22634	<i>ovo</i>		0.000503	Group 1	27.6
GA19264	<i>woc</i>		0.002863	Group 1	10.8
GA18636	<i>croc</i>		0.009234	Group 1	20.4
GA11197	<i>Hr96</i>		0.007574	Group 2	47.0
GA28498	<i>Pdp1</i>		0.048497	Group 1	6.7
GA18732	<i>Dad</i>		0.010695	Group 1	8.1
GA17066	<i>phtf</i>		4.16E-05	Group 1	11.6
GA15016	<i>REPTOR-BP</i>		0.032794	Group 1	7.4
GA10314	<i>CG10431</i>		0.020162	Group 1	11.1
GA14905	<i>Mondo</i>		0.018462	Group 1	5.1

GA18658	CG5098		0.025042	Group 2	21.5
GA24806	CG3328		0.004961	Group 1	8.1
GA10210	CG10274		0.004164	Group 2	69.0
GA21903	<i>Pbp45</i>		0.043376	Group 1	6.7
GA30460	<i>Camta</i>		0.026603	Group 2	20.0
GA26409	CG43347		0.039305	Group 2	16.9
GA18735	<i>kmg</i>		0.024343	Group 2	89.7
GA19370	<i>piwi</i>	Spermatogenesis	0.048212	Group 1	5.1
GA23025	*		0.00916	Group 2	22.4
GA27003	<i>mil</i>		0.009169	Group 2	14.6
GA23808	<i>bb8</i>		0.016748	Group 2	24.0
GA24140	<i>Pkd2</i>		0.019148	Group 2	36.0
GA27927	<i>tbrd-2</i>		0.004809	Group 2	45.0
GA28467	<i>mael</i>		0.032811	Group 1	7.4
GA28096	<i>Lar</i>		0.015647	Group 2	20.0
<i>exu2</i> (GA28347)	<i>exu2</i>	Spermatogenesis, spermiogenesis	0.002473	Group 2	101.7
GA18558	<i>twe</i>		0.028595	Group 2	26.0
GA22296	<i>unc</i>		0.010802	Group 2	37.0
GA18412	<i>bol</i>		0.012463	Group 2	19.1
GA27688	<i>Psi</i>		0.001631	Group 1	13.3
GA21002	<i>sfl</i>		0.001349	Group 1	17.8
GA18272	<i>Nup154</i>		0.034358	Group 2	17.7
GA28835	<i>Hsp60A</i> <i>Hsp60C</i>		0.019586	Group 2	26.5
GA12828	<i>Rcd-1</i>		0.009415	Group 1	6.3
GA19706	<i>spag4</i>		0.033263	Group 2	30.0
GA12730	<i>U2A</i>		0.020332	Group 1	9.6
GA14055	<i>dila</i>		0.005707	Group 2	179.7
GA21384	<i>Prp8</i>		0.017177	Group 1	5.1
GA25581	<i>Sf3b5</i>		0.009304	Group 1	7.7
GA10370	<i>park</i>		0.043887	Group 2	29.4

GA25911	<i>Mst36Fb</i> <i>CG43339</i> <i>Mst36Fa</i>		0.015487	Group 2	68.9
GA19239	<i>l(3)72Ab</i>		0.024558	Group 1	6.5
GA20060	<i>Dic61B</i>		0.006508	Group 2	42.0
GA18699	<i>Trs20</i>		0.014058	Group 2	26.6
GA21874	<i>chic</i>	Spermiogenesis	0.002763	Group 1	5.8
GA10082	<i>W-Cup</i>	Post meiotic transcription, localised RNA	0.02755	Group 2	19.5
GA26457	<i>asl</i>	Axoneme assembly, cell cycle	0.04323	Group 2	32.8
GA19739	<i>ana1</i>	Axoneme assembly	0.033175	Group 2	23.3
GA14315	<i>Cep135</i>		0.009414	Group 2	49.0
GA15128	<i>alphaTub84D</i>		0.028341	Group 1	3.6

Table 4.7: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq based on hierarchical cluster analysis clustering into two cyst groups. Group 1: MVII, MX, RVII. Group 2: MI-MVI, MVIII, MIX, RI-RVI, RVIII-RX. *Orthologue of milkah (mil) previously supported, model updated and previous orthology no longer supported (Gramates et al. 2017; Larkin et al. 2021).

D. <i>pseudoobscura</i> Gene Name	D. <i>melanogaster</i> Gene Name	Gene Function	p-value	Highest Mean Normalised Count	Fold Change
GA21345	<i>achi vis</i>	tMAC, spermatogenesis, spermiogenesis	0.012652	Group 2.2	17.2
GA18885	<i>Taf4</i>	TFIID, spermatogenesis, spermiogenesis	0.034485	Group 2.2	3.5
GA16761	<i>Taf6</i>		0.047069	Group 2.1	168.0
GA25290	<i>can</i>		0.019971	Group 1	4.9
GA19104	<i>Hsf</i>	Transcription factor	0.002548	Group 1	8.2
GA11197	<i>Hr96</i>		0.026968	Group 2.2	43.1
GA18732	<i>Dad</i>		0.039782	Group 1	5.2

GA17066	<i>phtf</i>		0.014981	Group 1	2.6
GA30460	<i>Camta</i>		0.03176	Group 1	14.9
GA11505	<i>gcm</i>		0.040849	Group 2.1	2.7
GA12197	<i>Mnt</i>		0.027168	Group 2.1	5.0
<i>exu2 (GA28347)</i>	<i>exu2</i>	Spermatogenesis, spermiogenesis	0.034836	Group 1	10.8
GA18558	<i>twe</i>		0.042832	Group 1	5.5
GA12828	<i>Rcd-1</i>		0.039923	Group 2.1	3.9
GA14055	<i>dila</i>		0.040973	Group 2.2	5.4
GA18699	<i>Trs20</i>		0.004178	Group 1	7.2
GA25980	<i>Cd1c2</i> <i>ctp</i>		0.039118	Group 1	5.6
GA27927	<i>tbrd-2</i>	Spermatogenesis	0.006408	Group 2.2	9.8
GA15128	<i>alphaTub84D</i> <i>alphaTub84B</i>	Axoneme Assembly	0.032689	Group 2.1	3.4

Table 4.8: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of primary spermatocyte cyst RNA-seq based on hierarchical cluster analysis clustering into three cyst groups. Group 1: MVII, MX and RVII. Group 2.1: MII, MVI and MVIII. Group 2.2: MI, MIII, MIV, MV, MIX, RI, RII, RIII, RIV, RV, RVI, RVIII, RIX and RX.

4.3 Spermatid Cyst Differential Gene Expression Analysis

4.3.1 Spermatid cyst dissection for cDNA library preparation

Eusperm and parasperm morphs differ in length; eusperm are 304 μ m in length, parasperm 1 are 54 μ m, and parasperm 2 are 101 μ m. Eusperm and parasperm spermatocyte cysts cannot be distinguished visually, however spermatid cysts vary in length according to spermatid morph. RNA-seq of spermatid cysts may reveal transcriptional variation between the eusperm and parasperm morphs late in development, as they develop their distinct morphologies.

Spermatid cysts were dissected from *D. pseudoobscura* SLOB3 wild type testes. Pairs of long and short spermatids were dissected from each male. Six pairs of long and short spermatid cysts were dissected in total. Spermatid cysts were imaged prior to collection and RNA extraction. Long spermatid cysts were assumed to contain elongated eusperm spermatids, short cyst were assumed to contain parasperm spermatids. Parasperm morphs 1 and 2 were confirmed after spermatid RNA-seq was completed, and therefore short cysts may represent either morph.

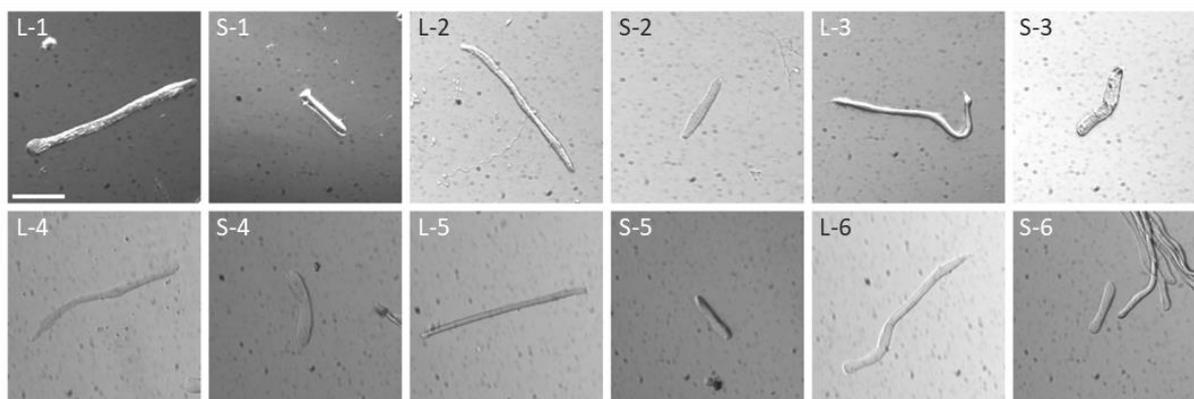


Figure 4.9: Spermatid cysts dissected from six *D. pseudoobscura* *SLOB3* wild-type males. Cysts were collected into individual tubes and RNA extracted for preparation of cDNA libraries. Scale = 50 μ m. L-1 – L-6: Long spermatid cysts, sample IDs L1-L6. S-1 – S-6: Short spermatid cysts, sample IDs S1-S6.

4.3.2 Sequence quality

RNA extraction, cDNA and library synthesis from individual spermatids was carried out using the QIAseq FX Single Cell RNA Library kit (Qiagen). Libraries were then sequenced by HiSeq (Illumina). All twenty cyst libraries were successfully sequenced to generate 75 base pair paired-end reads.

Raw sequence quality was high (mean Phred >20) with sequence lengths above 70 base-pairs. Sequences were trimmed to remove adaptor sequences and sequence regions of low quality (Phred <20), then filtered to remove reads less than 36 base pairs in length and of low quality (mean Phred <20 across 4bp). Paired-end read counts before and after trimming are summarised in Table 4.9.

Reads were aligned to the *D. pseudoobscura* reference genome version 3.04 (Gramates *et al.* 2017). The data were filtered to remove unmapped and duplicate reads. Aligned reads were counted to give a total count for each annotated gene per spermatid cyst. Unmapped reads were also counted. Count data are summarised in Table 4.10. A total of 14539 genes were represented in spermatid cyst RNA sequencing data.

As was observed in the second spermatocyte cyst RNA-seq dataset, there was a substantial decrease in the number of mapped reads compared to the number of paired-end reads after filtering, as a result of filtering after alignment. Low mapped read counts may have impacted subsequent cluster analysis.

Raw and normalised count data are available in Appendix 7.

Spermatid Cyst Sample ID	Raw Data Paired-End Reads	Filtered Paired-End Reads	% Sequences Filtered
L1	4568836	3581722	21.61

S1	4786284	4240936	11.39
L2	5543532	4908783	11.45
S2	5332375	4804359	9.90
L3	5469920	5019206	8.24
S3	6103621	5457518	10.59
L4	5577370	4658457	16.48
S4	5763178	4834037	16.12
L5	4716260	4184175	11.28
S5	5493276	5012410	8.75
L6	5724723	4768768	16.70
S6	5399020	4481257	17.00

Table 4.9: RNA-seq analysis of *D. pseudoobscura* SLOB3 wild type spermatid cysts. Paired-end read counts of *D. pseudoobscura* SLOB3 wild type spermatid cyst RNA sequence raw data of 12 spermatid cysts dissected from six males. Read counts are provided for data prior to and post trimming and filtering by Trimmomatic (Bolger et al. 2014). The percentage of sequences filtered was calculated.

Sample	Mapped Read Count	Unmapped Read Count	Genes Represented
L1	832416	294375	5464
S1	1264406	405357	7319
L2	1715963	483677	9843
S2	1905267	457980	11020
L3	324678	399157	4220
S3	597350	361383	6942
L4	1355509	542695	8460
S4	1790822	378247	10316
L5	873155	570111	9983
S5	1336767	558164	11604
L6	1548528	434224	8533
S6	1310809	421055	8984
Total			14539

Table 4.10: Counts of mapped and unmapped reads from *D. pseudoobscura* SLOB3 wild type single spermatid cyst RNA sequencing. Reads were aligned to *D. pseudoobscura* reference genome. Unmapped reads include reads for which no feature is mapped, or for which alignment was ambiguous. Read counts performed by htseq (Anders et al. 2015). The total number of genes represented in each spermatid dataset is given.

4.3.3 Hierarchical cluster analysis did not distinguish between long and short spermatid cyst morphs

Spermatid cysts are distinguishable by appearance (cyst length indicates morph), and as such differential gene expression was based on morph appearance instead of hierarchical cluster analysis, as was required for primary spermatocyte cyst RNA-seq analysis.

Hierarchical cluster analysis of normalised spermatid cyst transcription data did not show clustering by apparent morph (Figure 4.10). Restricting HCA to transcription data of the 10% most highly expressed genes altered clustering, but also did not cluster according to apparent morph. Similarly, clustering based on transcription of the post-meiotic expressed cup genes altered clustering but did not cluster according to morph. HCA of the whole data set and the 10% most highly expressed genes showed similar clustering patterns, the most significant change being the removal of short spermatid cyst 3 (S3).

This result suggests that HCA may not be a robust method for distinguishing between spermatid cyst morphs, and may show that in fact, HCA as a method is not reliable for distinguishing between cyst morphs at any stage of development without larger sample sizes, or more indication of genes which are known to be highly differentially expressed between morphs.

Furthermore, while the aim of spermatid RNA-seq was to collect long and short spermatid cysts, there may be further confounding factors which are more relevant to transcription profile, which were not easy to distinguish by bright-field microscopy alone. 'Short' spermatid cysts could be elongated parasperm 1 or 2, or elongating parasperm 2 or eusperm. 'Long' spermatid cysts could be at various stages of elongation or individualisation. It is not certain which of these factors is relevant for each spermatid cyst sample in this study.

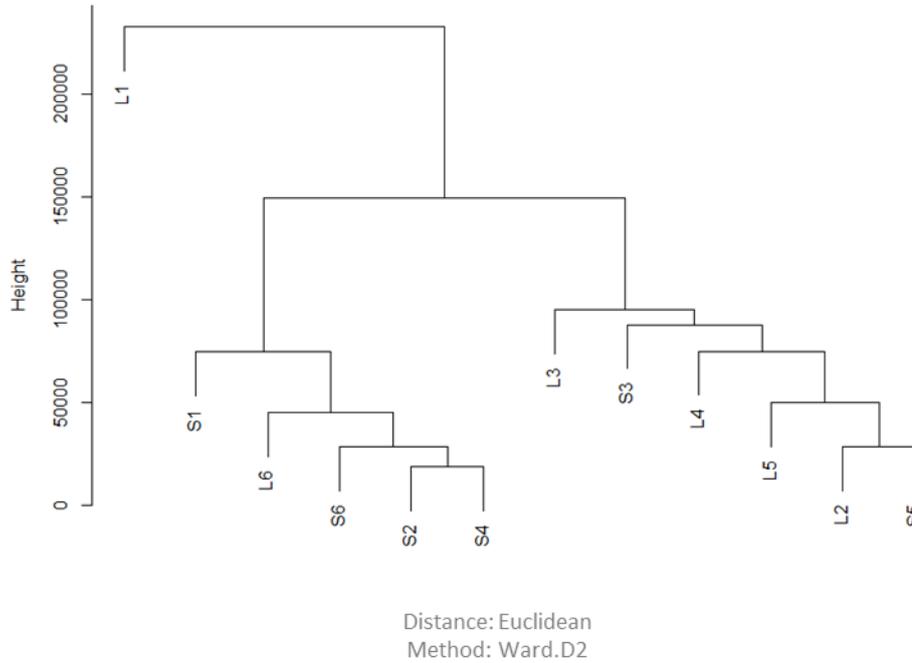


Figure 4.10: Hierarchical cluster analysis dendrogram of *D. pseudoobscura* *SLOB3* wild type spermatid cyst RNA sequencing data. Normalisation by TMM. Pairwise distances Euclidean. Clustering method Ward D2. No filtering.

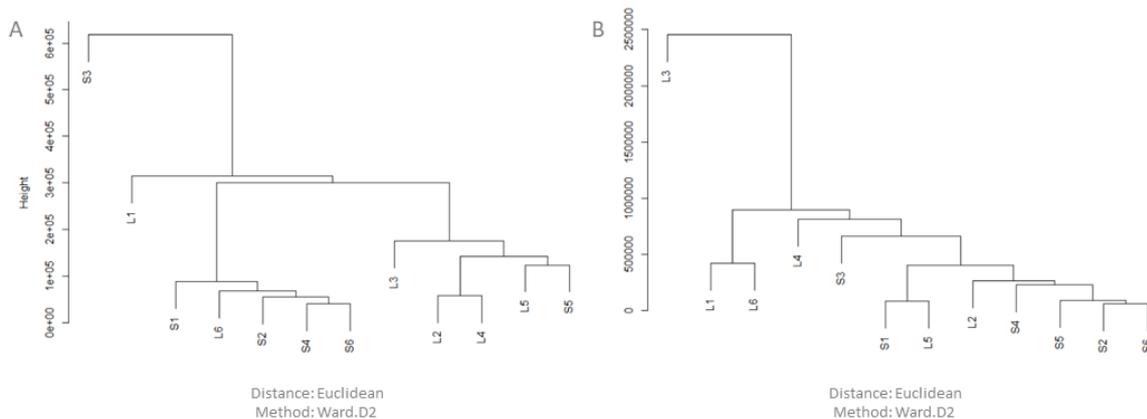


Figure 4.11: Hierarchical cluster analysis dendrogram of *D. pseudoobscura* *SLOB3* wild type spermatid cyst RNA sequencing data. Normalisation by TMM. Pairwise distances Euclidean. Clustering method Ward D2. A: Clustering based on 10% highest expressed genes in spermatid cysts. B: Clustering based on cup gene expression.

4.3.4 Differential gene expression analysis of long and short spermatid cysts

Differential gene expression analysis of spermatid cysts was performed on two data groups, based on cyst morph – long or short. Analysis is limited as short cyst data may relate to any of short or medium elongated cysts, or medium or long elongating cysts. The analysis still has value as the DGE analysis can still identify differentially expressed genes in the elongated long cysts, compared to other morphs and stages (if the short cyst data is not uniformly parasperm 1 morph data), giving some indications

of gene expression specific to the elongated eusperm morph and stage. Genes found to be higher in the short cysts may indicate degradation of those transcripts in the long cysts, previously expressed in the spermatocyte stage then translated in the elongating spermatid stage.

1401 genes were found to be significantly differentially expressed in long and short spermatid cysts by differential gene expression analysis ($p < 0.05$). The full list of differentially expressed genes is available in Appendix 8. Differentially expressed genes with functions in spermiogenesis and cell cycle are summarised in Table 4.11.

A greater proportion of the differentially expressed genes were more highly expressed in short cysts. 900 differentially expressed genes were more highly expressed in short cysts, whereas 501 differentially expressed genes were more highly expressed in long cysts. Long spermatid cysts had greater overall expression than short spermatid cysts, on average. Long cysts had a median total normalised mapped read count of 1138538. Short cysts had a median total normalised mapped read count of 877165.9.

<i>D. pseudoobscura</i> Gene Name	<i>D. melanogaster</i> Gene Name	Highest Expression Morph	Gene Function	p-value	Fold Change
GA11961	<i>presidents-cup</i>	L	Post meiotic transcription, localised RNA	0.026282	6.2
GA13552	<i>sungrazer</i>	S		0.002709	6.2
GA10082	<i>walker-cup</i>	S		7.01E-05	15.3
GA15601	<i>stanley-cup</i>	L		0.022061	6.0
GA12730	<i>U2A</i>	L	Spermiogenesis	0.04358	8.1
GA14608	<i>dia</i>	L		0.033864	24.6
GA18820	<i>SmB</i>	L		0.023487	7.1
GA19299	<i>Hip14</i>	L		0.00351	20.2
GA23656	<i>Trl</i>	L		0.025705	4.4
GA28096	<i>Lar</i>	L		0.014955	13.8
GA27209	<i>tbrd-1</i>	S		0.005223	36.7
GA28949	<i>ACXE, ACXA, ACXB, ACXC</i>	S		0.045937	3.5
GA28966	<i>eIF4E3</i>	S		0.041543	3.7
<i>exu2 (GA28347)</i>	<i>exu2</i>	S		8.60E-05	13.5
GA12828	<i>Rcd-1</i>	S		0.008316	7.0
GA18412	<i>bol</i>	S		0.000837	5.9

GA18673	<i>pAbp</i>	S		0.029695	3.5	
GA18757	<i>Tasp1</i>	S		0.006535	22.2	
GA19020	<i>nos</i>	S		0.024499	35.9	
GA20022	<i>fwd</i>	S		0.008562	6.0	
GA21079	<i>Lis-1</i>	S		0.027888	7.3	
GA21156	<i>Kap-α1</i>	S		0.023539	8.5	
GA21506	<i>Ddx1</i>	S		0.039034	15.1	
GA23025	*	S		0.001265	7.8	
GA24628	<i>eIF4E3</i>	S		0.008469	5.0	
GA25290	<i>can</i>	S		0.032069	14.8	
GA26552	<i>eIF4G2</i>	S		0.022786	3.8	
GA28835	<i>Hsp60A, Hsp60C</i>	S		0.013745	4.5	
GA21613	<i>Gas41</i>	S		TFIID, transcription factor	0.018133	4.7
GA10095	<i>ndl</i>	S		Cell cycle	0.005388	11.0
GA11545	<i>polo</i>	S	0.015656		9.8	
GA19337	<i>sra</i>	L	0.024651		7.1	
GA22911	<i>asl</i>	S	0.029229		1.9	

Table 4.11: Summary of differentially expressed genes ($p < 0.05$) from differential gene expression analysis of spermatid cyst RNA-seq based on long and short cyst groups. *Orthologue of milkah (mil) previously supported, model updated and previous orthology no longer supported (Gramates et al. 2017; Larkin et al. 2021).

4.4 Comparison of Spermatocyte Cyst RNA-Seq Datasets

4.4.1 RNA-seq 1 vs. RNA-seq 2

Two different methods of RNA extraction and library preparation were used for the first and second rounds of primary spermatocyte RNA-seq. Differential gene expression analysis of these primary spermatocyte RNA-seq datasets identified significantly differentially expressed genes based on the grouping by hierarchical cluster analysis. Initial analysis shows that genes identified as significantly differentially expressed were not consistent across datasets, potentially as a result of variation in RNA extraction, cDNA synthesis, library preparation and sequencing methods.

Some genes were found to be consistently significantly differentially expressed across both primary spermatocyte cyst datasets. Where two cyst groups were defined, DGE analysis of A spermatocyte cyst data found 1399 significant genes, and DGE analysis of M and R spermatocyte cyst data found

1613 significant genes. Comparison of the significantly differentially expressed genes of both spermatocyte cyst RNA-seq data sets showed 241 genes appeared on both lists of significant genes.

Where three cyst groups were defined, DGE analysis of A spermatocyte cyst data found 1528 significant genes, and DGE analysis of M and R spermatocyte cyst data found 747 significant genes. 110 genes were found to be significantly differentially expressed in both datasets.

DGE analysis of spermatid cyst data found 1401 significant genes, of which 189 were also significant in M and R spermatocyte cyst DGE analysis.

4.4.2 Unique expression in spermatid cysts

The post-meiotic transcriptome is mostly comprised of genes expressed prior to meiosis. Gene expression during sperm development occurs mostly during the spermatocyte stage, although some additional gene expression also occurs after meiosis (Barreau *et al.* 2008b). Transcripts identified by spermatid RNA-seq are likely to reflect this. It is not possible to determine the stage at which transcripts are first expressed from these datasets, however comparison of primary spermatocyte cyst and spermatid cyst data identifies genes which appear only in the spermatid datasets.

After filtering, spermatocyte cyst M and R datasets contained a total of 14420 expressed genes. The filtered spermatid cyst data contained 12122 expressed genes. 948 genes were present only in the spermatid expression list, and not in spermatocyte RNA-seq 2 expression data. Of these, 124 were also differentially expressed between long and short spermatid cysts (Appendix 9).

It is unlikely that all of the genes identified only in the spermatid data are expressed post-meiotically. The spermatocytes are the most transcriptionally active cells in adult *Drosophila*, and are more transcriptionally complex than spermatids. Spermatid cysts have a smaller transcriptome in comparison to spermatocyte cysts, but were sequenced to a similar depth in this experiment. The presence of genes in the spermatid dataset, which were not present in the spermatocyte datasets, may be the result of the increased chance of detecting a given gene whose mRNA perdures from pre-meiotic transcription. Some genes identified in spermatid cysts, but not spermatocyte cysts, may have post-meiotic expression, however further analysis would be needed to confirm this.

4.5 Further Investigation of Differential Gene Expression by RNA *In Situ* Hybridisation Staining of *D. pseudoobscura* Testes: Prioritisation of Candidate Genes

RNA-seq of *D. pseudoobscura* spermatocyte and spermatid cysts generated large datasets of significantly differentially expressed genes. To further validate the RNA-seq analysis and to narrow down the lists of candidate genes to follow up in subsequent experiments, an *in situ* hybridisation method was used to examine expression of differentially expressed genes with functions in spermatogenesis, spermiogenesis and transcription, and other functions of interest.

In situ hybridisation of candidate genes showed varied results. Expression ranged from little detectable expression, to ubiquitous expression across one or more stages of development, to clear differential expression.

4.5.1 Differential gene expression of *GA18735*, an orthologue of *kumgang*.

GA18735 is the *D. pseudoobscura* orthologue of the *D. melanogaster* gene *kumgang* (*CG5204*), hereafter *kmg*. *D. melanogaster* *Kmg* has been shown to contribute to the maintenance of the germline-specific gene expression programme. *Kmg* blocks Aly (a component of the testis meiotic arrest complex) binding to cryptic promoters of somatic genes, preventing expression in the testes (Kim *et al.* 2017). The function of *D. pseudoobscura* *Kmg* is not known.

In situ hybridisation showed *D. pseudoobscura* *kmg* to have strong transcript staining in some spermatocyte cysts, with no detectable staining in other spermatocyte cysts (Figure 4.12). There was no transcript detection in the hub region or early spermatogonia, or in the post-meiotic cysts.

It is not possible to provide an absolute quantitation of gene expression based on *in situ* hybridisation, however it is valid to compare expression levels of a gene within testes. Figure 4.12 indicates that the expression of *kmg* is high in some spermatocyte cysts, but low if not completely absent in other spermatocyte cysts. *kmg* may have a role in the differential development of sperm morphs in *D. pseudoobscura*.

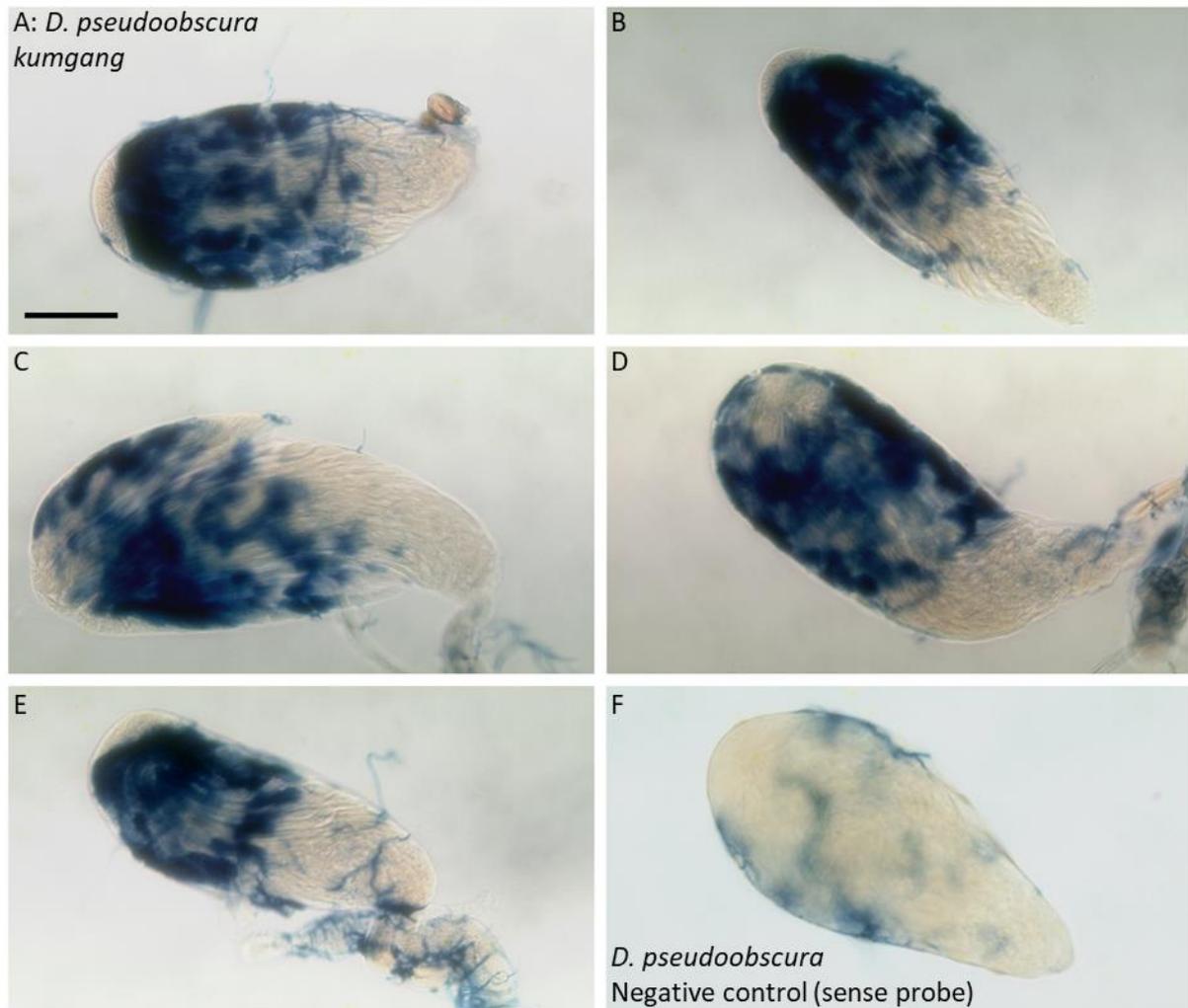


Figure 4.12 A-E: *In situ* hybridisation of GA18735, orthologous to the *D. melanogaster* gene *kumgang* (*kmg*). Left – right, apical – proximal. Scale = 100 μ m. Staining appears patchy throughout the primary and secondary spermatocyte regions of the testis. In *D. melanogaster*, *Kumgang* is a transcriptional repressor, blocking *Aly* from accessing promiscuous promoters of somatic genes, therefore maintaining germline-specific transcription (Kim *et al.* 2017). F: Negative control using sense probe for GA21874 (*chickadee*).

4.5.2 Some components of the testis meiotic arrest complex show differential gene expression.

4.5.2.1 GA21345 (*achi/vis*) differential gene expression is isoform dependent

GA21345 is orthologous to the *D. melanogaster* tMAC genes *achintya* (*achi*) and *vismay* (*vis*). *Achi* and *Vis* are DNA binding proteins, likely to be involved in transcription in spermatocytes (Ayyar *et al.* 2003; Wang and Mann 2003; Beall *et al.* 2007). There are three isoforms of *D. pseudoobscura* *achi/vis*; RA, RB and RC. RA and RC have the same coding sequence, with alternative splicing in the 5' untranslated region. The 5' UTR of RB is the same as RC, but RB coding sequence does not contain exon 3 (Figure

4.13) (Gramates *et al.* 2017; Yang *et al.* 2018). *D. melanogaster achi* and *vis* also contain an exon (exon 4) expressed only in the testes (Gramates *et al.* 2017; Thurmond *et al.* 2019; Larkin *et al.* 2021).

Figure 4.14 shows *in situ* hybridisation of the *D. pseudoobscura achi/vis* isoforms. Figure 4.14 A shows staining of all three isoforms with a single probe. Staining for all three isoforms simultaneously shows consistent, strong staining in the regions of the testes containing the early primary spermatocytes, with weaker staining in the earlier stages – hub and mitotic spermatogonia regions – and in the late stage primary spermatocytes, with little to no staining after the onset of meiosis. There is no obvious differential expression in any of the stages.

The probe for *achi/vis*-RB was designed to target exon 2 to exon 4 splicing. Staining for *achi/vis*-RB showed a similar pattern to staining with a probe recognising all three isoforms, with strong staining in the early primary spermatocytes and weaker staining in later spermatocyte stages (Figure 4.14; B). There was no detectable staining in earlier stages.

Staining with a probe targeting exon 3 – which is present in the coding sequence of isoforms RA and RC – showed strong expression in the spermatogonia stages, which appeared uniform, with expression decreasing in spermatocytes. A patchy staining pattern, similar to that observed in *D. pseudoobscura kmg*, indicated that *achi/vis*-RA/C expression was higher in a subset of spermatocyte cysts.

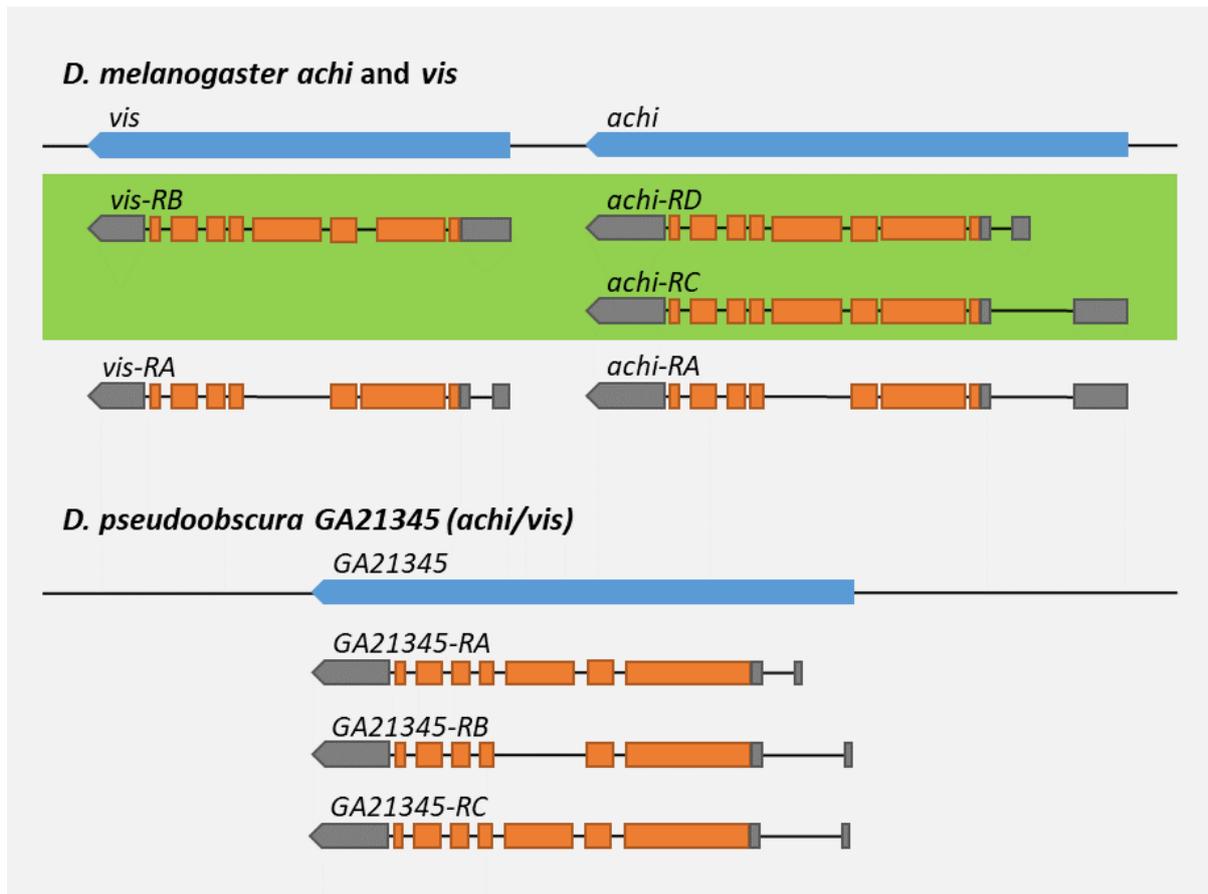


Figure 4.13: Alternative splicing of *D. melanogaster* *achintya* (*achi*) and *vismay* (*vis*), a duplicate pair of meiotic arrest genes, components of the testis meiotic arrest complex (tMAC). Green box indicates testis-specific isoforms. *D. pseudoobscura* has a single orthologous copy of *achi/vis*; GA21345. Exon 3 of *D. pseudoobscura* GA21345 is testis specific (Gramates et al. 2017; Yang et al. 2018).

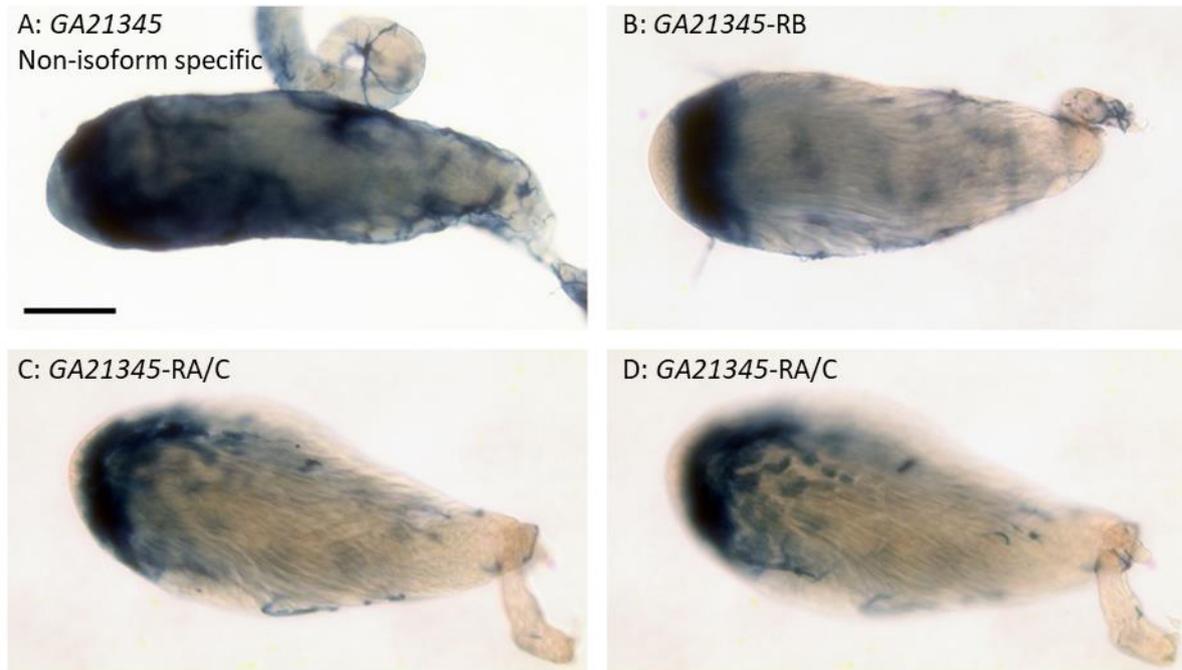


Figure 4.14: In situ hybridisation of GA21345, orthologous to the *D. melanogaster* genes *achintya* and *vismay* (*achi/vis*). GA21345 has three isoforms. Left – right, apical – proximal. Scale = 100 μ m. A: GA21345, non-isoform specific probe. B: Probe targeting GA21345 isoform RB. C-D: GA21345 isoforms RA and RC.

4.5.2.2 *D. pseudoobscura* *Caf1A* (GA18051) and *Caf1B* (GA26389) – A duplicated pair of the *D. melanogaster* tMAC component; *Caf1*

D. pseudoobscura has two orthologues of the *D. melanogaster* tMAC component *Chromatin assembly factor 1* (*Caf1*) (Calvo-Martin *et al.* 2017). The syntenic orthologue, GA18051, or *Caf1A*, and its paralogue, GA26389, or *Caf1B*, showed different staining patterns within the testis (Figure 4.15). *Caf1A* was detected in the early stages of sperm development, in the hub region or spermatogonia and the spermatocyte stages. Staining appeared patchy in the spermatocyte cysts, with stronger staining in a subset of cysts.

Caf1B staining was limited to a smaller region of the testis, indicating that this paralogue was expressed at fewer stages of spermatogenesis. *Caf1B* staining appeared to be localised to the spermatogonia and early spermatocytes. Staining for *Caf1B* did not appear patchy.

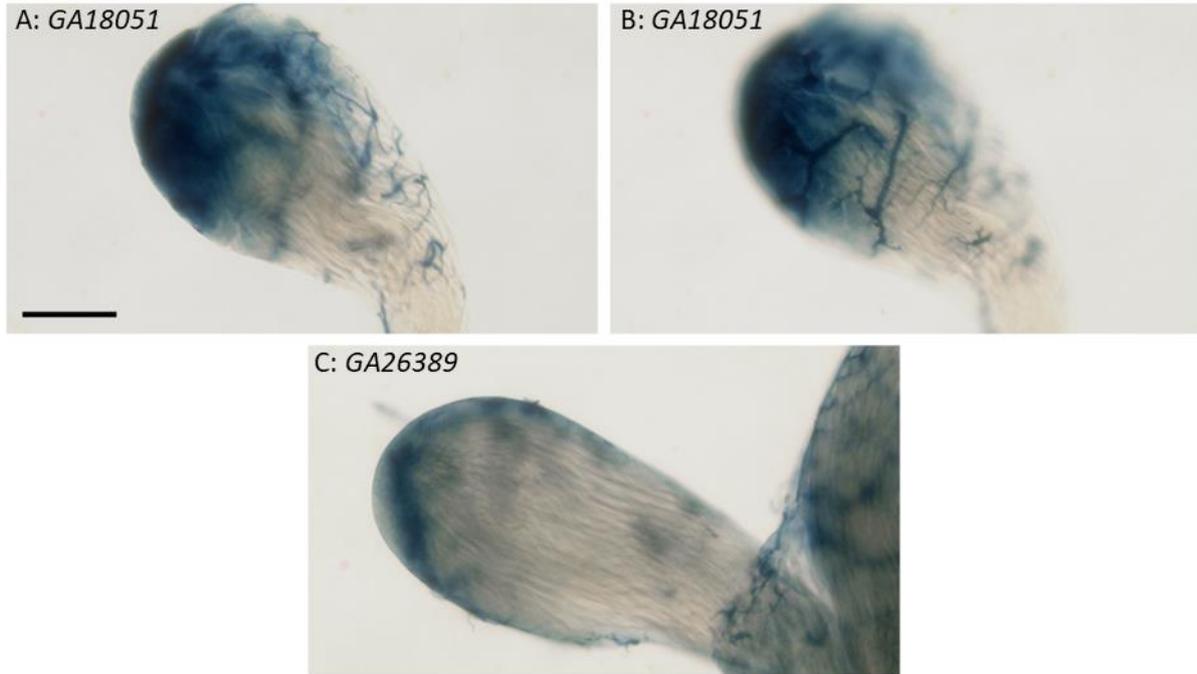


Figure 4.15: *In situ* hybridisation of GA18051 and GA26389, orthologous to the *D. melanogaster* gene Chromatin assembly factor 1 (Caf1). Left – right, apical – proximal. Scale = 100 μ m. A-B: GA18051, Caf1A. C: GA26389, Caf1B.

4.5.2.3 Other tMAC components

The testis meiotic arrest complex (tMAC) is essential for the correct transcription of thousands of genes within the germline (Jiang and White-Cooper 2003; Beall *et al.* 2007; Laktionov *et al.* 2014; Laktionov *et al.* 2018). The *D. pseudoobscura* tMAC orthologues of *achi-vis* and *Caf1A* showed evidence of differential gene expression in *in situ* hybridisation staining of the testes. Other tMAC orthologues did not show similar patchy staining, which would suggest differential gene expression in testes (Figure 4.16).

GA28313, orthologous to *always early (aly)* was detected in spermatogonia through to the late spermatocytes, with staining less detectable in elongating spermatids (Figure 4.16; A). GA12700, orthologous to *tombola (tomb)* had stronger staining in the spermatogonia to early spermatocyte stages, with staining weaker in later spermatocytes (Figure 4.16; B). GA12326, orthologous to *cookie monster (comr)*, showed strong staining in the late spermatogonia to early spermatocyte stage (Figure 4.16; C), but was also detectable in the spermatid cysts, and appeared to be localised to the elongating tail ends of the spermatid cysts (Figure 4.16; D). GA11638, orthologous to *wake-up call (wuc)*, showed staining from the hub to the primary spermatocyte stages, with an increase in staining intensity at the late spermatogonia/early spermatocyte stage (Figure 4.16; E).

The *D. melanogaster* *Myb-interacting protein 40 (mip40)* gene has two *D. pseudoobscura* orthologues; GA13507, the syntenic orthologue, or *mip40A*, and GA23669, or *mip40B*. The *D. pseudoobscura* *mip40*

paralogues show different staining patterns within the testis, as observed with the Caf1 paralogues. *mip40A* staining was detectable in pre-meiotic stages, from the early spermatogonia to the late spermatocytes, but there was no staining in the post-meiotic cysts (Figure 4.16; F). *mip40B* was detectable only in the spermatids, shown by the presence of staining in the central region of the testis, with no expression in the distal regions of the testis (Figure 4.16; G-H).

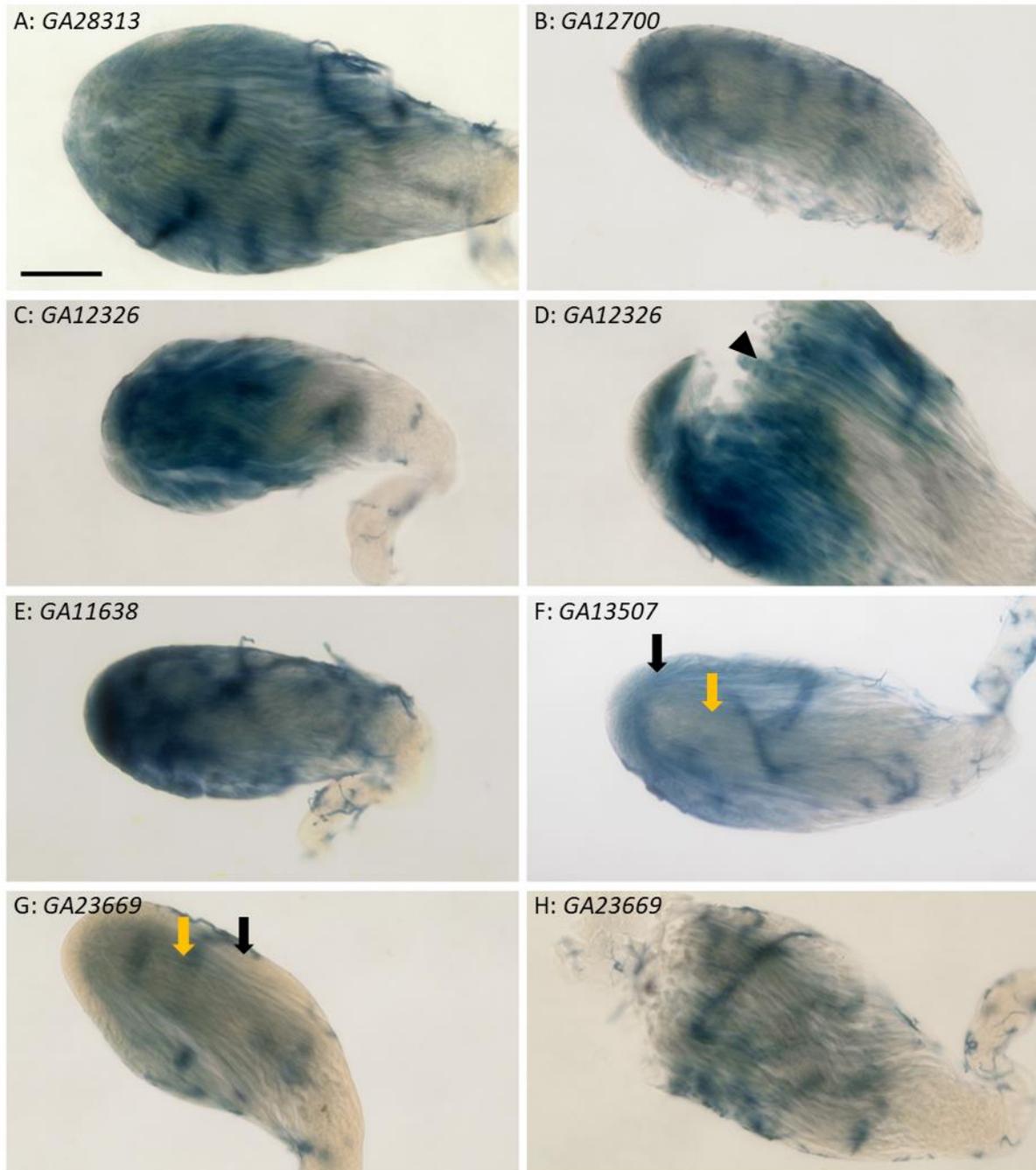


Figure 4.16: *In situ* hybridisation of tMAC component genes, orthologous to *D. melanogaster* tMAC genes. Left – right, apical – proximal. Scale = 100 μ m. A: GA28313, always early (aly). B: GA12700, tombola (tomb). C-D: GA12326, cookie monster (comr). Black arrowhead indicates localisation in spermatid cysts. E: GA11638, wake-

up call (*wuc*). *F-H: Orthologues of D. melanogaster Myb-interacting protein 40 (mip40)*. Black arrows indicate pre-meiotic cyst regions of testes. Yellow arrows indicate post-meiotic regions of testes. *F: GA13507, mip40A. G-H: GA23669, mip40B.*

4.5.3 Transcription factor genes

Other genes orthologous to *D. melanogaster* transcription factors identified by RNA-seq were also investigated by RNA *in situ* hybridisation to examine expression within the testes. Transcription factor genes other than those contributing to tMAC were detected in the testes (Figure 4.17). This indicates that these transcription factor orthologues are expressed in the testes of *D. pseudoobscura*, and may have functions in spermatogenesis. None showed the patchy staining pattern observed in *D. pseudoobscura kmg, achi/vis* and *Caf1A* staining.

GA18636 is orthologous to *D. melanogaster crocodile (croc)*, which has been found to prevent germline differentiation in ovaries by preventing BMP pathway signalling (Tu *et al.* 2020). Staining for *GA18636* was uniform from the hub region, throughout spermatogenesis, up to meiosis. It was not possible to determine the staining of the spermatids. This staining suggests that *GA18636* is expressed early in sperm development and that the mRNA perdures through meiosis.

GA12828 is orthologous to *D. melanogaster Required for cell differentiation 1 (Rcd-1)*, a component of the CCR4-NOT complex, involved in negative regulation of mRNA translation (Sgromo *et al.* 2018). Staining was weak in the spermatogonia, increasing in strength during the early spermatocyte stage. Staining was present at least up to meiosis, but was weaker post-meiosis. Staining suggests that *D. pseudoobscura Rcd-1* is expressed in the spermatogonia, with increasing expression in the spermatocytes.

GA17585 is orthologous to *D. melanogaster shuttlecraft (stc)*, a zinc-finger transcription factor with functions in embryo neurogenesis and adult lifespan (Stroumbakis *et al.* 1996; Pasyukova *et al.* 2004). *GA17585* staining was weak in the spermatogonia, increasing in strength in the early spermatocyte stage, then decreasing in strength throughout the spermatocyte stage. Staining was not detectable in late stage primary spermatocytes.

GA27927 is orthologous to *D. melanogaster testis-specifically expressed bromodomain containing protein-2 (tbrd-2)*, a spermatogenesis transcription factor that interacts with the testis TAFs (Theofel *et al.* 2014; Theofel *et al.* 2017). *GA27927* staining was detected in the spermatogonia and spermatocytes, with decreased staining in the later spermatocytes and no detectable post-meiotic staining.

GA19264 is orthologous to *D. melanogaster without children (woc)*, which has functions in GSC differentiation in the ovaries, and may regulate expression of *zfh1* and *Stat92E* (Maimon *et al.* 2014). Different staining patterns were detected between two *GA19264* isoforms; RD and RE. Staining for *woc*-RD was detected in the hub, with decreasing staining in subsequent stages. *woc*-RE showing no detectable staining.

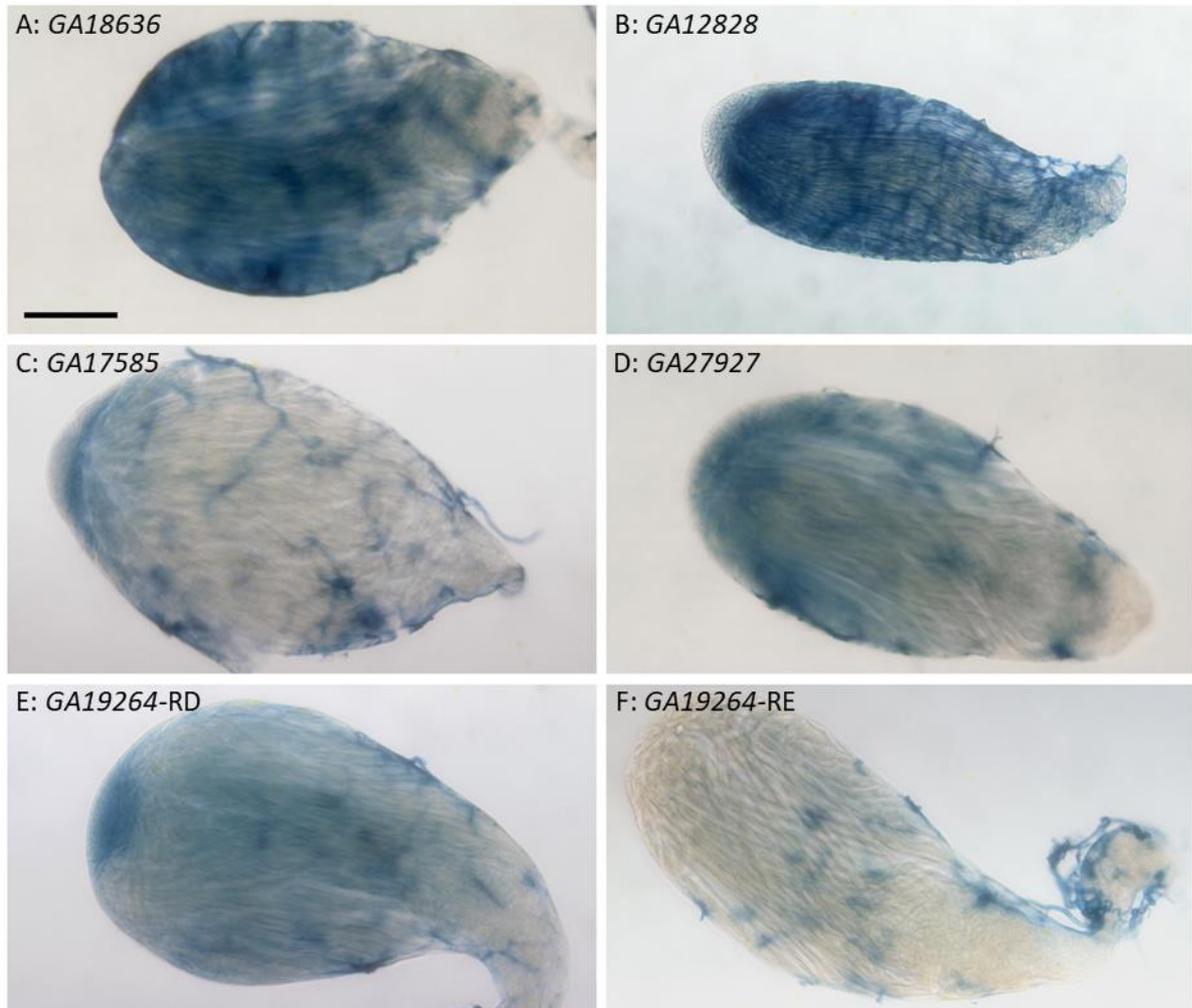


Figure 4.17: In situ hybridisation of transcription factor genes. Left – right, apical – proximal. Scale = 100 μ m. A: GA18636, crocodile (*croc*). B: GA12828, Required for cell differentiation 1 (*Rcd-1*). C: GA17585, shuttlecraft (*stc*). D: GA27927, testis-specifically expressed bromodomain containing protein-2 (*tbrd-2*). E-F: Isoforms of GA19264, without children (*woc*). E: GA19264 isoform RD. F: GA19264 isoform RE.

4.5.4 Meiosis genes *twine* and *boule* show differential gene expression

boule is a translational regulator of *twine*, required for entry into meiosis and normal spermatid elongation (Maines and Wasserman 1999). The *D. pseudoobscura boule (bol)* and *twine (twe)* orthologues, *GA18412* and *GA18558* respectively, showed patchy staining (Figure 4.18 and Figure 4.19). Detection of *bol* and *twe* was lower in spermatogonia, with higher detection in primary

spermatocytes. *bol* was detected in all primary spermatocyte cysts, with stronger staining in a subset of spermatocyte cysts. *twe* was detectable in a subset of spermatocyte cysts.

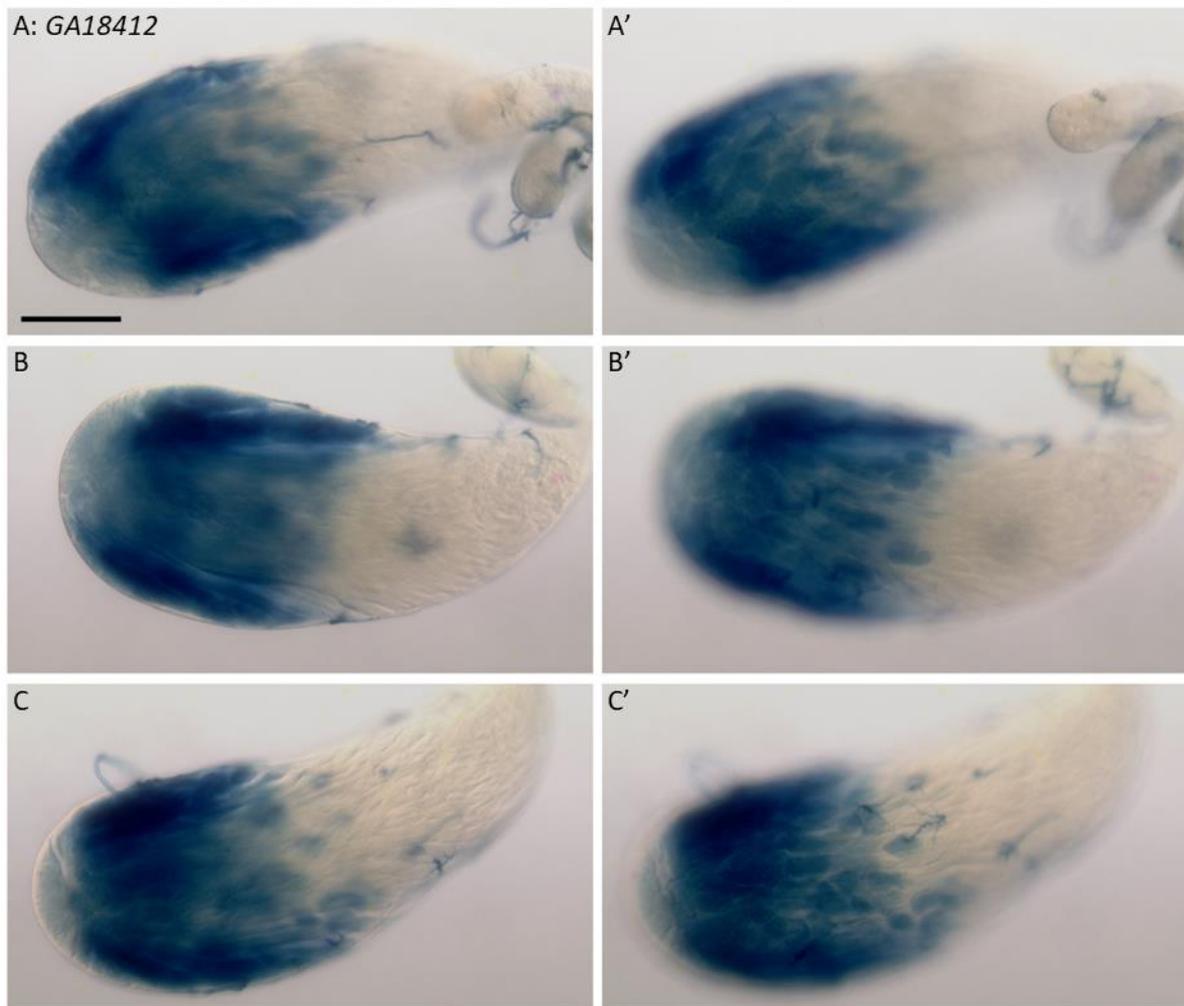


Figure 4.18: In situ hybridisation of GA18412, orthologous to *D. melanogaster* meiosis gene *boule* (*bol*). Stained testes are shown at two different focal planes, showing surface and more internal staining. Left – right, apical – proximal. Scale = 100 μ m. A-C: Surface staining of GA18412. A'-C': Internal staining of GA18412.

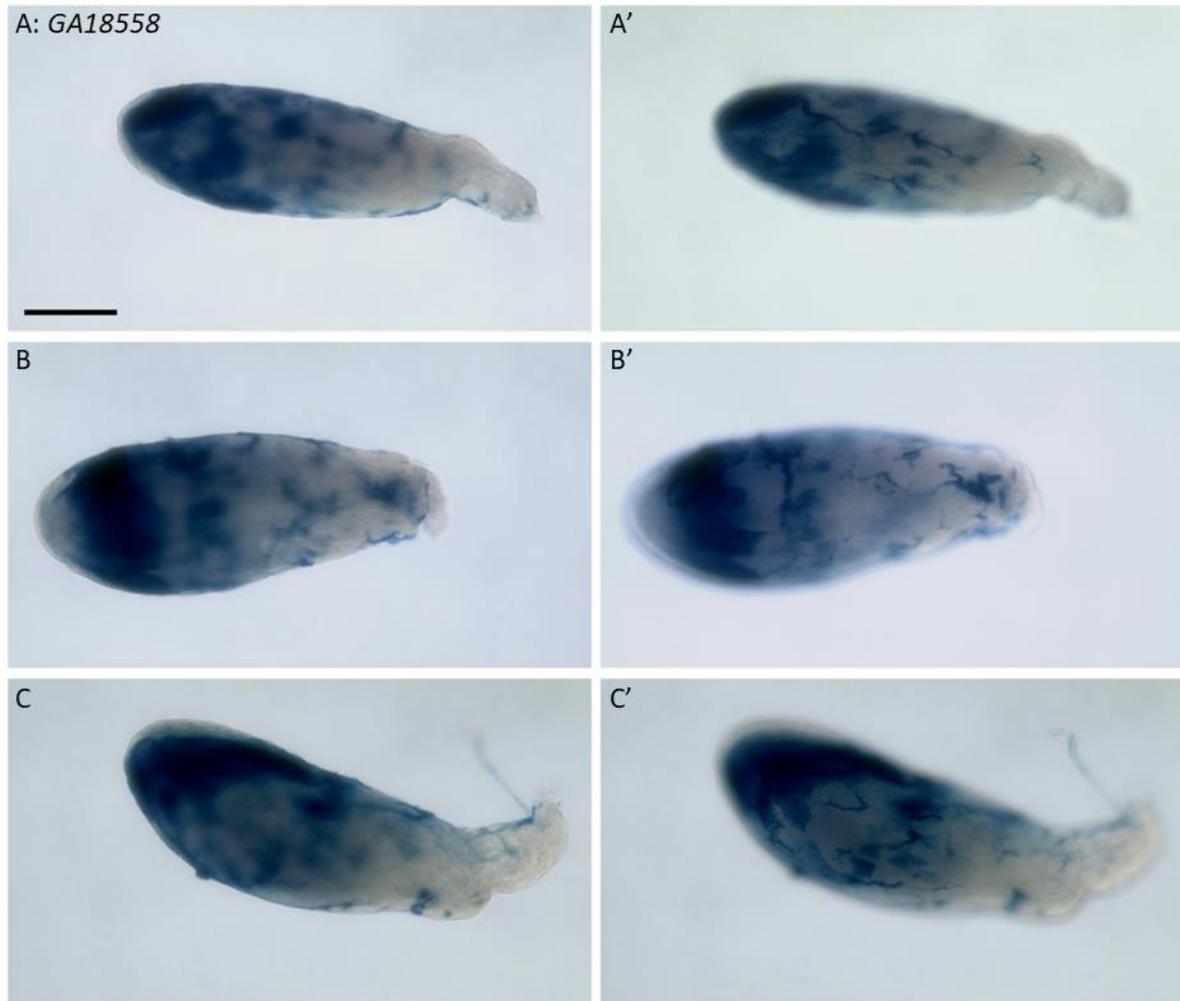


Figure 4.19: In situ hybridisation of GA18558, orthologous to *D. melanogaster* *meiosis gene twine (twe)*. Stained testes are shown at two different focal planes, showing surface and more internal staining. Left – right, apical – proximal. Scale = 100 μ m. A-C: Surface staining of GA18558. A'-C': Internal staining of GA18558.

4.5.5 Axoneme assembly gene *Dynein intermediate chain 61B (dic61B)*

Dynein intermediate chain at 61B (dic61B) is required for axoneme assembly (Fatima 2011). The *D. pseudoobscura* *dic61B* orthologue, GA20060, was detected in the late primary spermatocyte stage, entering into meiosis. Staining was patchy, indicating expression was higher in a subset of cysts (Figure 4.20).

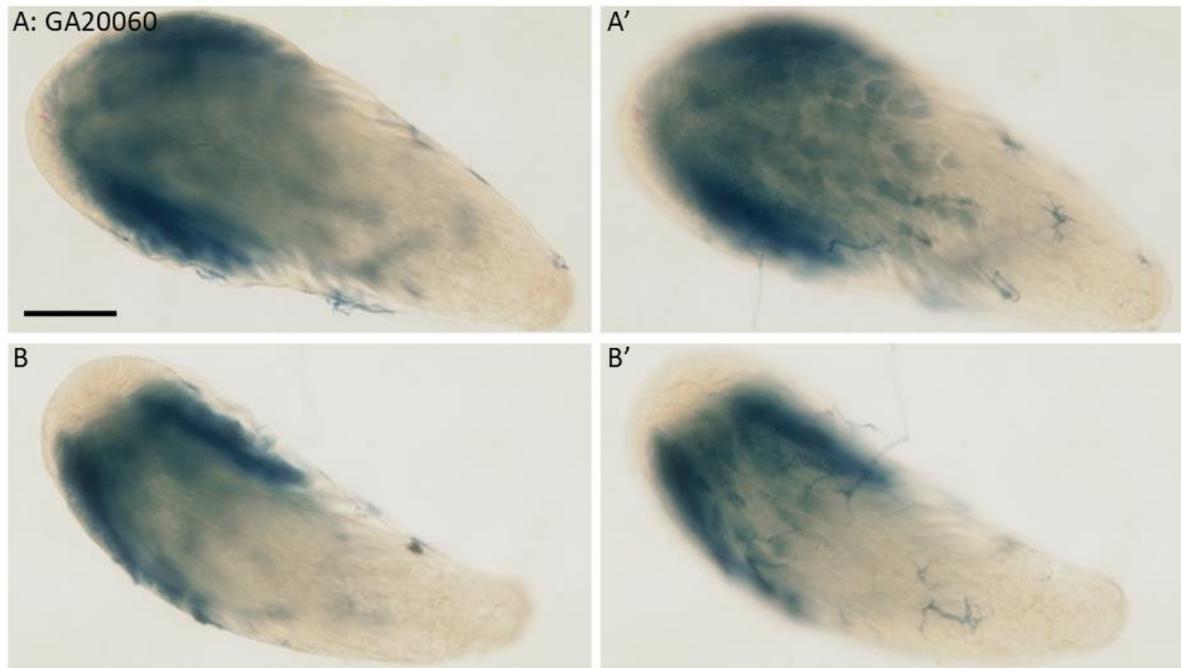


Figure 4.20: *In situ* hybridisation of GA20060, orthologous to *D. melanogaster* spermatogenesis gene Dynein intermediate chain at 61B (Dic61B). Stained testes are shown at two different focal planes, showing surface and more internal staining. Left – right, apical – proximal. Scale = 100 μ m. A-B: Surface staining of GA20060. A'-B': Internal staining of GA20060.

4.5.6 *milkah* (*mil*) is required for localisation and shaping of the nucleus in the elongating spermatid

GA27003 is orthologous to *D. melanogaster* *milkah* (*mil*). In *D. melanogaster* *mil* is required for the correct localisation of nuclei in the spermatids and nuclear shaping during elongation (Kimura 2013). GA27003 appeared to show patchy staining in late primary spermatocyte cysts and in the post-meiotic early spermatids. GA27003 may have higher expression in a subset of cysts in *D. pseudoobscura*.

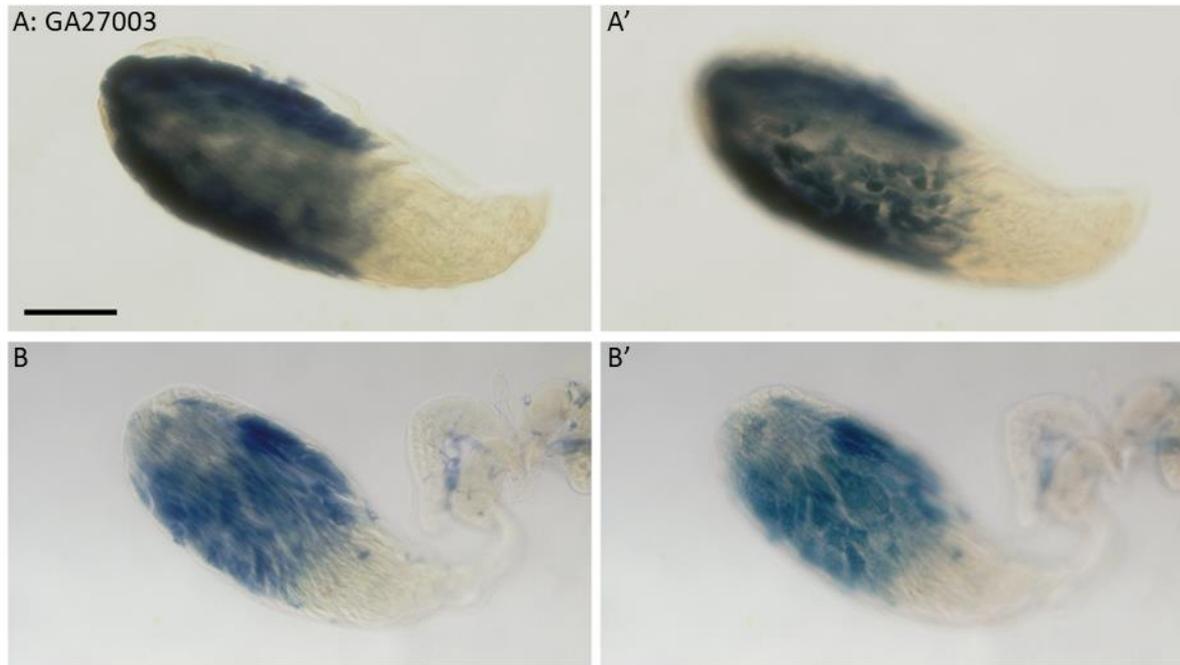


Figure 4.21: *In situ* hybridisation of GA27003, orthologous to *D. melanogaster* spermatogenesis gene *milka* (*mil*). Stained testes are shown at two different focal planes, showing surface and more internal staining. Left – right, apical – proximal. Scale = 100 μ m. A-B: Surface staining of GA27003. A'-B': Internal staining of GA27003.

4.5.7 Germline stem cell maintenance and proliferation genes

Germline stem cell maintenance is dependent on signalling from the testis hub activating the JAK-STAT pathway, with displacement from the hub reducing signal strength and allowing cell differentiation (Tulina and Matunis 2001; Herrera and Bach 2019).

chickadee (*chic*) is a highly conserved gene with multiple functions in development, including germ cell development. It is required for correct encapsulation of the germline by somatic cyst cells and formation of the permeability barrier, which in turn is essential for mediating signalling between the soma and the germline (Shields *et al.* 2014; Fairchild *et al.* 2015; Rockwell and Hongay 2020). The *D. pseudoobscura* *chic* orthologue, GA21874, was detected throughout the testis (Figure 4.22; A). Strong staining was observed in the hub region and spermatogonia, with staining decreasing in the spermatocyte regions. Staining was also present at the base of the testis, where mature sperm are released from the spermatid cysts prior to entry into the seminal vesicle. Staining at the base of the testis may be the result of somatic expression, rather than germline expression.

Daughters against dpp (*Dad*) encodes an inhibitory component of the BMP/Dpp signalling pathway (Attisano and Wrana 2000; Harris and Ashe 2011). The BMP/Dpp signalling pathway is involved in many processes throughout development (Hamaratoglu *et al.* 2014). In the testis, Dpp signalling is

required for maintenance of the germline stem cells, repressing expression of *bag of marbles (bam)*. Overexpression of *bam* results in the loss of the germline stem cells (Kawase *et al.* 2004). Staining for the *D. pseudoobscura* *Dad* orthologue, *GA18732*, showed uniform transcript detection throughout the pre-meiotic stages of sperm development (Figure 4.22; B). The *GA18732* transcript may have been present in the post-meiotic cysts, however it was not possible to determine by *in situ* hybridisation staining. Staining was also present at the base of the testis, which may indicate somatic expression.



Figure 4.22: *In situ* hybridisation of germline stem cell maintenance genes. Left – right, apical – proximal. Scale = 100 μ m. A: GA21874, chickadee (chic). B: GA18732, Daughters against dpp (*Dad*) isoform RA.

4.5.8 Post-meiotic expression in spermatid cysts

In situ hybridisation of several genes showed a characteristic staining pattern in the testis, of little to no staining around the edge of the testis, and darker staining in the centre of the testis. This staining pattern indicated that the transcript was present in the spermatid stages, post-meiosis. Furthermore, a lack of staining in the outer region of the testis indicated that these transcripts were not present prior to meiosis, suggesting post-meiotic expression. This was observed previously in *Mip40B* (Figure 4.16). A post-meiotic expression pattern was also observed in several genes orthologous to known *D. melanogaster* spermiogenesis genes (Figure 4.23).

D. melanogaster asl is required for correct duplication, localisation and elongation of the centriole in spermatids, on which formation of the basal body and axoneme is dependent (Blachon *et al.* 2008; Galletta *et al.* 2016). The *D. pseudoobscura* orthologue of *asterless (asl)*, *GA26457*, was not detected in the spermatogonia or spermatocyte stages, although expression may have been present in the stages immediately prior to meiosis (Figure 4.23; A). Staining did not appear to be present in the elongated spermatid cysts, indicating that the *GA26457* transcript may be limited to meiosis or immediately post-meiosis, consistent with a role in the formation of the meiotic spindle.

D. melanogaster dila is involved in the formation of the flagellum, with potential roles in transport of components required for flagellum and axoneme formation (Ma and Jarman 2011; Vieillard *et al.*

2016). The *Dilatory (dila)* orthologue, *GA14055*, was detected in the pre-meiotic stages, with light staining around the outer region of the testis. Stronger staining was detected in the post-meiotic stages, in the central region of the testis, disappearing at the basal region where mature sperm are released from the cyst cells (Figure 4.23; B). This indicates the presence of *GA14055* throughout the spermatid stage, consistent with a function in flagellum and axoneme assembly.

walker cup (wa-cup) has been previously shown to be expressed in post-meiotic spermatid cysts, localising at the distal elongating end of the spermatid cyst (Barreau *et al.* 2008a; Barreau *et al.* 2008b). This pattern was also observed in the *D. pseudoobscura* orthologue *GA10082*, with staining localised to the elongating spermatids. Staining appeared to have more intense ‘spots’ which appear to be localisation at the distal elongating end of the spermatid cysts (Figure 4.23; C and D).

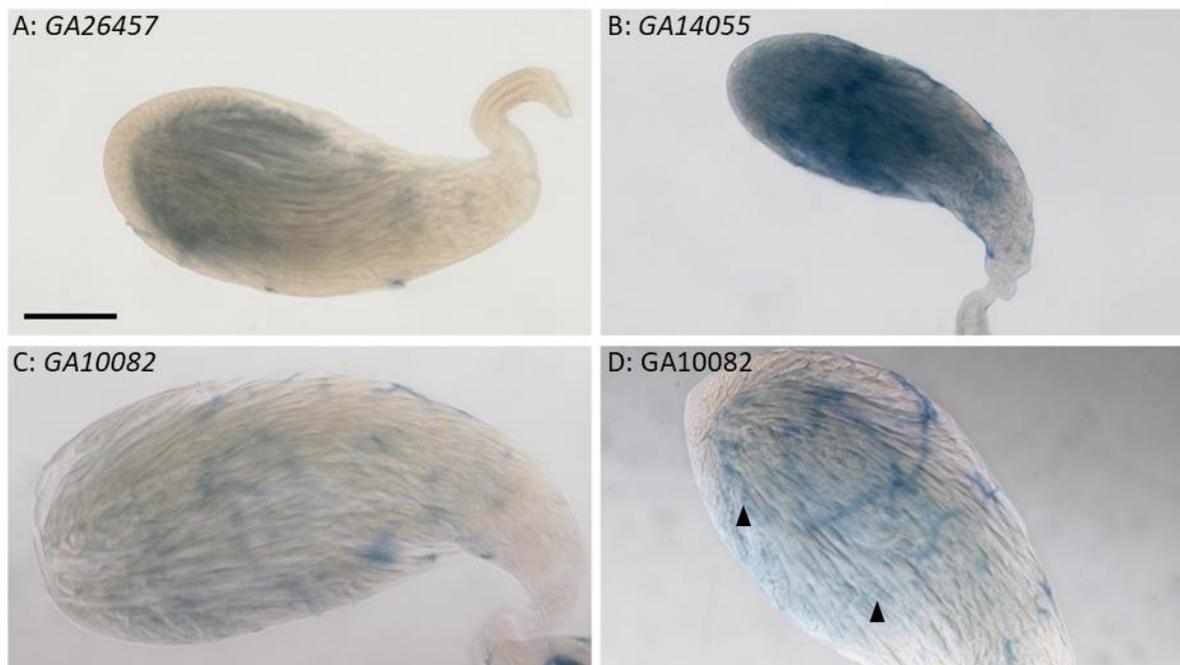


Figure 4.23: *In situ* hybridisation of spermiogenesis genes. Left – right, apical – proximal. Scale = 100 μ m. A: *GA26457*, *asterless (asl)*. B: *GA14055*, *dilatory (dila)*. C-D: *GA10082*, *walker cup (wa-cup)*. Black arrowheads indicate stronger staining at the distal ends of elongating spermatid cysts.

4.5.9 Splicing Regulators

There is some evidence of isoform-dependent differential gene expression in *D. pseudoobscura* spermatogenesis, notably *GA21345 (achi/vis)* (Figure 4.14). *D. melanogaster pre-mRNA processing factor 8 (prp8)* is a highly conserved splicing factor, predicted to have a function in the G2/M transition in spermatocytes (Wu *et al.* 2016). The *D. pseudoobscura prp8* orthologue, *GA21384*, was detected throughout the testis, with a band of strong staining around the spermatogonia to primary

spermatocyte region, indicating increased expression at this point (Figure 4.24; A). *GA21384* may have a role in the splicing of genes expressed in *D. pseudoobscura* primary spermatocytes.

The *D. pseudoobscura* orthologue of *U2A*, *GA12730* did not show staining higher than background (Figure 4.24; B). *U2A* was identified as significantly differentially expressed in the second spermatocyte and spermatid RNA-seq data. *In situ* hybridisation staining may reflect that *GA12730* expression is low, and not detected by this method, or may indicate that *GA12730* does not contribute to splicing in *D. pseudoobscura* testes.

D. melanogaster Splicing factor 3b subunit 5 (*Sf3b5*) is a splicing factor required for spermatogenesis. Similar to *prp8*, it is predicted to function in the G2/M transition in spermatocytes (Wu *et al.* 2016). *GA25581*, the *D. pseudoobscura* orthologue of *Sf3b5*, was not detected in spermatocytes, as the *D. melanogaster* function might predict. *GA25581* showed a similar post-meiotic staining pattern to that observed in other post-meiotic genes (Figure 4.24; C). *GA25581* may contribute to splicing of mRNA expressed post-meiosis in *D. pseudoobscura*.

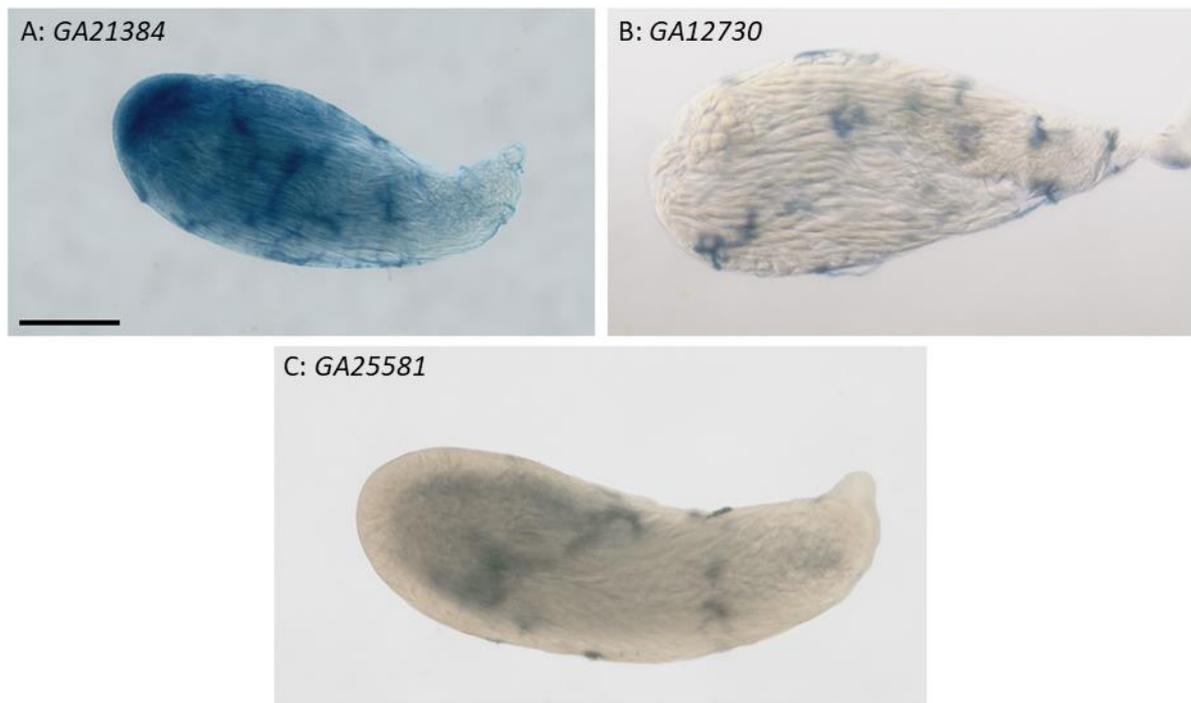


Figure 4.24: *In situ* hybridisation of spliceosome genes. Left – right, apical – proximal. Scale = 100 μ m. A: *GA21384*, pre-mRNA processing factor 8 (*prp8*). B: *GA12730*, *U2A*. C: *GA25581*, Splicing factor 3b subunit 5 (*Sf3b5*).

4.5.10 Other genes with known testis functions

D. melanogaster exuperantia has a role in mRNA localisation in male and female germline cells (Crowley and Hazelrigg 1995). *D. pseudoobscura exuperantia 2 (exu2)* was found to be significantly

differentially expressed in all spermatocyte and spermatid RNA-seq datasets. *exu2* was not detected in the hub or spermatogonia regions of the testis by *in situ* hybridisation. Strong staining was observed in the primary spermatocyte cysts, which may have been stronger in some cysts, indicating that there may be differential expression at this stage. Transcript detection decreased in later stage spermatocytes, with no detection in post-meiotic cysts (Figure 4.25; A).

D. melanogaster Nucleoporin 154kD (*Nup154*) has functions in nuclear localisation of the BMP signalling pathway, and is required for the formation of cysts, control of mitotic divisions of the spermatogonia and entry into the spermatocyte phase (Colozza *et al.* 2011). *Nup154* is also required for localisation of nucleoporins within the nuclear pore complex at the nuclear envelope (Gigliotti *et al.* 1998). The *D. pseudoobscura* *Nup154* orthologue, *GA18272*, was detected in the spermatogonia and spermatocytes, but was not detected in post-meiotic cysts. *GA18272* may have a similar function in *D. pseudoobscura* in regulation of mitosis and entry into meiosis, and the organisation of the nuclear pore complex in the germline and somatic cyst cells.

D. melanogaster P-element induced wimpy testis (*piwi*) is required for the regulation of spermatogenesis and silencing transposons (Ku *et al.* 2016). *GA19370*, orthologous to *D. melanogaster piwi* was detected in the spermatid cysts, although detection was weak. *GA19370* appeared more localised within the post-meiotic cyst region of the testis compared to other than other post-meiotic transcripts, for example *mip40B*, indicating that it may be expressed only, or more strongly, in longer spermatid cysts.

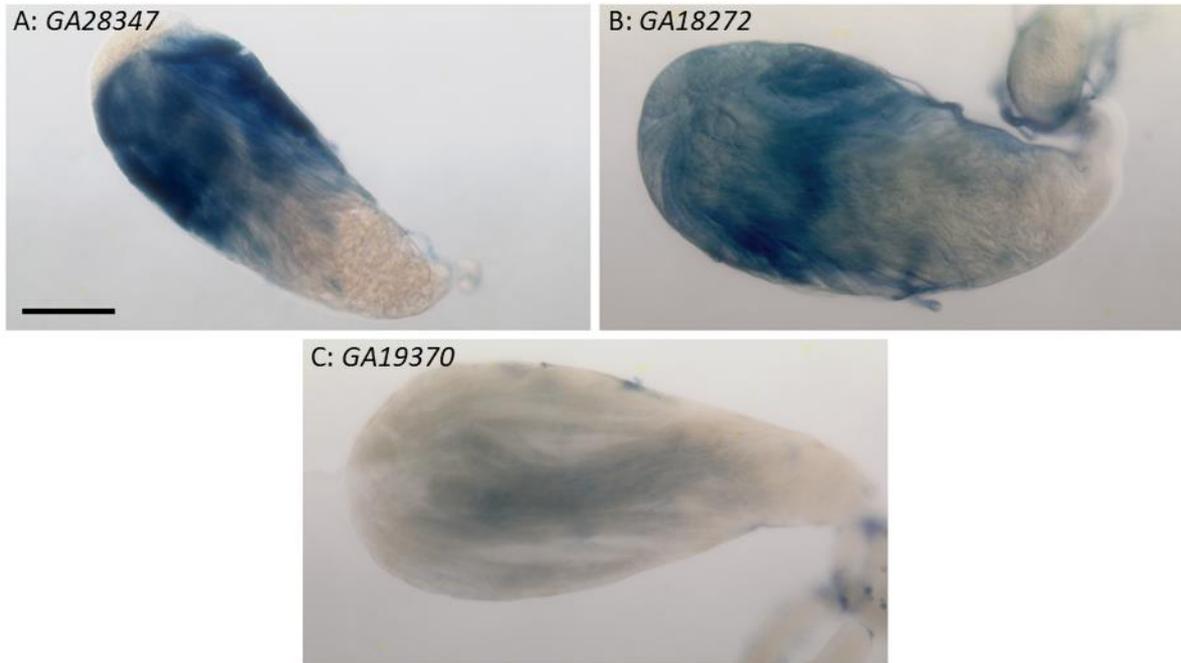


Figure 4.25: A: GA28347, exuperantia 2 (*exu2*). B: GA18272, Nucleoporin 154 (*Nup154*). C: GA19370, P-element induced wimpy testis (*piwi*). Left – right, apical – proximal. Scale = 100 μ m.

4.5.11 Unknown functions

GA23025 is a *D. pseudoobscura* gene also with no known *D. melanogaster* orthologue. Previously supported models have predicted that GA23025 is an orthologue of *mil*, although this model is no longer supported (Larkin *et al.* 2021). The GA23025 protein is predicted to have a nucleosome assembly protein (NAP) domain, and has been previously found to have testis specific expression (Gramates *et al.* 2017). GA23025 was detected in the spermatocyte cysts, with stronger staining in a subset of cysts (Figure 4.26). GA23025 may be differentially expressed in *D. pseudoobscura* primary spermatocytes.

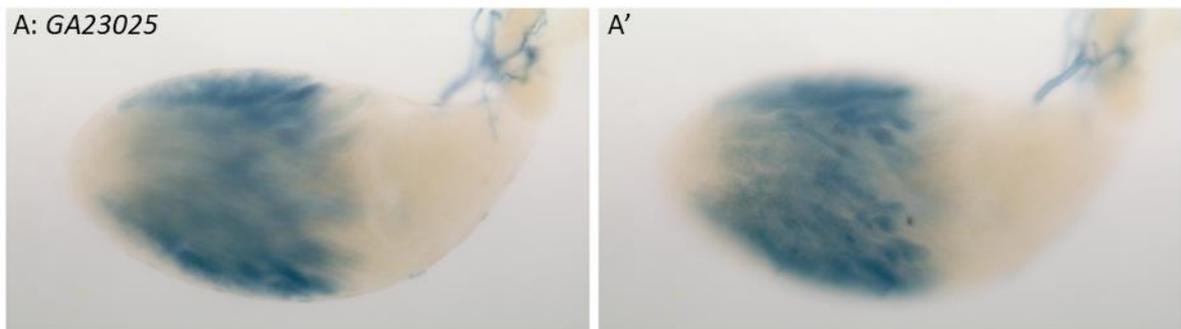


Figure 4.26: In situ hybridisation of GA23025, a *D. pseudoobscura* gene with no known orthologue in *D. melanogaster*. Stained testes are shown at two different focal plains, showing surface and more internal staining. Left – right, apical – proximal. Scale = 100 μ m. A: Surface staining of GA23025. A': Internal staining of GA23025.

4.6 Summary

The aim of this chapter was to investigate whether spermatocyte cysts have transcriptional variation which may be associated with morph-specific development in the spermatid cyst stage, and to characterise transcriptomes of both spermatocyte and spermatid cysts. Cluster analysis of spermatocyte cysts was used to analyse grouping based on transcription, and differential gene expression analysis analysed data to identify genes for which there were significant differences in transcript abundance. For each spermatocyte cyst data set, over 1400 genes were identified as significantly differentially expressed between two groups.

Candidate genes were identified from the RNA-seq datasets, for further analysis by RNA *in situ* hybridisation. Components of the testis meiotic arrest complex were also investigated by *in situ* hybridisation, as they are known to have an essential role in transcriptional control in *D. melanogaster* spermatogenesis. *In situ* hybridisation revealed further evidence of transcriptional variation between *D. pseudoobscura* spermatocyte cysts, providing further support for the presence of multiple spermatocyte cyst populations, each contributing to one of the three mature sperm morphs. A summary of genes identified by RNA-seq and followed up with *in situ* hybridisation is available in Table 4.12.

In situ hybridisation of a subset of the candidate genes showed a ‘patchy’ staining pattern, indicating differential transcript abundance in subsets of spermatocyte cysts. Of these, the most striking were the genes *kumgang*, *achi/vis* and *Caf1A*, all of which have been shown to have essential functions in transcription control in *D. melanogaster*. *kmg* was selected for further study, by insertion of a tagged copy of *kmg* by the piggyBac transposon vector.

This transcriptional variation between distinct subsets of cysts suggests that morph differentiation is underway in the spermatocyte stage. Prior to this work, it had been hypothesised that eusperm and parasperm morph differentiation would take place before meiosis (Moore *et al.* 2013); RNA-seq and *in situ* hybridisation data presented in this chapter has provided the first evidence of this differentiation.

<i>D. pseudoobscura</i> Gene	<i>D. melanogaster</i> Orthologue	RNA-seq DGE Analysis	<i>In situ</i> Staining Patchy Expression	<i>In situ</i> Staining	Post meiotic expression
GA10082	<i>wa-cup</i>	Spermatocyte	Maybe	Spermatid (distal ends)	Yes

GA11638	<i>wuc</i>	Spermatocyte	No	Hub Spermatogonia Spermatocyte	No
GA12326	<i>comr</i>	None	No	Spermatogonia Spermatocyte Spermatid	No
GA12700	<i>tomb</i>	None	No	Spermatogonia Spermatocyte	No
GA12730	<i>U2A</i>	Spermatocyte Spermatid	No Staining	No staining	No
GA13507	<i>mip40A</i>	Spermatocyte	No	Spermatogonia Spermatocyte	No
GA14055	<i>dila</i>	Spermatocyte	No	Hub Spermatogonia Spermatocyte Spermatid	Yes
GA17585	<i>stc</i>	Spermatocyte	No	Spermatogonia Spermatocyte	No
GA18051	<i>Caf1a</i>	None	Yes	Hub Spermatogonia Spermatocyte	No
GA18272	<i>Nup154</i>	Spermatocyte	No	Hub Spermatogonia Spermatocyte	No
GA18412	<i>bol</i>	Spermatocyte Spermatid	Maybe	Spermatogonia Spermatocyte	No
GA18558	<i>twe</i>	Spermatocyte	Maybe	Spermatogonia Spermatocyte	No
GA18636	<i>croc</i>	Spermatocyte	No	Hub Spermatogenesis Spermatocyte	No
GA18732	<i>Dad</i>	Spermatocyte	No	Hub Spermatogonia Spermatocyte	No

				Somatic	
GA18735	<i>kmg</i>	Spermatocyte	Yes	Spermatocyte	No
GA19264	<i>woc</i>	Spermatocyte	No	Hub Spermatogonia	No
GA19370	<i>piwi</i>	Spermatocyte	Yes	Spermatid	Yes
GA20060	<i>Dic61B</i>	Spermatocyte	Yes	Spermatocyte Spermatid	No
GA21328	<i>Rcd-1</i>	None	No	Spermatogonia Spermatocyte Spermatid	No
GA21345	<i>achivis</i>	Spermatocyte	Yes	Hub Spermatogonia Spermatocyte	No
GA21384	<i>prp8</i>	Spermatocyte	No	Hub Spermatogonia Spermatocyte Spermatid	No
GA21874	<i>chic</i>	Spermatocyte	No	Hub Spermatogonia Spermatocyte Somatic	No
GA23025	.	Spermatocyte Spermatid	Yes	Spermatocyte	No
GA23669	<i>mip40B</i>	None	No	Spermatid	Yes
GA25581	<i>Sf3b5</i>	Spermatocyte	No	Spermatid	Yes
GA26389	<i>Caf1b</i>	None	No	Hub Spermatogonia	No
GA26457	<i>asl</i>	Spermatocyte Spermatid	No	Late spermatocyte or early spermatid	Maybe
GA27003	<i>mil</i>	Spermatocyte	Yes	Spermatocyte Spermatid	No
GA27927	<i>tbrd-2</i>	Spermatocyte	No	Spermatogonia Spermatocyte	No

GA28313	<i>aly</i>	None	No	Hub Spermatogonia Spermatocyte Spermatid	No
GA28347	<i>exu2</i>	Spermatocyte Spermatid	No	Spermatocyte	No

Table 4.12: Summary of gene expression analysis in *D. pseudoobscura* testes. A list of candidate genes was identified by RNA-seq analysis of spermatocyte and spermatid cysts. Candidate genes were tested for testis expression by RNA in situ hybridisation, and genes of interest identified based on staining patterns indicating differential expression in spermatocyte or spermatid cysts.

5 Localisation of GFP-Tagged Kumgang in *D. pseudoobscura* Spermatogenesis

5.1 Protein Sequence Analysis of *kumgang*, a Regulator of Germline-Specific Gene Expression

Kumgang (Kmg) is expressed in the spermatocytes of *D. melanogaster*, and has functions in the regulation of germline specific gene expression in *D. melanogaster*. Acting with the chromatin remodeller dMi-2, Kmg blocks expression of normally somatic-expressed genes by preventing Aly binding at cryptic promoters. Knockdown of Kmg in *D. melanogaster* results in enrichment of Aly at Kmg binding sites, an increase in somatic genes expressed in the spermatocytes, and meiotic arrest (Kim *et al.* 2017). Its function in *D. pseudoobscura* is not well understood, however RNA-seq and *in situ* hybridisation for the *kmg* orthologue, GA18735, suggests that it is differentially expressed in primary spermatocyte cysts.

D. pseudoobscura Kmg protein sequence was aligned to *D. melanogaster* Kmg protein sequence by Protein BLAST (Altschul *et al.* 1997; Altschul *et al.* 2005). The protein alignment showed *D. pseudoobscura* Kmg shared 59% amino acid identities with *D. melanogaster* Kmg (Figure 5.1; A). Analysis of *D. pseudoobscura* protein sequences by InterPro revealed seven predicted C2H2 DNA binding sites. These sites appear to be conserved, suggesting a similar function to that of *D. melanogaster* Kmg (Figure 5.1; B). Combined, the analyses suggest that the Kmg protein sequence is more highly conserved in the DNA binding regions.

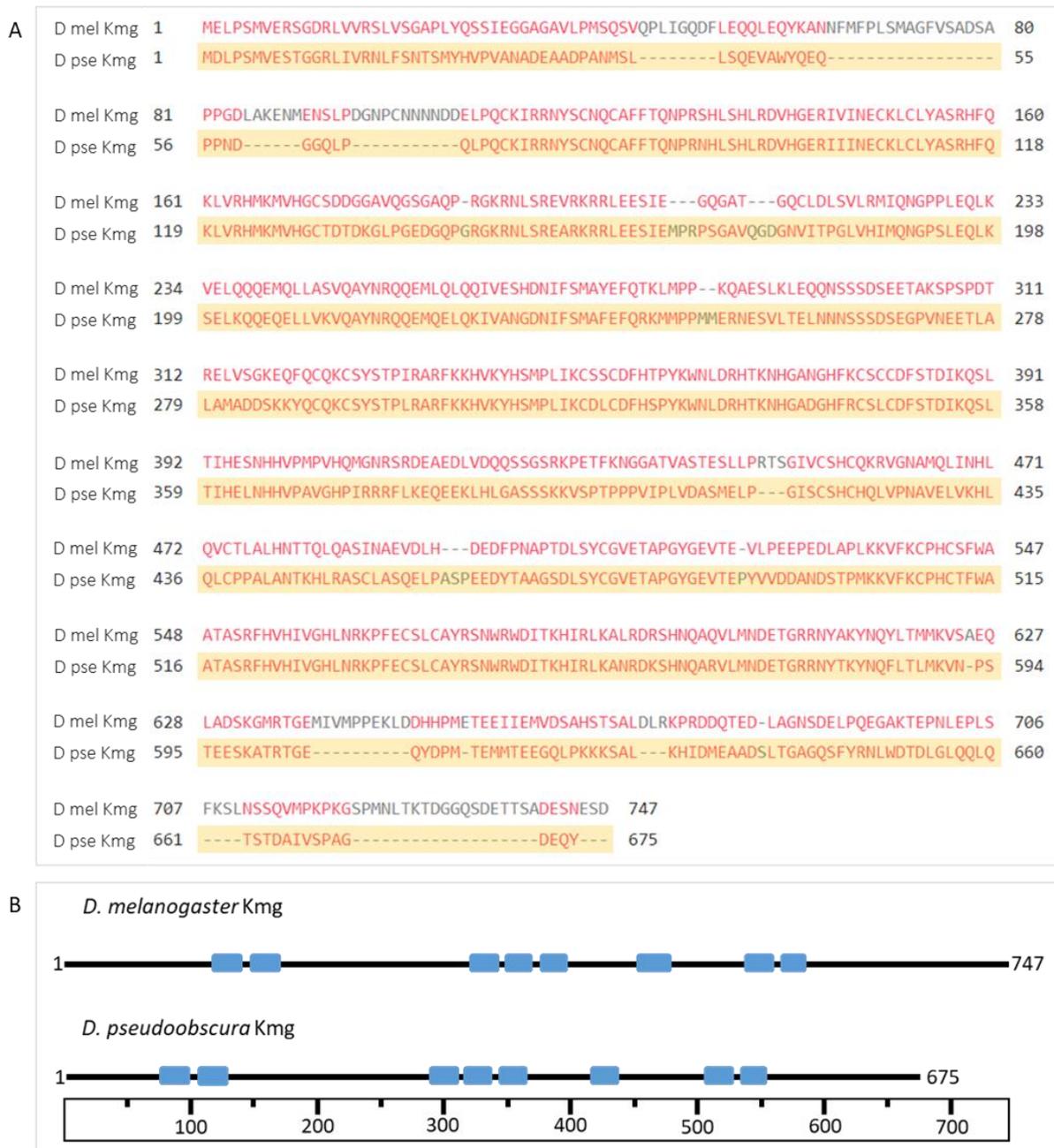


Figure 5.1: Alignment and analysis of *D. melanogaster* and *D. pseudoobscura* Kmg protein sequences. A: *D. melanogaster* and *D. pseudoobscura* Kmg protein sequence alignment. Red indicates aligned sequence. Grey indicates non-aligned sequence. *D. pseudoobscura* sequence is highlighted. Alignment performed by BLAST P (Altschul et al. 1997; Altschul et al. 2005). B: Predicted domains of *D. melanogaster* and *D. pseudoobscura* Kmg proteins. Black lines indicate protein. Blue boxes indicate C2H2 DNA binding domains. Scale indicates amino acid position. Protein domain analysis performed with InterPro (Blum et al. 2021).

5.2 Assembly of piggyBac-3xP3-AmCyan-Kmg-GFP for Genomic Insertion of Fluorescently Tagged *D. pseudoobscura* Kumgang

To further examine the localisation and potential roles of Kmg in *D. pseudoobscura* spermatogenesis, piggyBac constructs were designed for the insertion of GFP-tagged *kmg* into the genome. piggyBac was selected as it was previously used for insertion of the *cas9* sequence into *D. pseudoobscura*. C-terminal and N-terminal GFP tag constructs were designed and assembled.

Synthesised DNA was used as a template for PCR to produce *kmg* fragments which were cloned in to pGEM-T Easy by TA cloning. GFP PCR products were also cloned into pGEM-T Easy. pGEM-T Easy constructs were sequenced and sequences aligned to predicted sequences (Figure 5.2, Figure 5.3). The C-terminal tagged construct consisted of the piggyBac vector, the *kmg* promoter-5' UTR-CDS sequence, the GFP sequence, and *kmg* 3' UTR sequence (Figure 5.4; A). The N-terminal tagged construct was not completed; the piggyBac vector, *kmg* promoter-5' UTR sequence, and GFP sequence were assembled, but the CDS and 3'UTR was not successfully ligated into the final construct (Figure 5.4; B).

Assembled constructs were sequenced (Eurofins), to ensure assembly in the correct orientation. The presence of mutations was also assessed. Sequencing data were aligned to the predicted construct sequence (Figure 5.4).

Four mutations were present in the *kmg* coding sequence: two in Exon 1 and two in Exon 2 (Figure 5.5). All four mutations were non-synonymous, but none resulted in a frame shift or stop codon. Mutations did not affect any of the predicted DNA binding sites.

Full sequences of *kmg* and GFP components in pGEM-T Easy and piggyBac can be found in Appendix 10.

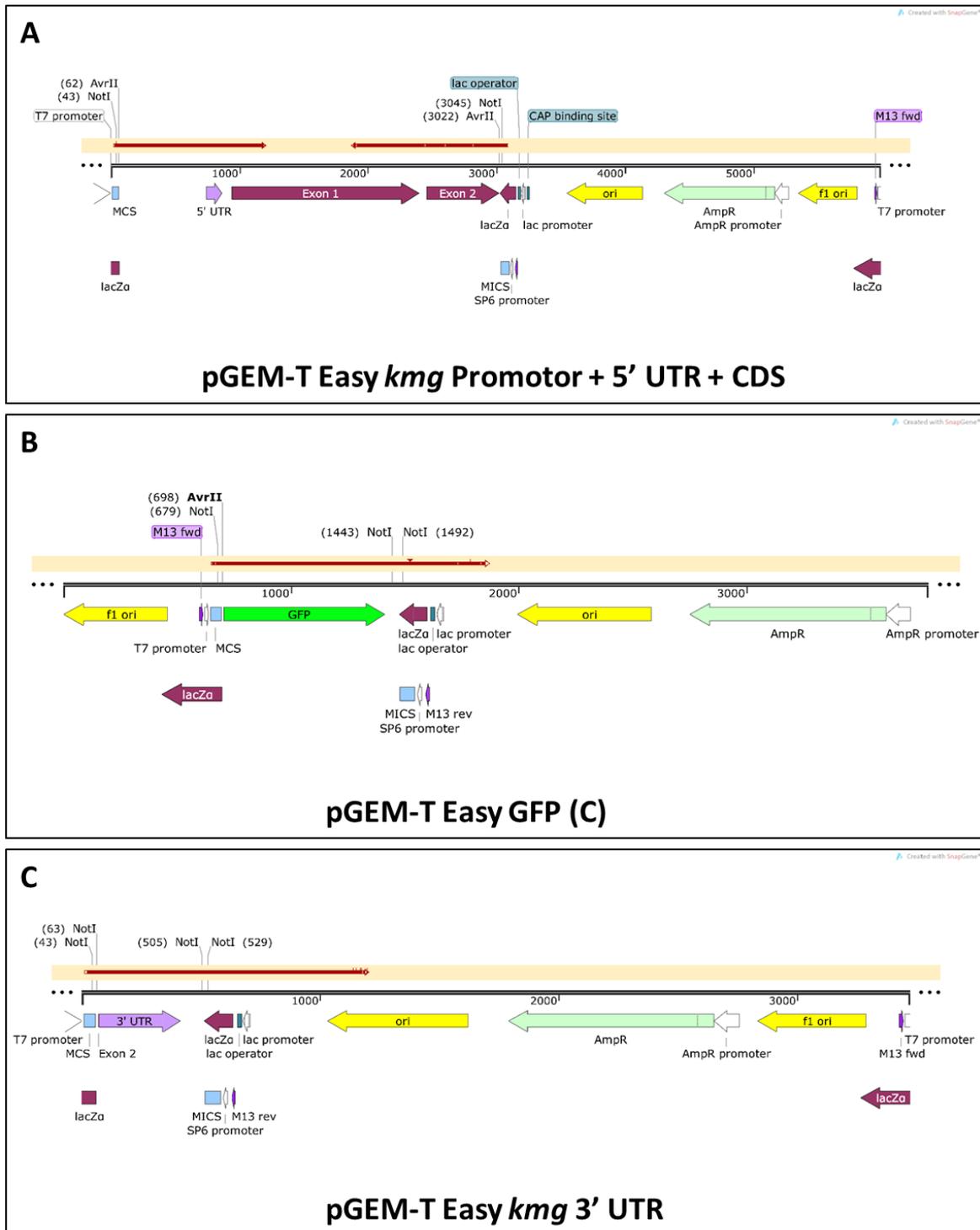


Figure 5.2: Kungang-GFP C-terminal tag construct components in pGEM-T Easy, for subsequent ligation into piggyBac vector. Sequencing results shown in red (Eurofins), aligned to predicted sequence. A: *kmg* promotor+5' UTR+CDS in pGEM-T Easy with T7 and SP6 sequencing aligned. B: GFP (C) in pGEM-T Easy with T7 sequencing aligned. C: *kmg* 3' UTR in pGEM-T Easy with T7 sequencing aligned.

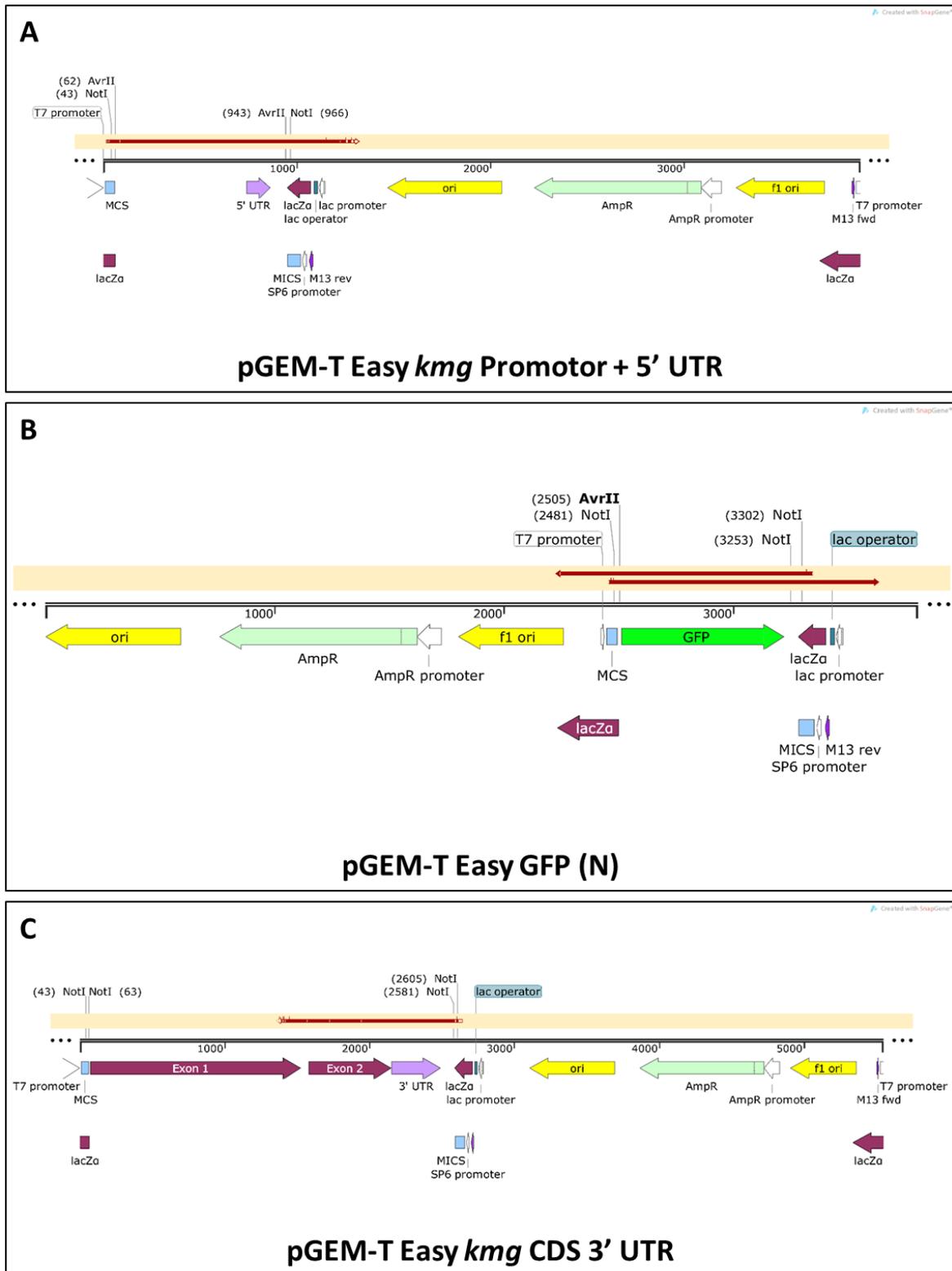


Figure 5.3: Kungang-GFP N-terminal tag construct components in pGEM-T Easy, for subsequent ligation into piggyBac vector. Sequencing results shown in red (Eurofins), aligned to predicted sequence. A: kmg promotor+5' UTR in pGEM-T Easy with T7 sequencing aligned. B: GFP (N) in pGEM-T Easy with T7 and SP6 sequencing aligned. C: kmg CDS+3' UTR in pGEM-T Easy with T7 sequencing aligned. Sequencing was not successful with the SP6 primer.

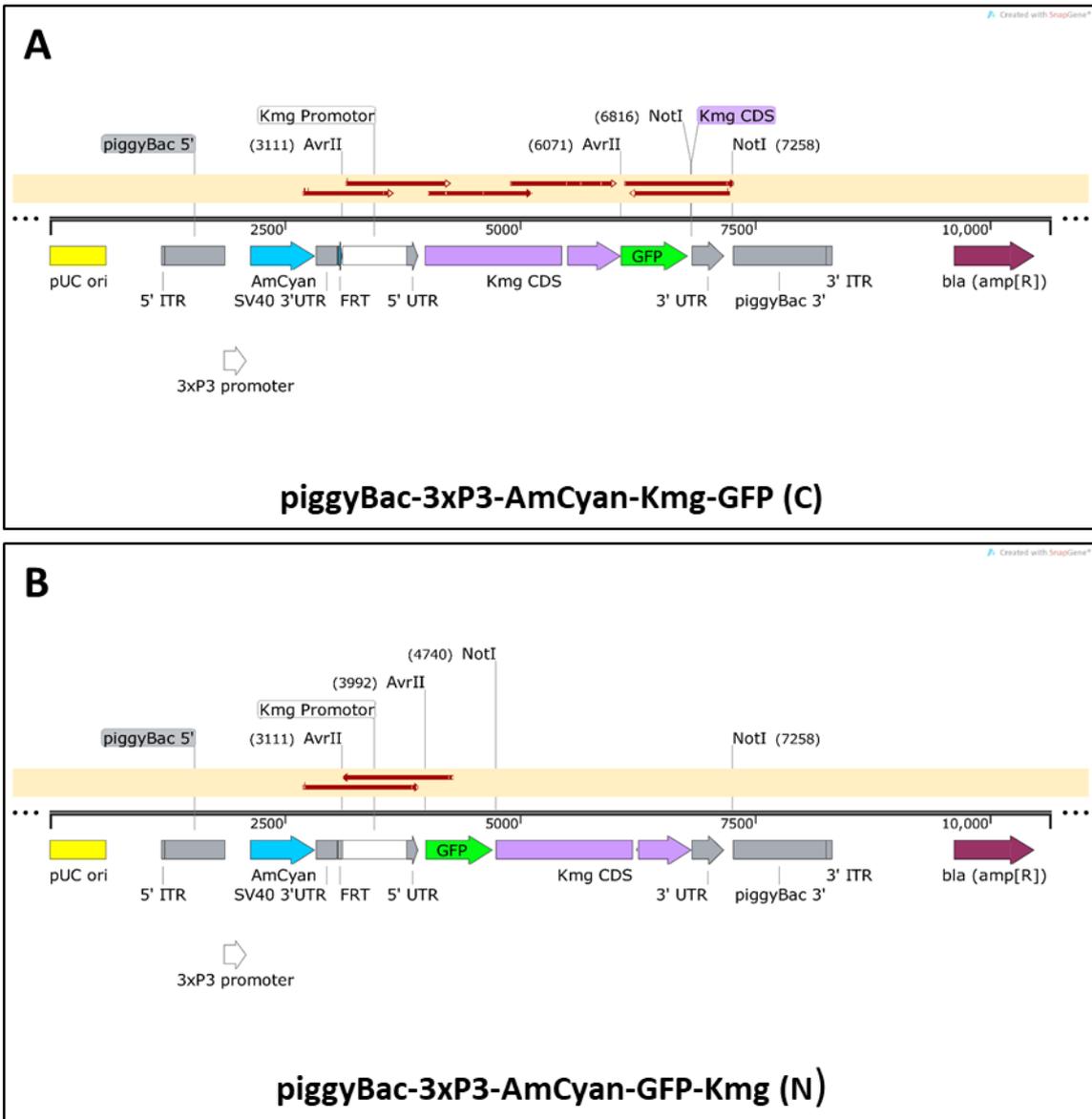


Figure 5.4: Construct assemblies and sequence alignments to predicted construct sequence for piggyBac-3xP3-AmCyan-Kmg-GFP C- and N-terminal constructs. Aligned sequences in red.

Amino Acid Position	Codon Mutation	Amino Acid
74	AAC>AGC	Asn>Ser
207	GAG>GGG	Glu>Gly
535	GAG>GGG	Glu>Gly
608	GAT>GTT	Asp>Val
1 1	MDLPSMVESTGGRLIVRNLFNNTSMYHVPVANADEAADPANMSLLSQEVAWYQEPPNDG	60 60
61 61	GQLPQLPQCKIRRNYSCNQCAFFTONPRNHLSHLRDVHGERIINECKLCLYASRHFQKLS.....	120 120
121 121	VRHMKMVHGCTDSDKGLPGEDGQPRGKRNLREARKRRLEESIEMPRPSGAVQGDGNVI	180 180
181 181	TPGLVHIMQNGPSLEQLKSELKQQEQELLVKVQAYNRQQEMQELQKIVANGDNIFSMAFEG.....	240 240
241 241	FQRKMMPPMMERNESVLTELNNNSSDSEGPVNEETLALAMADDSKKYQCQKCSYSTPLR	300 300
301 301	ARFKKHVKYHSMPLIKCDLDFHSPYKWNLDLDRHTKNHGADGHFRCSLDFSTDIKQSLTI	360 360
361 361	HELNHVPAVGHPIRRRFLKEQEEKLHLGASSSKKVSPTPPPVIPLVDASMELPGISCSH	420 420
421 421	CHQLVPNAVELVKHLQLCPPALANTKHLRASCLASQELPASPEEDYTAAGSDLSYCGVET	480 480
481 481	APGYGEVTEPYVDDANDSTPMKKVFKCPHCTFWAATASRFVHIVGHLNRKPFECSLCAG.....	540 540
541 541	YRSNWRWDITKHIRLKANRDKSHNQARVLMNDETGRNRYTKYNQFLTMKVNVPSTEEKSA	600 600
601 601	TRTGEQYDPMTEMMTEEGQLPKKKSALKHIDMEAADSLTGAGQSFYRNLWDTDLGLQLLQV.....	660 660
661 661	TSTDAIVSPAGDEQY 675 675	

Figure 5.5: Mutations present in *D. pseudoobscura* kmg sequence for piggyBac-3xP3-AmCyan-Kmg-GFP C terminal tag construct. Non-synonymous mutations are shown as red.

5.3 Collection of Injection Survivors, Crosses and Screening for Transgenic F1 and F2

D. pseudoobscura SLOB3 were injected with the piggyBac-3xP3-AmCyan-Kmg-GFP and piggyBac Helper constructs. 21 surviving larvae were collected after injection. Of the 21 surviving larvae, a total of 7 female and 5 male F0 adults were collected post-eclosion.

F0 adults were crossed with *D. pseudoobscura* SLOB3 wild-type. Of the F0 crosses, one produced transgenic F1 offspring (Table 5.1), all of which were female. Recovered transgenic F1 females likely contained the same insertion. It is likely that the piggyBac insertion is on the X chromosome in the transgenic F1 offspring.

F0 Adult	M/F	F1 Offspring (No Transgene)	F1 Transgenic Males	F1 Transgenic Females
A	F	11	0	0
B	M	116	0	7
C	M	152	0	0
D	F	0	0	0
E	F	61	0	0
F	M	114	0	0
G	F	39	0	0
H	F	87	0	0
I	F	1	0	0
J	F	29	0	0
K	M	50	0	0
L	M	123	0	0

Table 5.1: *D. pseudoobscura* F0 adult survivors of piggyBac-3xP3-AmCyan-Kmg-GFP injection and offspring from F0 crosses with *D. pseudoobscura* SLOB3 wild-type. F1 adults were screened for AmCyan eyes, indicating the presence of the piggyBac insertion.

D. pseudoobscura piggyBac-3xP3-AmCyan-Kmg-GFP F1 females were crossed with wild-type males. The F2 offspring were collected and screened. Female F1 crosses produced male and female transgenic offspring, further indicating that the piggyBac insertion is on the X chromosome, and is neither male lethal nor dominant female sterile (Table 5.2).

F2 males were collected and crossed with F2 transgenic females, or were dissected for imaging of cysts.

F1 Xs	Cyan Females (Transgene)	Cyan Males (Transgene)	Females (No Transgene)	Males (No Transgene)
F01	25	29	26	22
F02	32	29	33	37
F03	5	3	1	4

F04	0	0	0	0
F05	26	26	31	27
F06	19	21	20	19
F07	20	20	18	13

Table 5.2: F2 offspring from D. pseudoobscura piggyBac-3xP3-AmCyan-Kmg-GFP transgenic F1 adults crossed with SLOB3 wild-type. F2 adults were screened for AmCyan eyes, indicating the presence of the piggyBac insertion.

5.4 Expression of GFP-Tagged Kumgang in Spermatocyte Cysts

Testes were dissected from F2 males with the cyan fluorescence eye marker. The testes were cut near the base with a tungsten needle and the cysts emptied in to testis buffer. Cysts were imaged with phase contrast and fluorescence microscopy. Cysts of all stages were imaged.

5.4.1 Primary spermatocyte cysts show differential expression of the GFP-tagged Kmg protein

Primary spermatocyte cysts imaged under DIC and fluorescence showed varying levels of GFP fluorescence (Figure 5.6). Some cysts showed GFP fluorescence, indicating expression of the GFP-tagged Kmg protein (Figure 5.6; C and D). Some cysts had no detectable GFP fluorescence, indicating low levels of Kmg (Figure 5.6; E). Where GFP expression was observed, there appeared to be two levels of GFP fluorescence intensity, indicating that there may be two levels of Kmg-GFP expression within the spermatocyte cysts (Figure 5.6; A and D).

GFP fluorescence was localised to the nuclei of the germline cells, indicating the presence of Kmg within the nucleus. Figure 5.7 shows both Hoechst 33342, indicating DNA, and Kmg fluorescence within the spermatogonia and spermatocytes. Hoechst staining showed dispersed chromatin in the nuclei of a 16-cell spermatogonia cyst (Figure 5.7; A) and as condensed chromosomes prior to meiotic division in a late primary spermatocyte cyst (Figure 5.7; B). Kmg did not appear to be localised to the dispersed or condensed DNA, but was distributed throughout the nucleus.

There was some indication that Kmg was absent from the nucleolus. A darker 'spot' in the nucleus region of the spermatocytes indicates the position of the nucleolus. This 'spot' also appears in GFP images, indicating lower or a lack of GFP localised to this area (Figure 5.7; B, yellow arrowheads).

The nuclei of cyst cells, which encapsulate the developing germline cells, did not exhibit GFP fluorescence. The white arrows in Figure 5.7; B indicate the location of the cyst cell nucleus, which are seen in phase contrast and Hoechst images, but are not visible in GFP images. Kmg does not appear to be present in the somatic cyst cell nuclei; within the developing cyst it is specific to the nuclei of the germline cells.

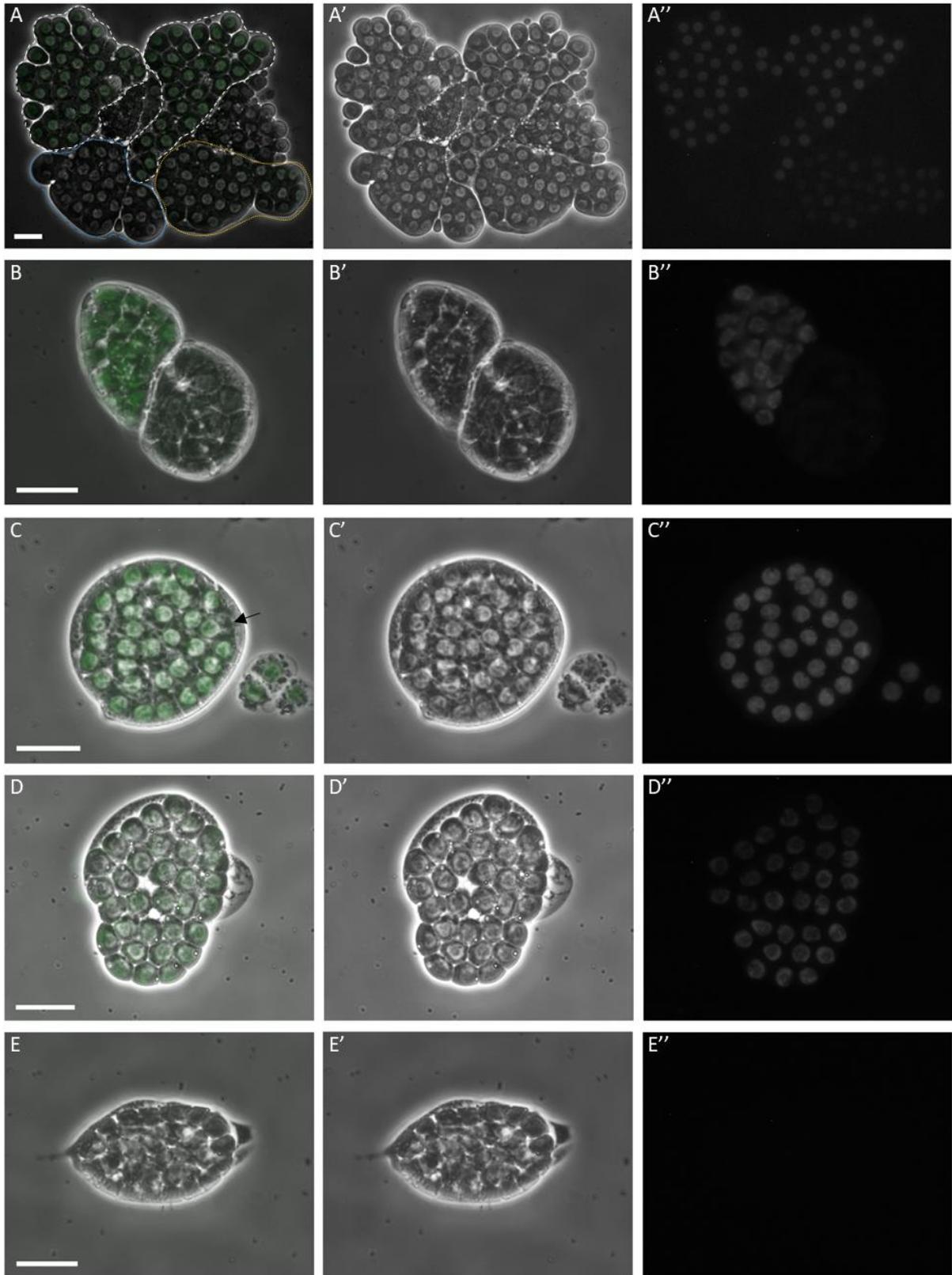


Figure 5.6: Primary spermatocyte cysts dissected from *D. pseudoobscura* SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735). Scale = 20 μ m. A: Five primary spermatocyte cysts, lysed from cyst cells. Two

spermatocyte cysts show high GFP fluorescence (white dashed lines). One spermatocyte cyst shows GFP fluorescence, but at a lower intensity (yellow dotted line). Two spermatocyte cysts do not show GFP fluorescence (one indicated with blue dashed line). B: Two spermatocyte cysts, one showing high GFP fluorescence, one showing no GFP fluorescence, indicating high and low Kmg protein, respectively. C: Spermatocyte cyst showing high GFP fluorescence. Arrow indicates cyst cell nucleus. Cyst cells do not exhibit GFP fluorescence. D: Spermatocyte cyst showing GFP fluorescence of lower intensity than C. E: Spermatocyte cyst showing no GFP fluorescence. A-E: Combined phase contrast and GFP images. A'-E': Phase contrast images. A''-E'': Single channel GFP.

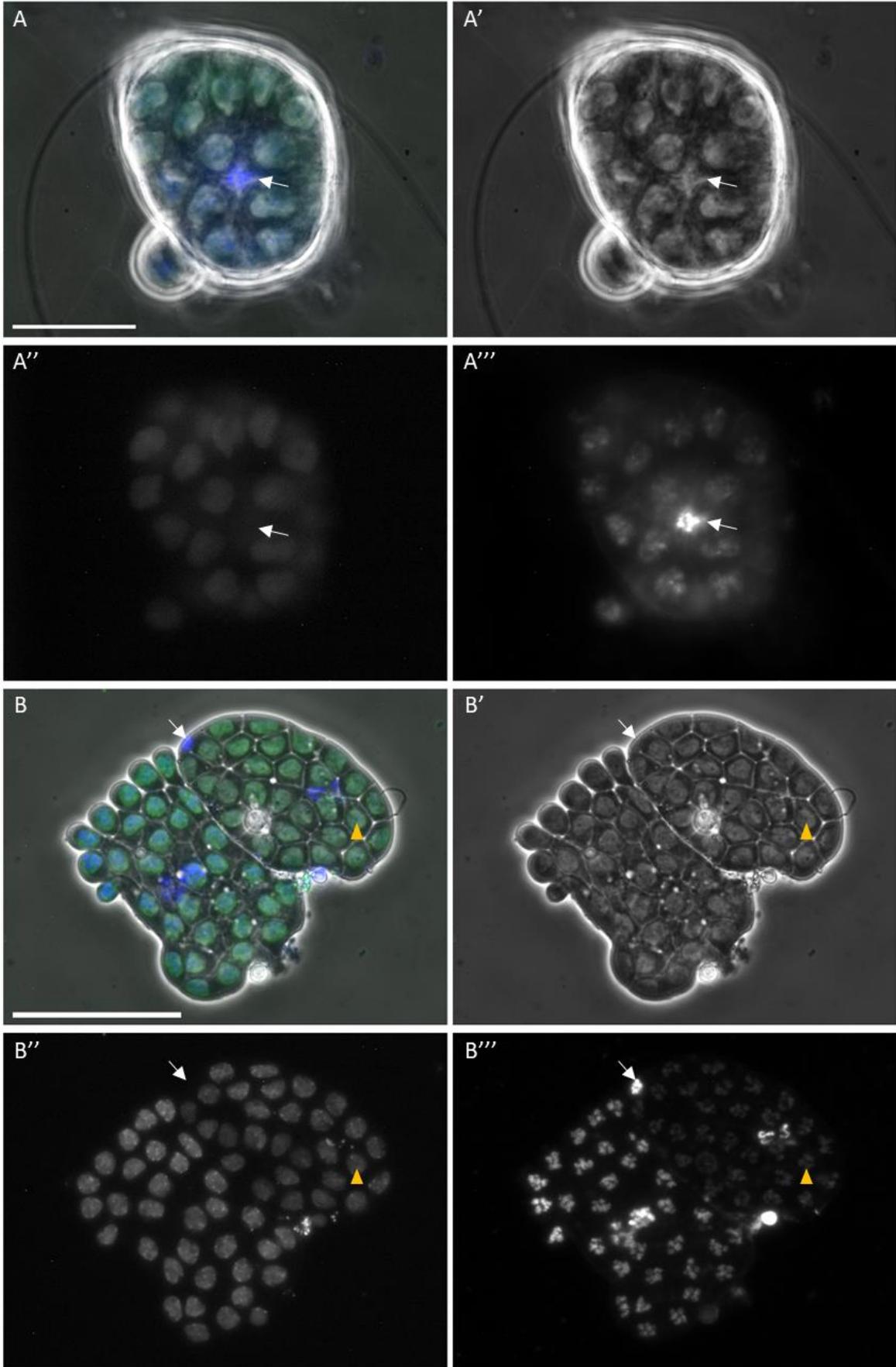


Figure 5.7: Primary spermatocyte cysts dissected from *D. pseudoobscura* SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735) (green) and Hoechst 33342 indicating DNA (blue). White arrows indicate location of cyst cell nucleus. Scale = 50 μ m. A: 16-cell spermatogonia cyst exhibiting GFP fluorescence indicating the presence of Kmg. Chromatin is dispersed. B: Two spermatocyte cysts exhibiting GFP fluorescence indicating high presence of Kmg. Yellow arrowheads indicate location of nucleolus. Chromatin is in condensed state. A-B: Combined phase contrast, GFP and Hoechst 33342 images. A'-B': Phase contrast images. A''- B'': Single channel GFP. A'''-B''': Single channel Hoechst 33342.

5.4.2 Kmg is translated during the mitotic divisions and is maintained at a high level in some, but not all cysts

Figure 5.8 and Figure 5.9 show spermatogonia and spermatocyte stages. Figure 5.8 shows early- and late-stage spermatogonia cysts and early spermatocyte cysts exhibiting no detectable GFP fluorescence, indicating that Kmg-GFP expression is low or not present in these stages of at least one morph. Enhanced brightness shows only low levels of background fluorescence detectable in these cysts.

Figure 5.9 shows spermatogonia and spermatocytes exhibiting GFP fluorescence, indicating that Kmg-GFP expression is maintained from the spermatogonia stage in at least one morph. GFP fluorescence appeared to increase in later-stage cysts, at the spermatocyte stage, indicating that Kmg-GFP expression is not just maintained but increased over the course of spermatogenesis, in morphs where Kmg is expressed.

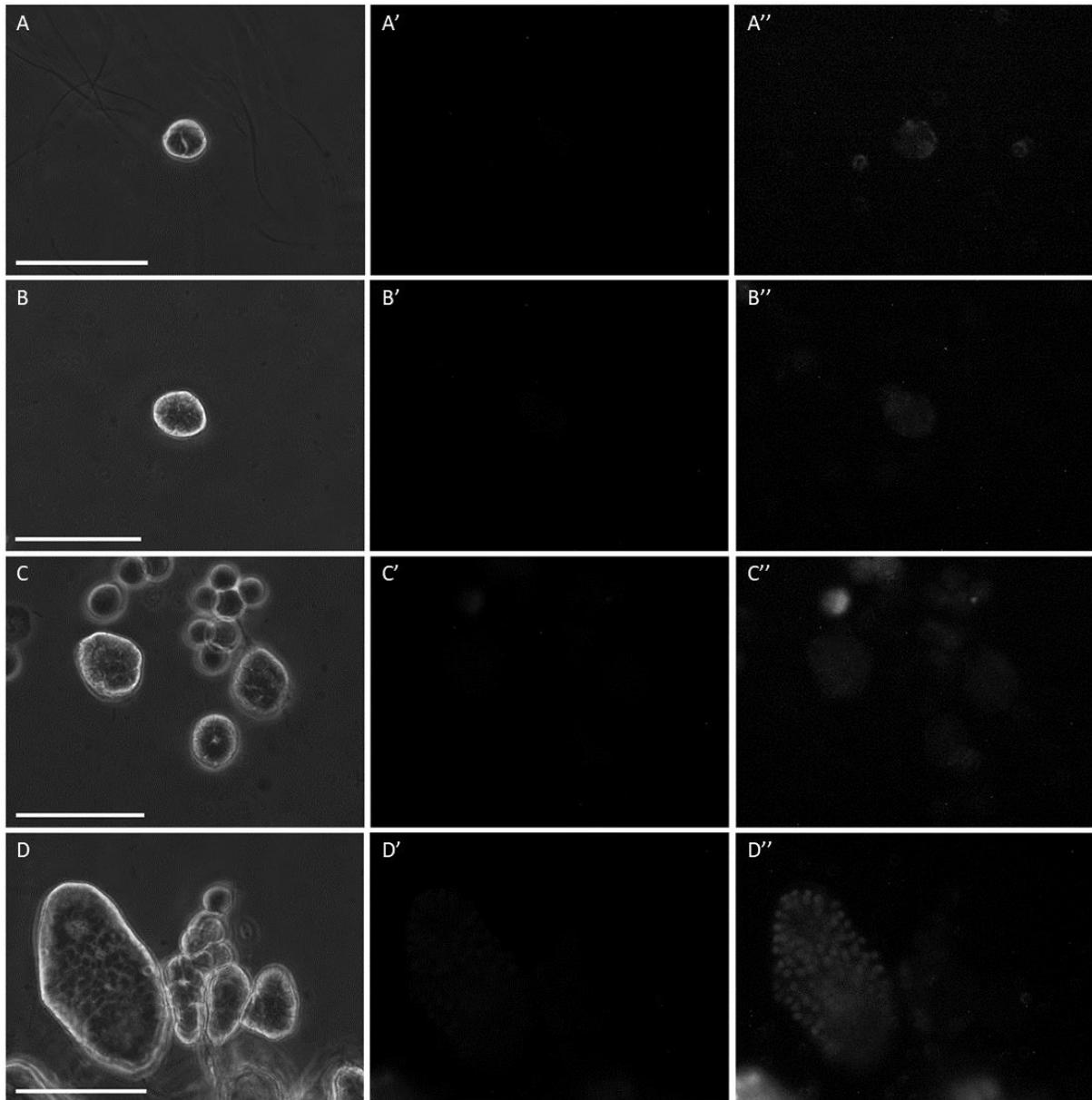


Figure 5.8: Spermatogonia and spermatocyte cysts dissected from *D. pseudoobscura* SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy. Cysts shown at stages through spermatogonia mitotic divisions. Spermatogonia do not exhibit GFP fluorescence from GFP tagged Kumgang. Scale = 50 μ m. A: Spermatogonia cyst containing at least two spermatogonia. B: Later stage spermatogonia cyst, containing 16 spermatogonia. C: (L-R) 32-, 8- and 16-cell cysts. D: (L-R) Meiosis II-stage spermatocyte cyst exhibiting low GFP expression, cysts from spermatogonia to spermatocyte stage, no GFP expression. A-D: Phase contrast images. A'-D': Single channel GFP. A''-D'': Single channel GFP, brightness enhanced.

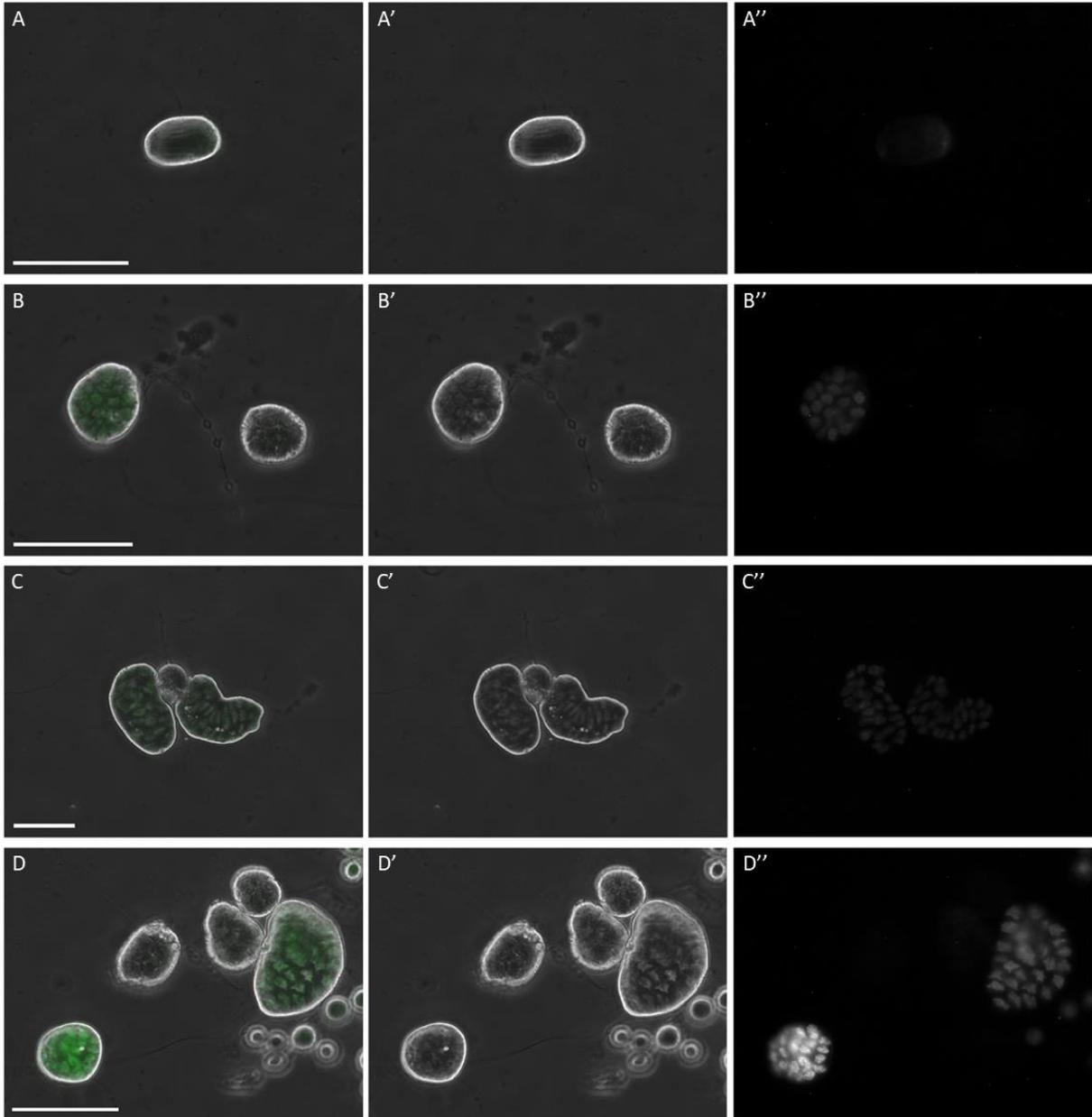


Figure 5.9: Spermatogonia and spermatocyte cysts dissected from *D. pseudoobscura* SLOB3 piggyBac-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kumgang (GA18735). Scale = 50 μ m. A: Spermatogonia undergoing mitosis, exhibiting low levels of GFP fluorescence localised to nuclei. B: Late spermatogonia cysts exhibiting high and low GFP fluorescence. C: Spermatocyte cysts exhibiting GFP fluorescence. D: Spermatocyte cysts undergoing growth phase exhibiting high and low GFP fluorescence. A-D: Combined phase contrast and GFP images. A'-D': Phase contrast images. A''-D'': Single channel GFP.

5.4.3 GFP fluorescence decreased in meiotic cysts and was undetectable in elongating spermatids

Between the primary spermatocyte and spermatid cyst stages, GFP fluorescence decreased (Figure 5.10). Figure 5.10; A shows primary spermatocyte, meiotic and post-meiotic cysts with variable levels of GFP fluorescence. Primary spermatocytes again showed high or low levels of GFP, indicating high or low levels of the tagged Kmg protein. The meiotic cyst did exhibit GFP fluorescence, however the intensity of even the brightest cysts was lower than that of the spermatocyte cysts.

Variable GFP fluorescence appears to persist throughout the growth phase of primary spermatocytes. Figure 5.10; B shows primary spermatocytes cysts (32 cell cysts), at various stages of cell growth showing both high and low GFP fluorescence. This indicates that differential Kmg expression is maintained throughout the growth phase of primary spermatocytes.

Imaging of early spermatid cysts showed the decrease in GFP intensity (Figure 5.11). GFP was visible in onion stage spermatids (Figure 5.11; A), with decreased GFP in cysts undergoing polarisation (Figure 5.11; B), the process of nuclear migration to one end of the cyst. Cysts undergoing flagellar elongation no longer had visible GFP (Figure 5.11). This indicates that Kmg persists in the nuclei throughout meiosis and into early spermiogenesis, but is no longer present after the spermatid cyst has become polarised and elongation of the flagellum has initiated.

No GFP fluorescence was observed in elongating (Figure 5.10; C and D) or individualising spermatid cysts (Figure 5.12), indicating that Kmg was very low or absent at these stages.

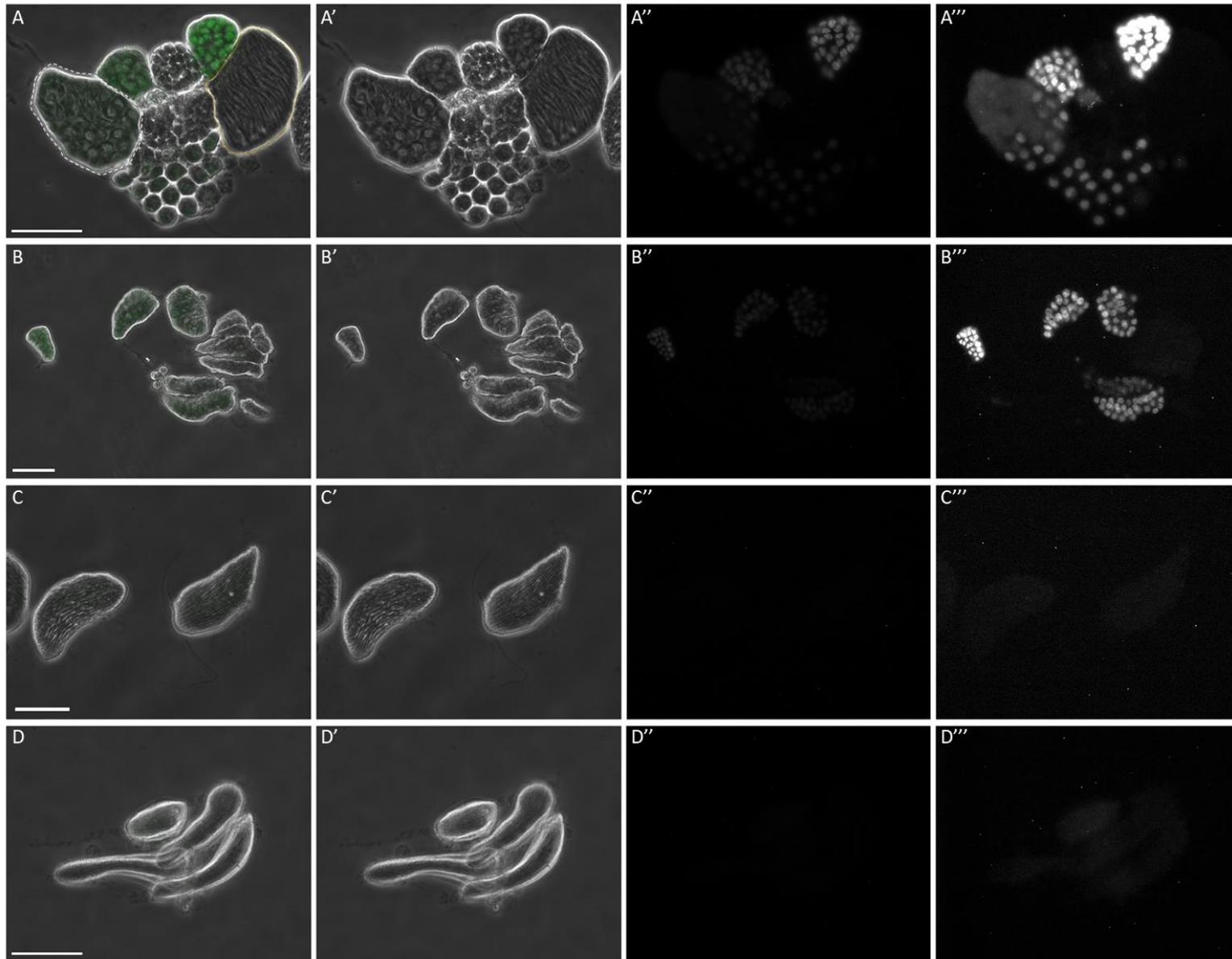


Figure 5.10: Primary spermatocyte to spermatid stages dissected from *D. pseudoobscura* SLOB3 piggyBac-3xP3-AmCyan-Kmg-GFP(C) imaged by differential interference contrast microscopy and fluorescence microscopy, showing GFP fluorescence from tagged Kmgang (GA18735). Scale = 50 μ m. A: Three primary spermatocyte cysts showing variable GFP fluorescence from tagged Kmg. White dashed line indicates cyst undergoing meiosis. GFP fluorescence is present, but less localised than pre-meiosis. Yellow dotted line indicates post-meiotic early spermatid cyst, beginning the process of elongation. Spermatid cyst does not exhibit GFP fluorescence. B: Primary spermatocyte cysts showing variable GFP fluorescence intensity undergoing cell growth. C: Early spermatid cysts undergoing elongation showing no GFP fluorescence. D: Later stage elongating or parasperm spermatid cysts showing no GFP fluorescence. A-D: Combined phase contrast and GFP images. A'-D': Phase contrast images. A''-D'': Single channel GFP. A'''-D''': Single channel GFP, brightness enhanced.

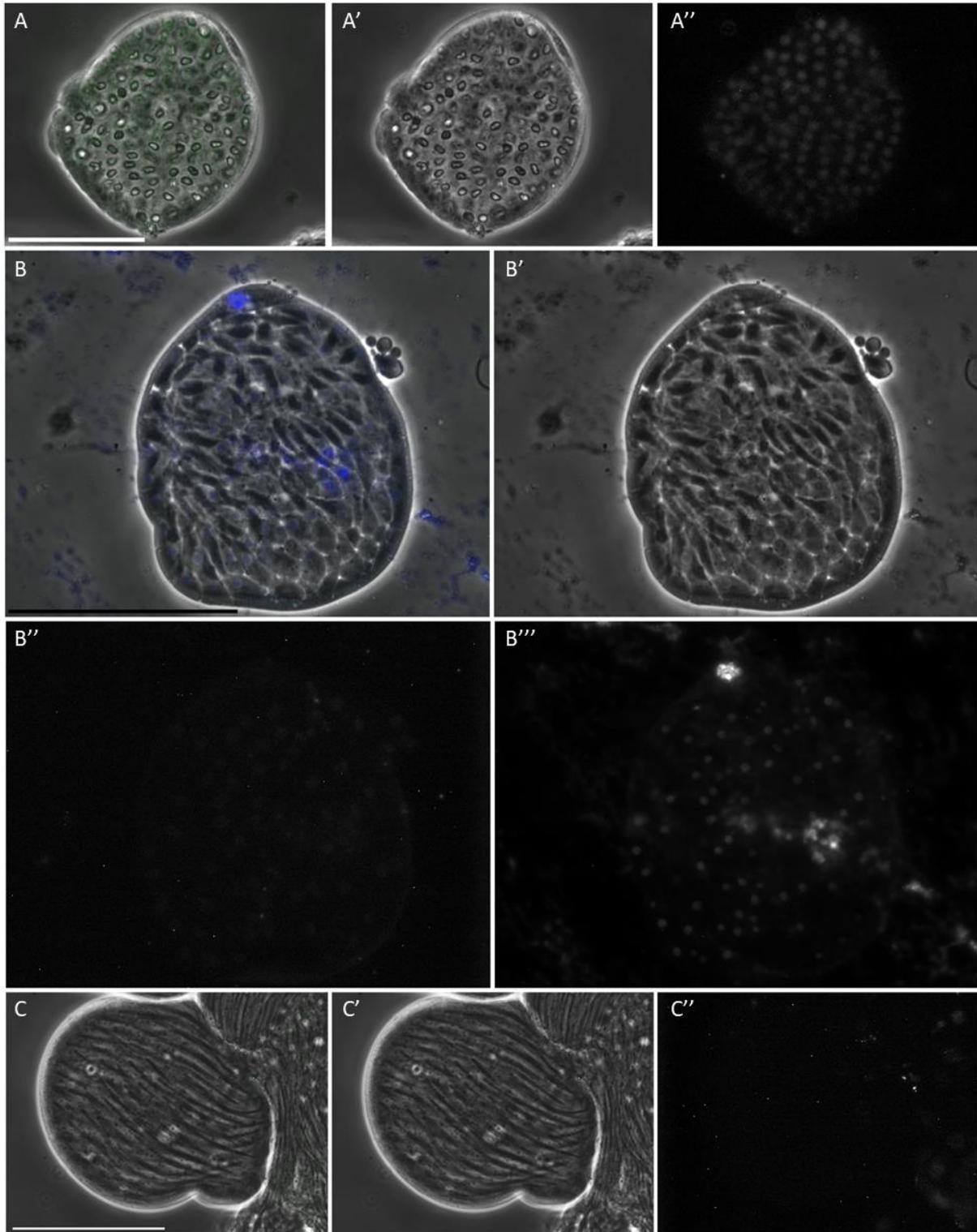


Figure 5.11: Early spermatid cysts dissected from *D. pseudoobscura* *SLOB3 piggyBac-3xP3-AmCyan-Kmg-GFP(C)* imaged by differential interference contrast microscopy and fluorescence microscopy. Scale = 50 μ m. A: Onion stage spermatid cyst. B: Spermatid cyst undergoing polarisation. C: Early elongating spermatid cyst. A-C: Combined phase contrast and GFP images. A'-C': Phase contrast images. A''-C'': Single channel GFP, brightness enhanced. B''': Single channel Hoechst 33342.

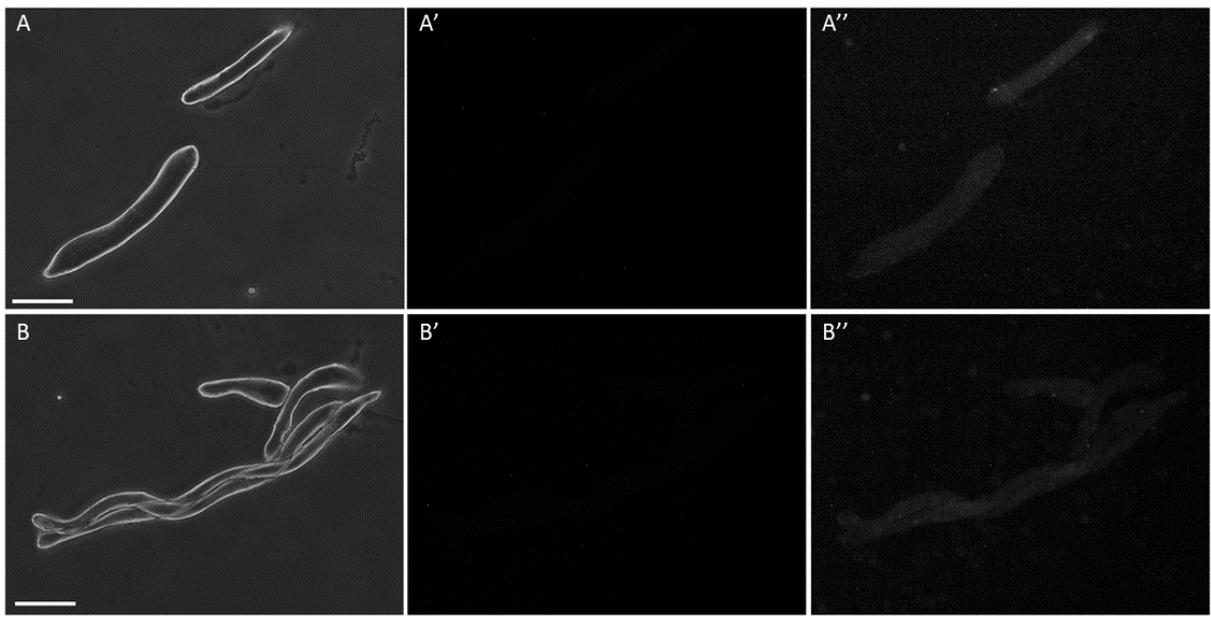


Figure 5.12: Spermatid cysts dissected from *D. pseudoobscura* *SLOB3 piggyBac-3xP3-AmCyan-Kmg-GFP(C)* imaged by differential interference contrast microscopy and fluorescence microscopy. Spermatids do not exhibit GFP fluorescence from GFP tagged Kungang. Scale = 50 μ m. A: (L-R) Elongating eusperm spermatid cyst, parasperm spermatid cyst. B: (L-R) Eusperm spermatid cysts, spermatid cysts (elongating or parasperm). A-B: Phase contrast images. A'-B': Single channel GFP. A''-B'': Single channel GFP, brightness enhanced.

5.4.4 Lattice lightsheet imaging revealed localisation of Kmg-expressing cysts within the testes

Imaging of whole *D. pseudoobscura* Kmg-GFP testes on the Lattice Lightsheet 7 (Zeiss) showed the localisation of cysts expressing GFP-tagged Kmg within the testis (Figure 5.13, Figure 5.14). All spermatocyte cysts appeared to express some level of Kmg-GFP, with some cysts showing higher expression. Cysts expressing high levels of Kmg-GFP were distributed throughout the apical portion of the testis, which contains the primary spermatocyte cysts. This finding is in accordance with imaging of individual cysts described above.

Kmg-GFP was localised to the nuclei of germline cells. Previous fluorescence imaging had shown some localisation within the nuclei (Figure 5.6; C, Figure 5.7; B). Lattice lightsheet images also show localisation within the nuclei.

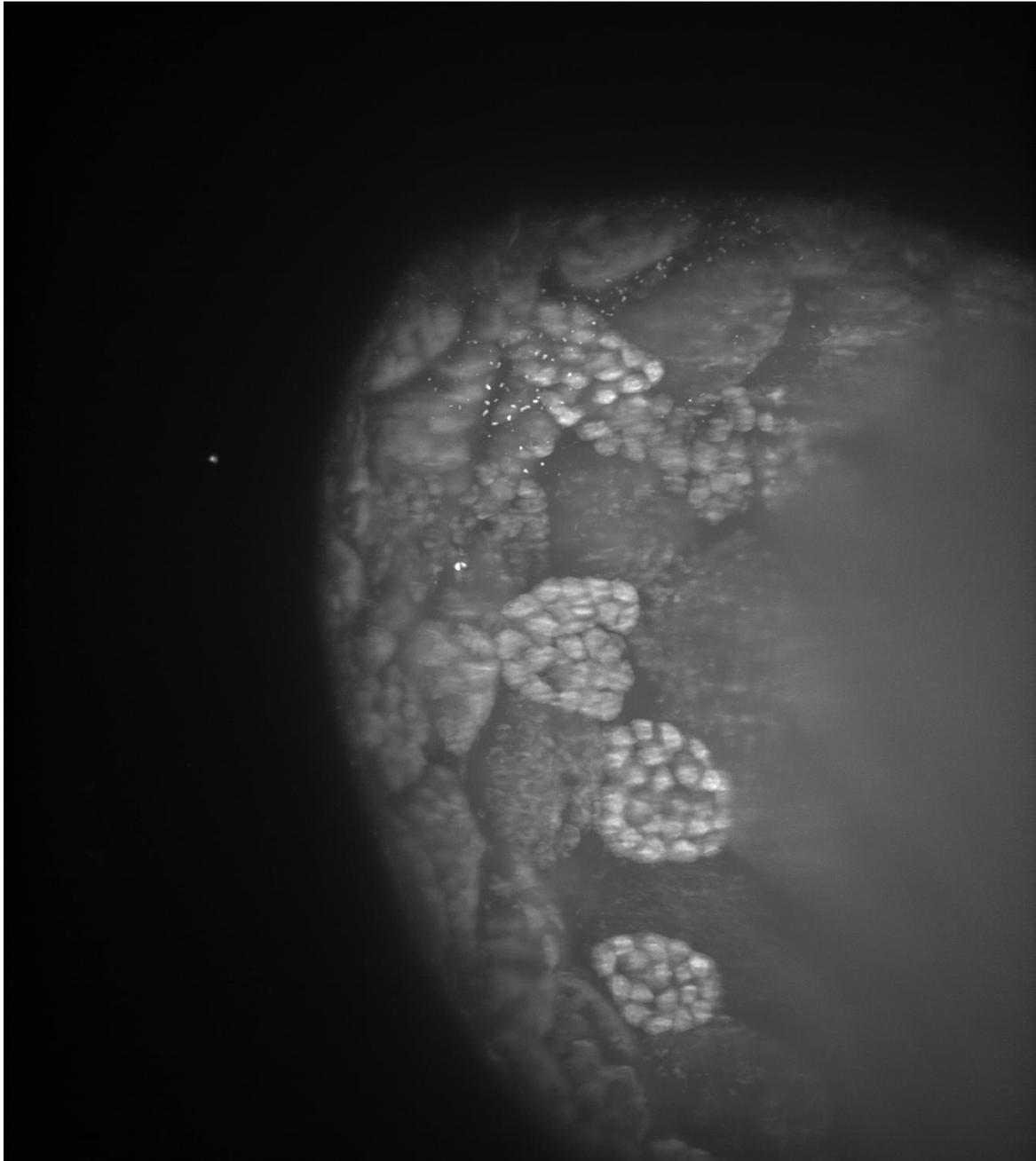


Figure 5.13: Lattice Lightsheet 7 (Zeiss) maximum intensity image showing GFP expression of Kmg-GFP fusion in whole D. pseudoobscura Kmg-GFP testis. All cysts appear to express GFP, with some cysts showing higher GFP expression. Spermatocyte cysts show highest Kmg-GFP expression, with some spermatocyte cysts showing higher signal than other spermatocyte cysts. Kmg-GFP is localised to the nuclei. There appears to be some localisation of GFP within the nuclei.

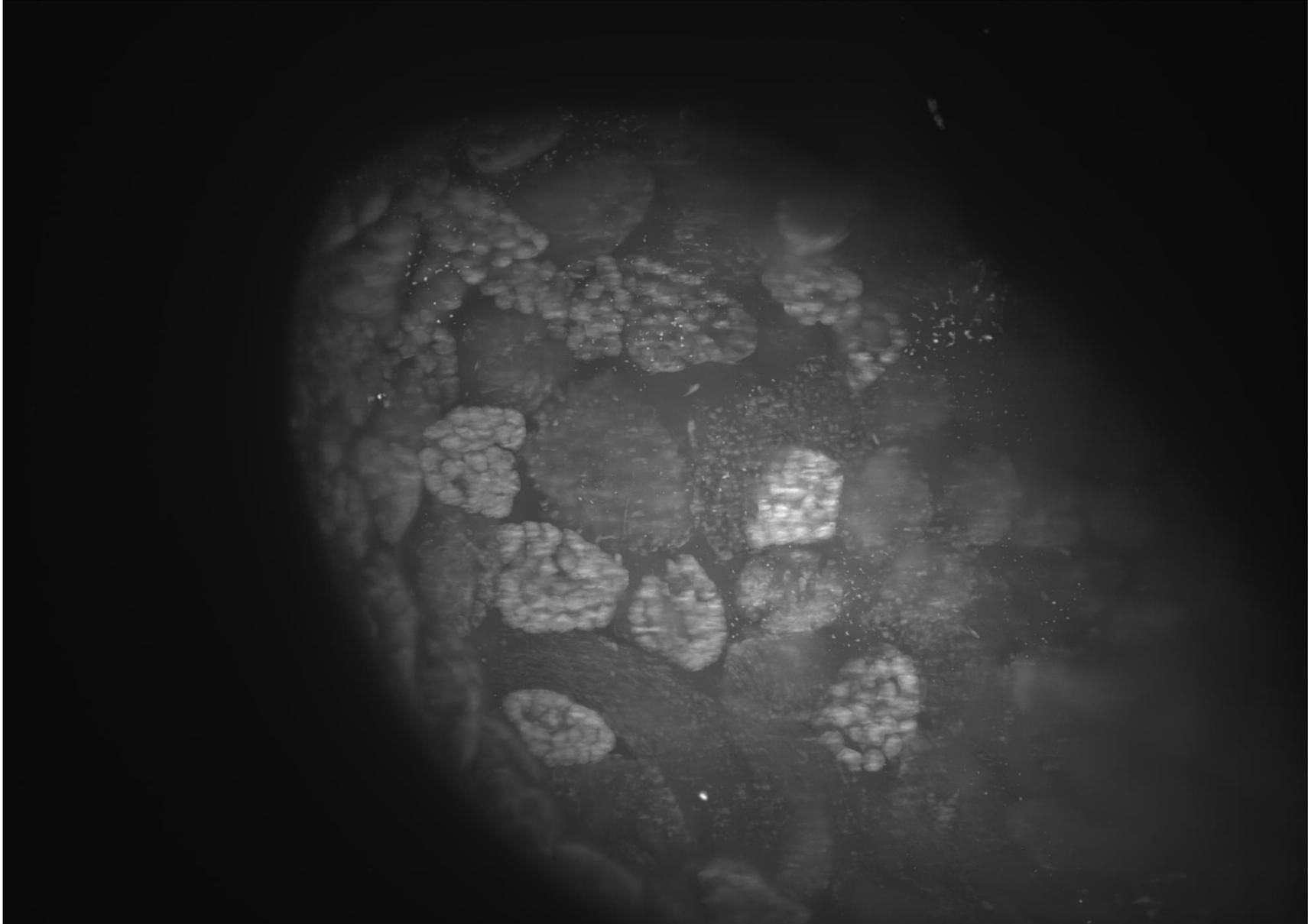


Figure 5.14: Lattice Lightsheet 7 (Zeiss) maximum intensity image showing GFP expression of Kumgang-GFP fusion in whole D. pseudoobscura Kmg-GFP testis. Spermatocyte cysts containing high levels of Kmg-GFP are distributed throughout the apical portion of the testis (spermatocyte cyst region).

6 Characterisation of *D. pseudoobscura* Testis Structure by Immunofluorescence and DNA Staining

6.1 Immunostaining *D. melanogaster* and *D. pseudoobscura* Testes for Hub and Germline Markers

6.1.1 Comparison between *D. melanogaster* and *D. pseudoobscura* germline stem cell populations with germline stem cell marker Vasa

The size and structure of the *D. pseudoobscura* testis apical proliferation centre, consisting of the hub and early germline, has not previously been well characterised. It is not known whether the number of germline stem cells (GSC) is similar to that in *D. melanogaster*. Furthermore, it is not known whether there is a single population of GSCs, all of which could contribute to the eusperm or parasperm spermatocyte population. The alternative is that there are distinct populations of GSCs, each of which contributes to a single morph. In examining the testes of *D. pseudoobscura*, a small number of GSCs would support the hypothesis that there is a single population of multipotent stem cells, which can contribute to multiple spermatocyte lineages. A larger number of GSCs could indicate multiple stem cell populations, but is not sufficient to conclude that this is the case.

Hub and germline marker antibodies were used in immunostaining in *D. pseudoobscura* testes to assess the size of the hub and the number of germline stem cells present. The *D. melanogaster* hub is approximately 30µm, surrounded by five to nine germline stem cells (Hardy *et al.* 1979).

Vasa is a marker for GSCs in *D. melanogaster*. Anti-Vasa was used to stain *D. pseudoobscura* and *D. melanogaster* testes to assess the number of GSCs surrounding the hub (Figure 6.1). Figure 6.1; A shows the apical tip of the *D. melanogaster* testis, stained for Vasa and counterstained for DNA. A small cluster of nuclei indicates the hub. Surrounding the hub are approximately eight germline stem cells, as indicated by Vasa (cyan). Immunostaining for Vasa showed a different organisation in *D. pseudoobscura*. Figure 6.1; B and C show the apical tip of the *D. pseudoobscura* testis, immunostained for Vasa and counterstained for DNA. Where *D. melanogaster* has a clear cluster of nuclei (Figure 6.1; A), there is a much larger region of compact nuclei at the apical tip of the *D. pseudoobscura* testis. Vasa does not appear to give a strong signal localised to the hub and GSCs, as it does in *D.*

melanogaster; instead Vasa staining is present in later stages of spermatogenesis, likely the spermatogonia and spermatocyte stage. Vasa does not appear to be enriched in the hub region, as is indicated by the region denser in nuclei.

Vasa does not give a clear indication of the number of GSCs in *D. pseudoobscura*. In conjunction with Hoechst 33258, the size and structure of the *D. pseudoobscura* testis hub could not be determined.

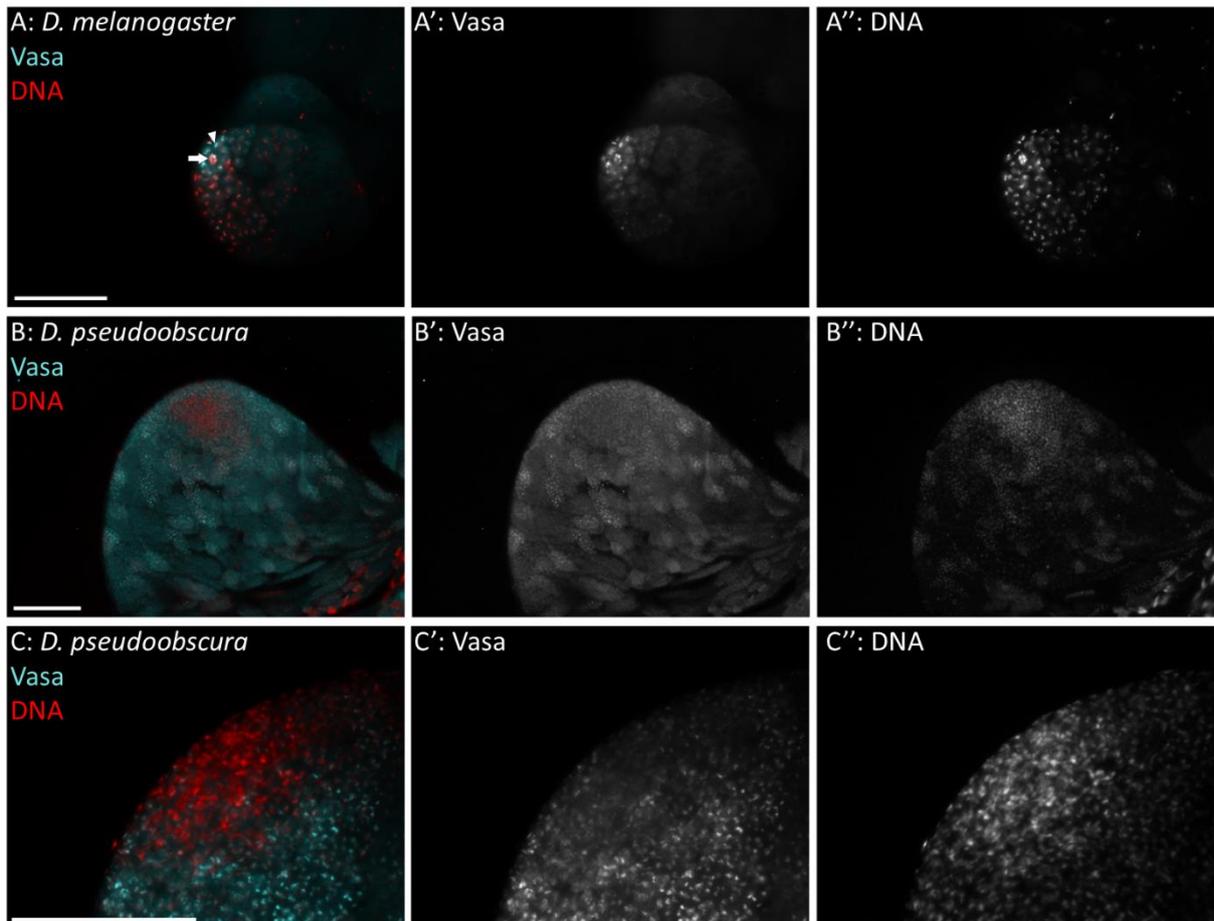


Figure 6.1: *D. melanogaster* and *D. pseudoobscura* testis hub and germline. Apical hub left. Scale = 100 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-Vasa (cyan, GSC) and counterstained with Hoechst 33258 (red, DNA). Hub is indicated by arrow, GSC indicated by arrowhead. A': Single channel anti-Vasa. A'': Single channel Hoechst 33258. B-C: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-Vasa (cyan) and counterstained with Hoechst 33258 (red, DNA). B' and C': Single channel anti-Vasa. B'' and C'': Single channel Hoechst 33258.

6.1.2 Comparison between *D. melanogaster* and *D. pseudoobscura* hub size and structure by immunostaining for Fasciclin III

Immunostaining against Vasa did not give clear indications of the number of GSCs present in *D. pseudoobscura* testes, nor give a clear indication of the structure of the hub. The presence of a large region, dense in nuclei, in the apical region of the testis indicated that the hub may be larger in *D. pseudoobscura* in comparison to *D. melanogaster*.

The hub marker Fasciclin III (FasIII) was used to further compare the size and structure of the hub between *D. melanogaster* and *D. pseudoobscura* (Figure 6.2). Figure 6.2; A shows *D. melanogaster* hub immunostained for FasIII and counterstained for DNA. The hub is shown by the FasIII marker (cyan), in a region dense with nuclei. FasIII staining shows the hub to be a small area at the apical tip of the testis, as previously indicated by Vasa/Hoechst staining. By contrast, *D. pseudoobscura* shows a large area of strong FasIII staining (Figure 6.2; B). The FasIII immunostained region is more extended in *D. pseudoobscura* than the comparable region in *D. melanogaster*, suggesting a larger hub region. The hub region appears to be surrounded by an area of weaker FasIII staining, potentially indicating GSCs in contact with the hub (Figure 6.2; B and B').

Figure 6.3 shows FasIII staining indicating the hub region over phase contrast images of *D. pseudoobscura* and *D. melanogaster* testes. Figure 6.3; A shows the strong FasIII staining indicating the hub region in *D. pseudoobscura*. Figure 6.3; A' shows the *D. pseudoobscura* testis imaged by phase contrast microscopy, in which the structure of the cells in the apical region of the testis can be seen. Figure 6.3; B shows the more compact *D. melanogaster* hub stained for FasIII. Figure 6.3; B' shows the *D. melanogaster* testis imaged by phase contrast microscopy, in which the cells in the apical region of the testis can be seen. The *D. melanogaster* hub appears as a small dark cluster of nuclei, whereas the *D. pseudoobscura* hub appears more dispersed, without a similar cluster of nuclei as observed in *D. melanogaster*.

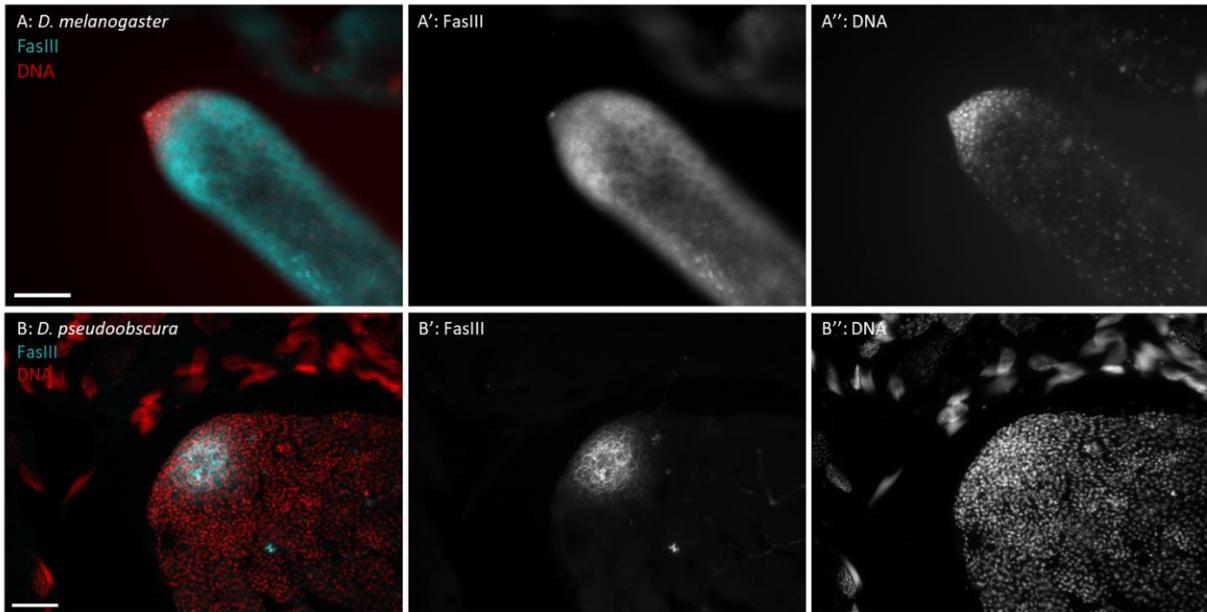


Figure 6.2: *D. melanogaster* and *D. pseudoobscura* testis hub stained for FasIII and counterstained for DNA. Apical hub left. Scale = 50 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-FasIII (cyan, hub) and counterstained with Hoechst 33258 (red, DNA). Hub is indicated by arrow. A': Single channel anti-FasIII. A'': Single channel Hoechst 33258. B: *D. pseudoobscura* SLOB3 WT testis hub immunostained for anti-FasIII (cyan, hub) and counterstained with Hoechst 33258 (red, DNA). B': Single channel anti-FasIII. B'': Single channel Hoechst 33258.

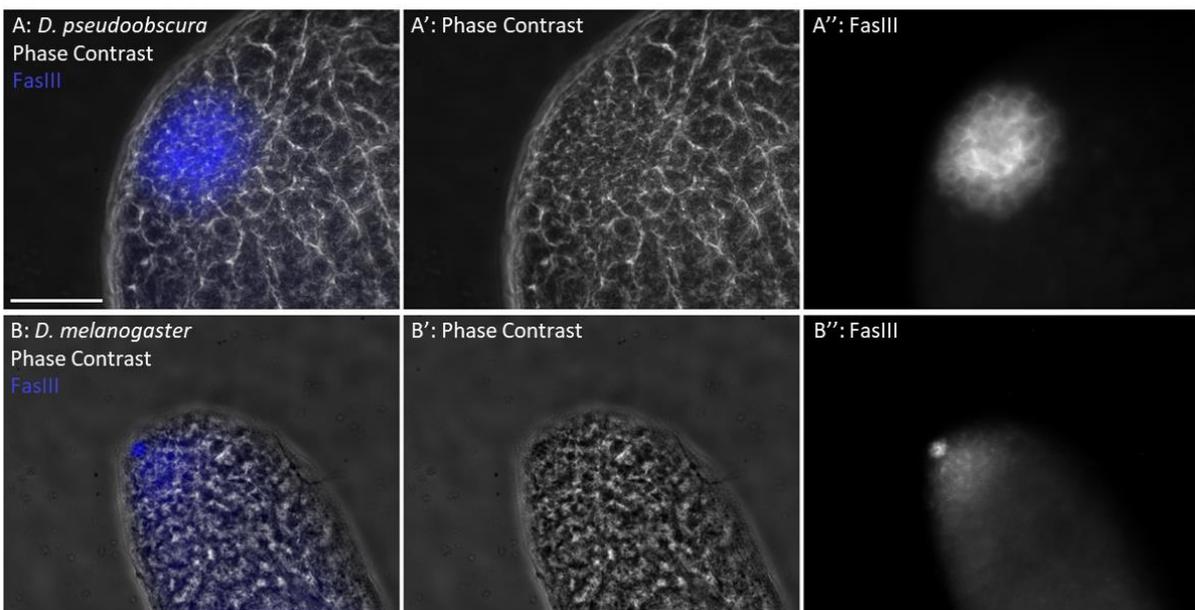


Figure 6.3: *D. pseudoobscura* and *D. melanogaster* testes imaged by differential interference contrast microscopy, immunostained for FasIII. Apical tip left. Scale = 50 μ m. A: *D. pseudoobscura* testis apical tip imaged by DIC microscopy overlaid with immunostaining with anti-FasIII (blue, hub). A': Single channel DIC. A'': Single

channel anti-FasIII. B: *D. melanogaster* testis apical tip imaged by DIC microscopy overlaid with immunostaining with anti-FasIII (blue, hub). B': Single channel DIC. B'': Single channel anti-FasIII.

6.1.3 Further investigation of *D. pseudoobscura* testis hub structure by immunostaining for hub markers

Immunostaining of *D. pseudoobscura* testes with Vasa and FasIII indicated a substantially different hub size and structure in comparison to the more familiar *D. melanogaster* model. To further investigate the structure of *D. pseudoobscura* testis hub, further immunostaining for hub markers was carried out on *D. pseudoobscura*, with comparisons to *D. melanogaster*.

6.1.3.1 Neural Cadherin showed a large hub structure in *D. pseudoobscura* testes

Neural Cadherin (N-Cad) is an adhesion molecule connecting the hub cells to each other and the GSCs to the hub (Epstein *et al.* 2017). Immunostaining for N-Cad shows the smaller hub region of the *D. melanogaster* testes, and the larger hub region in the *D. pseudoobscura* testes (Figure 6.4), consistent the results of staining for FasIII.

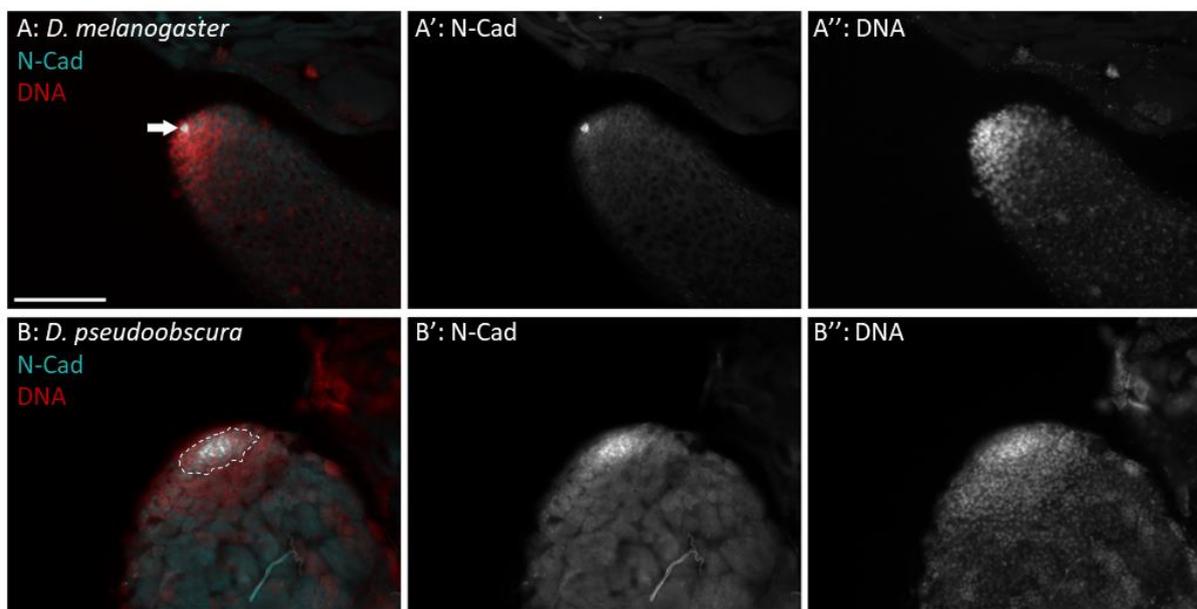


Figure 6.4: *D. melanogaster* and *D. pseudoobscura* testes immunostained for Neural Cadherin (N-Cad) and counterstained for DNA. Apical tip left. Scale = 100 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-N-Cad (cyan, hub) and counterstained with Hoechst 33258 (red, DNA). Hub is indicated by arrow. A': Single channel anti-N-Cad. A'': Single channel Hoechst 33258. B: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-N-Cad (cyan, hub) and counterstained with Hoechst 33258 (red, DNA). Hub region indicated by white dashed line. B': Single channel anti-N-Cad. B'': Single channel Hoechst 33258.

6.1.3.2 Escargot indicated the presence of substructures within the *D. pseudoobscura* testis hub region

Escargot (Esg) is a zinc-finger protein with multiple known functions. It is expressed in the tip cells of trachea, positively regulating DE-Cadherin (Tanaka-Matakatsu *et al.* 1996), is essential for maintenance of diploidy in imaginal discs (Hayashi *et al.* 1993), and is expressed in intestinal stem cells, in which it maintains the stem cell state (Korzelius *et al.* 2014; Loza-Coll *et al.* 2014). In testes, Esg is expressed in the hub, cyst stem cells, and germline stem cells (Bunt and Hime 2004; Voog *et al.* 2008; Voog *et al.* 2014).

Immunostaining for Esg shows it is present in the *D. melanogaster* hub and germline (Figure 6.5; A). The hub can be clearly seen at the apical tip of the testis, with a similar expression pattern to those observed with immunostaining for other hub markers, FasIII and N-Cad. Immunostaining also showed Esg signal in the spermatocytes, although this signal is similar to the cross-reaction observed between some secondary antibodies and Y loop structures, and may be an artefact of the Esg antibody (Redhouse *et al.* 2011).

While Esg shows a similar pattern within the hub in *D. melanogaster* to other hub markers, immunostaining for Esg in *D. pseudoobscura* appears to show a different localisation of the protein within the hub region. Within the *D. pseudoobscura* testis hub region, Esg staining appears as a 'ring', with patches of stronger expression. This suggests that the *D. pseudoobscura* hub region contains substructures in which Esg protein expression is greater.

Given the role of Esg in regulating stem cell maintenance, notably in regulation of the adhesion molecule DE-Cadherin, substructures within the *D. pseudoobscura* testis hub expressing higher levels of Esg may be involved in hub or GSC maintenance.

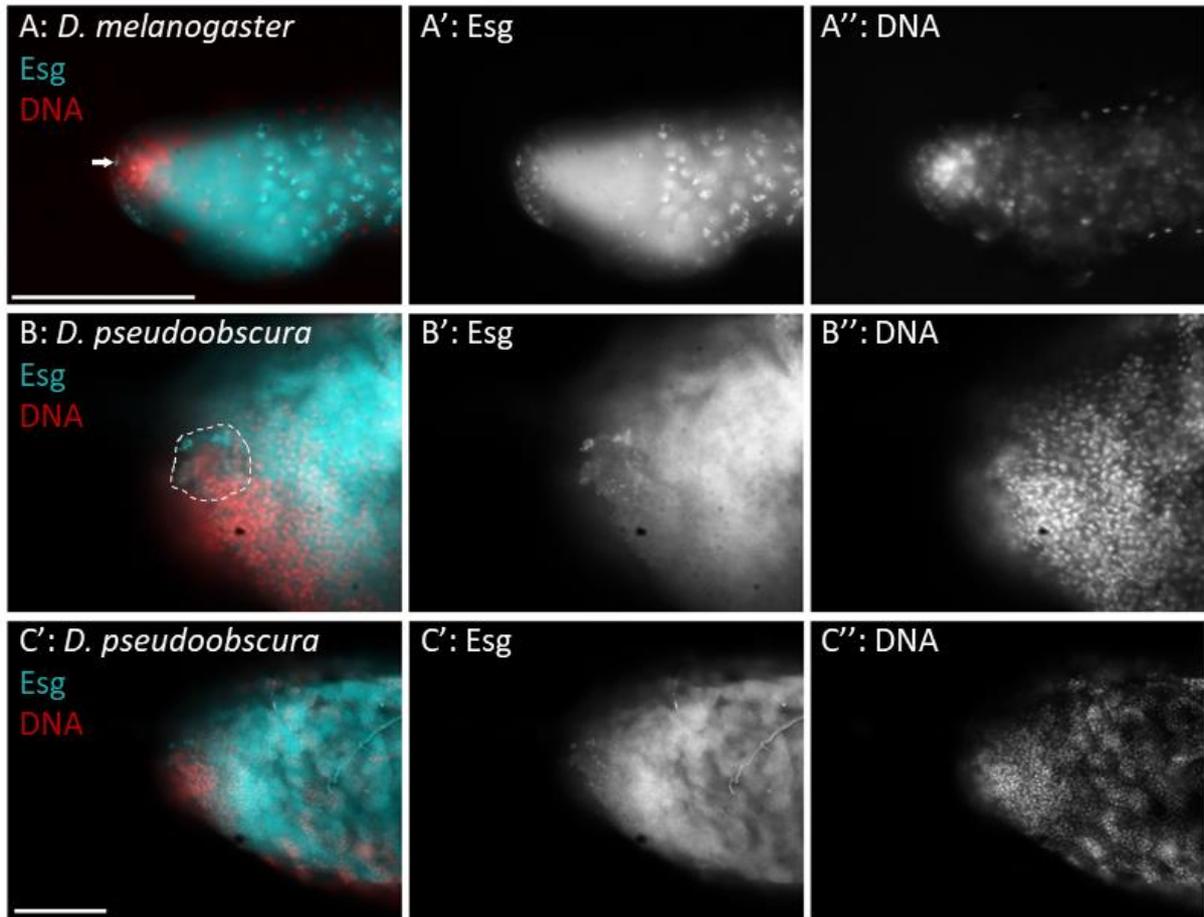


Figure 6.5: *D. melanogaster* and *D. pseudoobscura* testis hub immunostained for Escargot (Esg). Apical tip left. Scale = 100 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-Esg (cyan, hub and GSC) and counterstained with Hoechst 33258 (red, DNA). Hub is indicated by arrow. A': Single channel anti-Esg. A'': Single channel Hoechst 33258. B-C: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-Esg (cyan) and counterstained with Hoechst 33258 (red, DNA). Hub region indicated by white dashed line. B' and C': Single channel anti-Esg. B'' and C'': Single channel Hoechst 33258.

6.1.3.3 Combined immunostaining for Hu-li tai shao and DE-Cadherin

D. melanogaster and *D. pseudoobscura* testes were stained for DE-Cadherin (DE-Cad) and Hu-li tai shao (Hts), and counterstained for DNA (Figure 6.6). DE-Cad is an epithelial adhesion molecule, regulated by Escargot, which is a marker for the testis hub. Hts is a marker for the fusome, which can be used to identify GSCs.

Immunostaining for DE-Cad again labelled the hub of *D. melanogaster*. In *D. pseudoobscura* DE-Cad labelling differed from that observed with FasIII, N-Cad or Esg immunostaining (Figure 6.6). Within the hub region there appear to be smaller areas of higher expression, similar to the regions of higher expression in Esg immunostaining, but there does not appear to be the 'ring'-like structure observed

with Esg. These ‘spots’ of higher DE-Cad did not appear consistent in number between testes, as seen in Figure 6.6; B and C.

Hts shows similar staining between *D. melanogaster* and *D. pseudoobscura*, with greater branching present as distance from the hub increases (Figure 6.6, Figure 6.7). Around the hub, Hts shows the spectrosome of GSCs, where there is no branching. These can be seen as individual ‘dots’ in *D. melanogaster* (Figure 6.6; A and A’, Figure 6.7; A and A’). The *D. pseudoobscura* testis hub region also contains unbranched fusome, which appears to be more dispersed throughout the hub region, rather than surrounding the hub as is observed in *D. melanogaster* testes (Figure 6.7; B and B’).

Hts persists into the later stages of germline differentiation; Figure 6.8 shows Hts labelling the fusome in primary spermatocyte cysts, demonstrating the branching connecting the later germline cells.

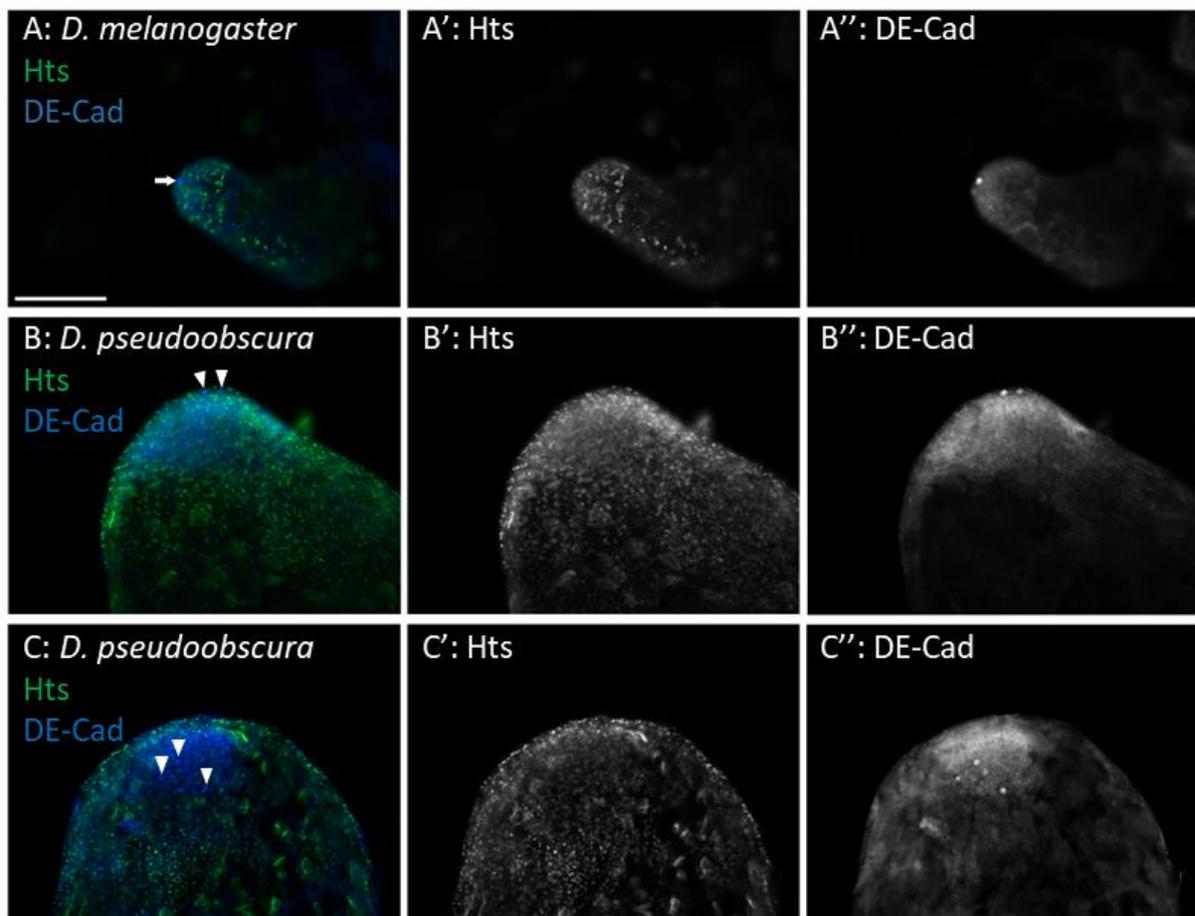


Figure 6.6: *D. melanogaster* and *D. pseudoobscura* testis hub immunostained for *Hu-li tai shao* (Hts) (green) and DE-Cadherin (DE-Cad) (blue). Scale = 100 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-Hts (green, fusome) and anti-DE-Cad (blue, hub). Hub is indicated by arrow. A': Single channel anti-Hts. A'': Single channel anti-DE-Cad. B-C: *D. pseudoobscura* SLOB3 WT testis hub immunostained for Hts (green, fusome) and DE-Cad (blue). Arrowheads indicate possible mini hub structures. B' and C': Single channel anti-Hts. B'' and C'': Single channel anti-DE-Cad.

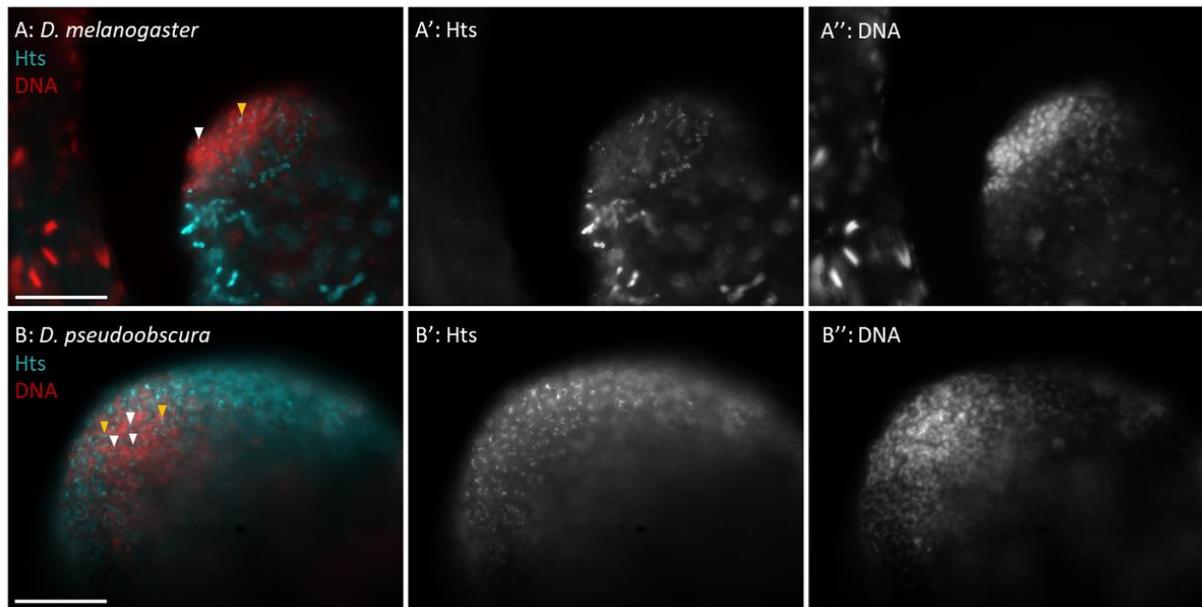


Figure 6.7: *D. melanogaster* and *D. pseudoobscura* testis hub immunostained for *Hu-li tai shao* (*Hts*) (cyan) and counterstained for DNA (red). Unbranched fusome (spectrosome) indicated with white arrowheads. Branched fusome indicated with yellow arrowheads. Scale = 50µm. A: *D. melanogaster* w1118 testis hub immunostained with anti-*Hts* (cyan, fusome) and DNA (red). A': Single channel anti-*Hts*. A'': Single channel Hoechst 33258. B-C: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-*Hts* (cyan, fusome) and DNA (red). B' and C': Single channel anti-*Hts*. B'' and C'': Single channel Hoechst 33258.

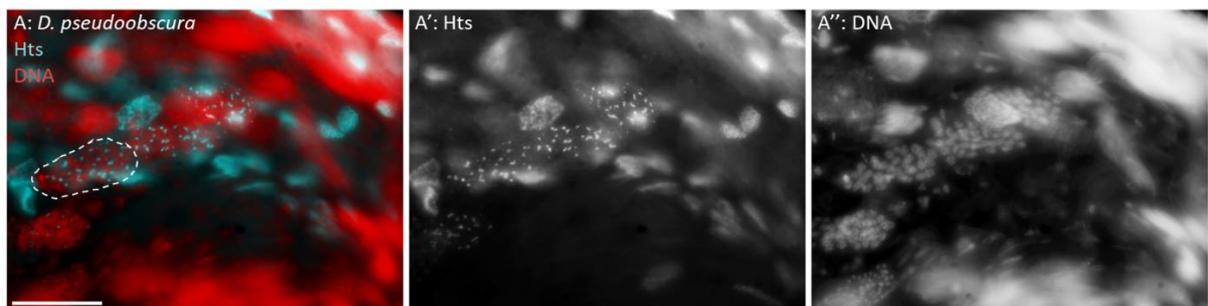


Figure 6.8: *D. pseudoobscura* primary spermatocyte cysts immunostained for *Hu-li tai shao* (*Hts*) (cyan) and counterstained for DNA (red). White dashed line indicates spermatocyte cyst. Scale = 50µm. A': Single channel anti-*Hts*. A'': Single channel Hoechst 33258.

6.1.3.4 Armadillo and Discs large are not markers of the *D. pseudoobscura* testis hub

Armadillo (Arm) was enriched in the *D. melanogaster* hub. Arm was present in the *D. pseudoobscura* testis, but was not enriched in the hub (Figure 6.9). Arm is also a marker for cyst cells, which can be seen in both *D. melanogaster* and *D. pseudoobscura* (Figure 6.9), confirming that the anti-Arm did cross-react with the *D. pseudoobscura* Arm protein.

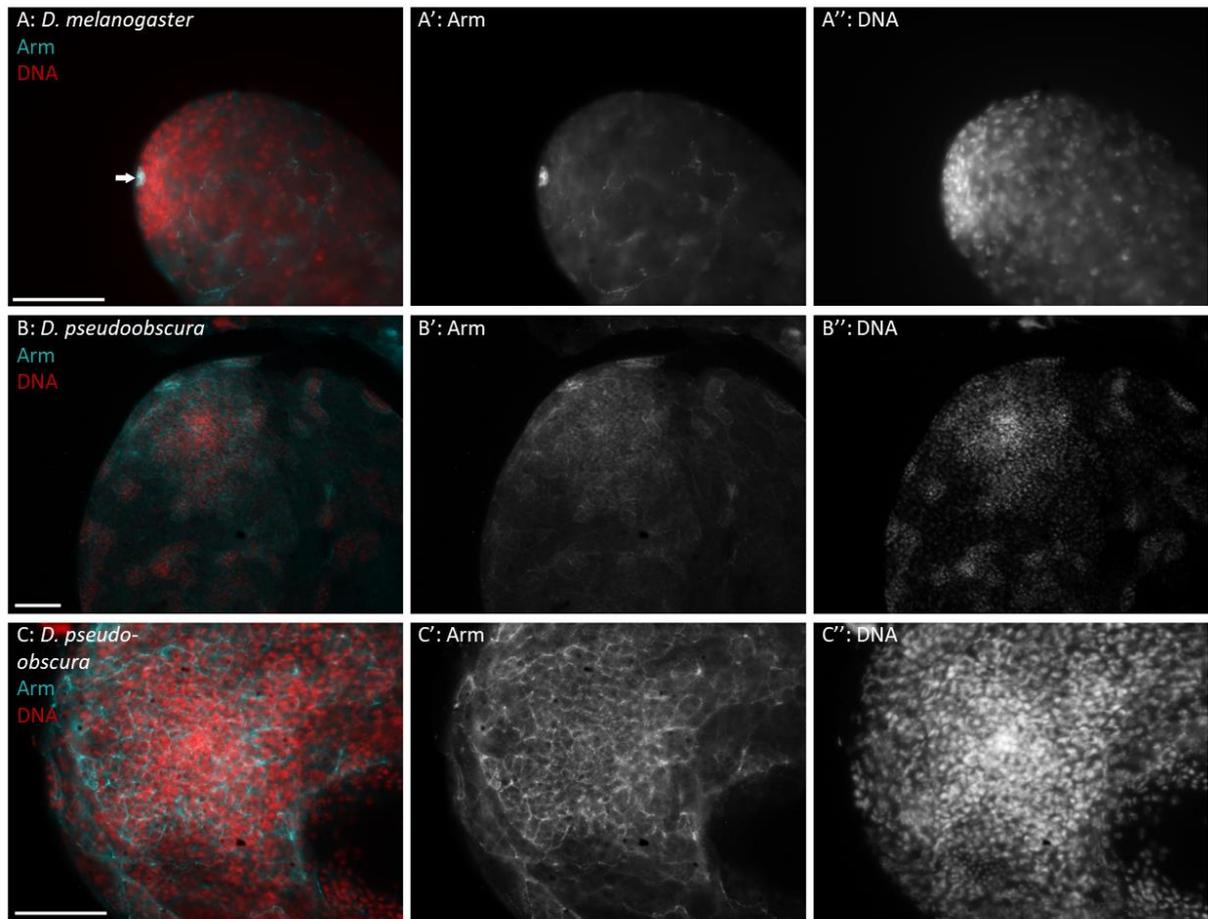


Figure 6.9: *D. melanogaster* and *D. pseudoobscura* testis hub immunostained for Armadillo (Arm) (cyan) and counterstained for DNA (red). Scale = 50 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-Arm (cyan, hub) and DNA (red). Hub is indicated with white arrow. A': Single channel anti-Arm. A'': Single channel Hoechst 33258. B-C: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-Arm (cyan) and DNA (red). B' and C': Single channel anti-Arm. B'' and C'': Single channel Hoechst 33258.

Discs large (Dlg) also marks the hub and cyst cells in *D. melanogaster*, but does not exhibit the same pattern of staining in *D. pseudoobscura* (Figure 6.10). In *D. pseudoobscura* testes, Dlg appears to be localised to cell membranes as well as showing weak localisation to the cyst cells. Interestingly, there appear to be 'bands' of increased Dlg as the cysts are pushed away from the hub (Figure 6.10; B and C). Unlike *D. melanogaster*, there does not appear to be Dlg in the hub region.

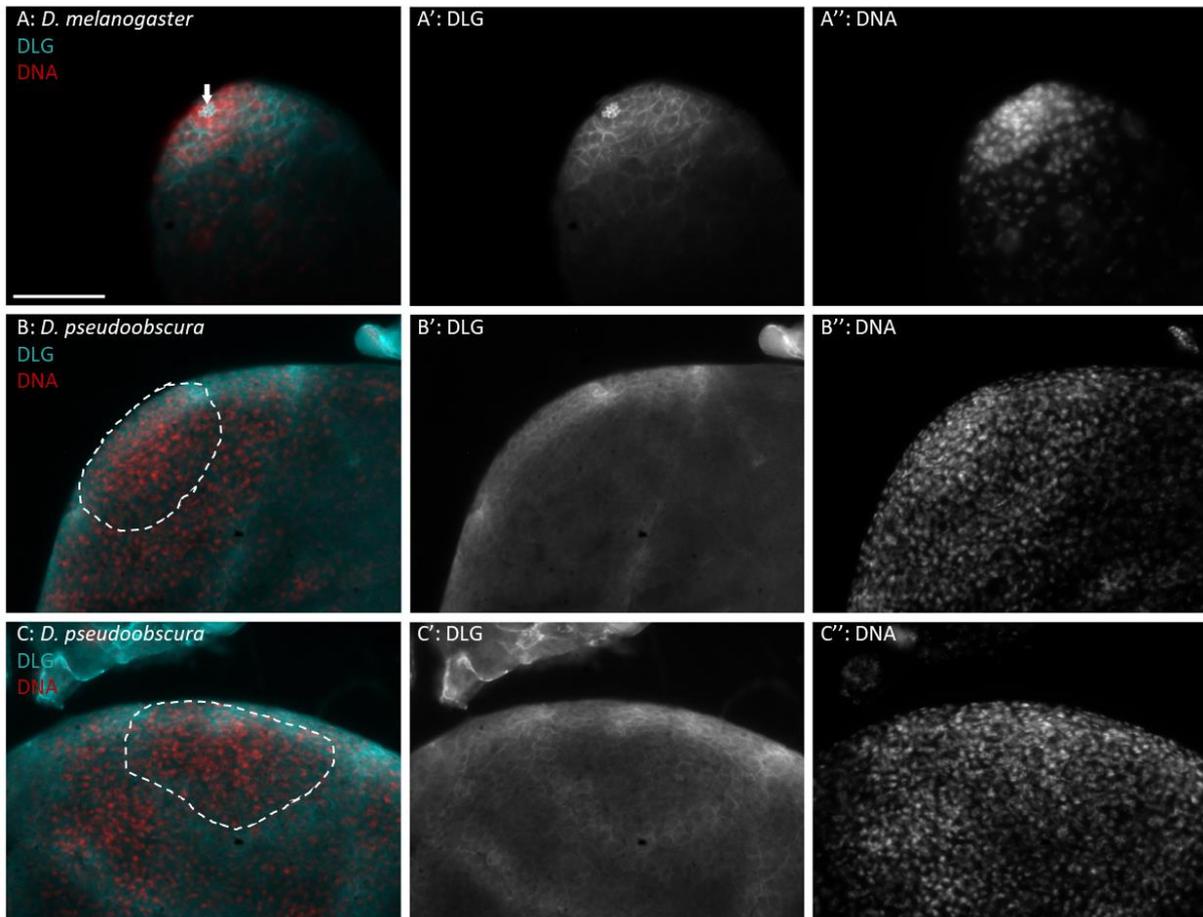


Figure 6.10: *D. melanogaster* and *D. pseudoobscura* testis hub immunostained for Discs large (Dlg) (cyan) and counterstained for DNA (red). Scale = 50 μ m. A: *D. melanogaster* w1118 testis hub immunostained with anti-Dlg (cyan, hub and cyst cells) and DNA (red). Hub is indicated with white arrow. A': Single channel anti-Dlg. A'': Single channel Hoechst 33258. B-C: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-Dlg (cyan) and DNA (red). Hub region indicated with white dashed line. B' and C': Single channel anti-Dlg. B'' and C'': Single channel Hoechst 33258.

6.1.4 Double staining for adhesins and FasIII did not reveal hub sub-structures

FasIII immunostaining demonstrated the presence of a large hub region in the *D. pseudoobscura* testis and DE-Cad showed that there may be sub-structures within the *D. pseudoobscura* testis hub region. However, combining FasIII and DE-Cad immunostaining did not show localisation of DE-Cad within the hub region (Figure 6.11). FasIII showed the hub region as previously observed. However, DE-Cad did not show localisation to the smaller areas. Images of DE-Cad immunostaining appear to show bleed-through from Hoechst 33258 DNA staining (Figure 6.11; A''). It was not possible further clarify the presence of the DE-Cad substructures within the hub region. FasIII and DE-Cad double staining showed only background in *D. melanogaster*.

Combined immunostaining for FasIII and N-Cad was successful in *D. melanogaster*, with both clearly showing the hub (Figure 6.12; A). In *D. pseudoobscura*, combined FasIII and N-Cad immunostaining shows only FasIII localisation, but does not show the N-Cad localisation as previously observed (Figure 6.12; B).

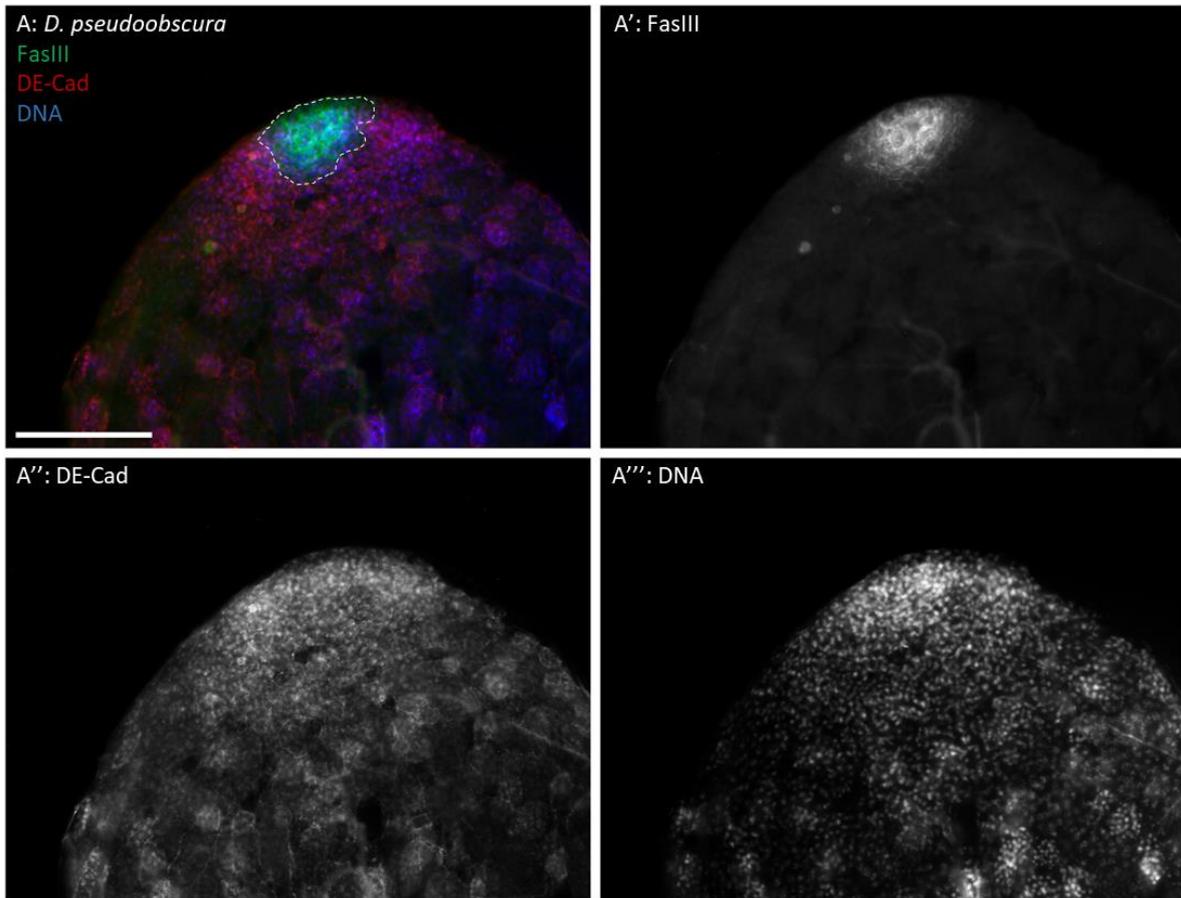


Figure 6.11: A: *D. pseudoobscura* testis hub immunostained for Fasciclin III (FasIII) (green), DE-Cadherin (DE-Cad) (red) and counterstained for DNA (blue). Scale = 100 μ m. Hub region indicated by white dashed line. A': Single channel anti-FasIII. A'': Single channel anti-DE-Cad. A''': Single channel Hoechst 33258.

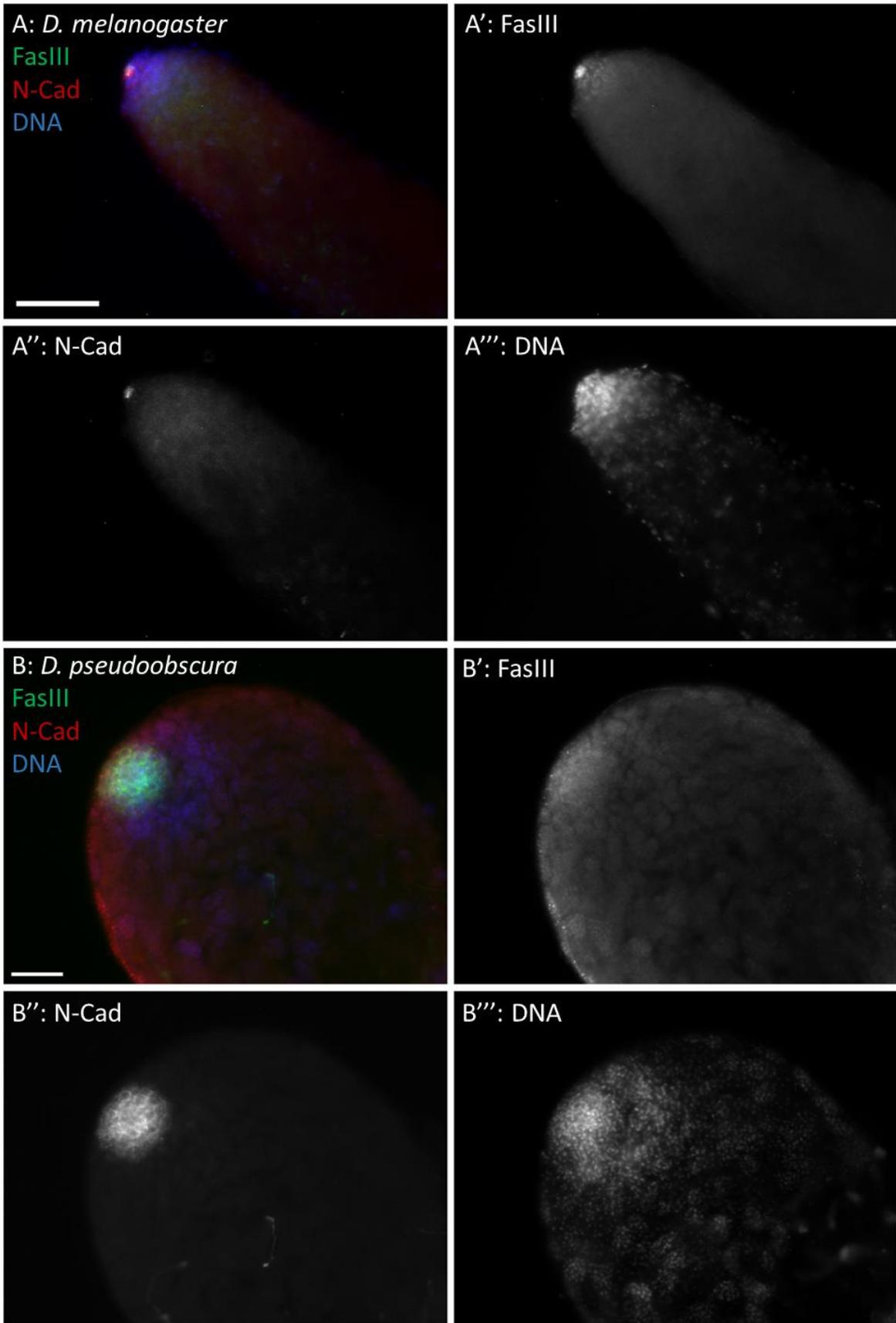


Figure 6.12: *D. melanogaster* and *D. pseudoobscura* testis hub immunostained for N-Cadherin (N-Cad) (red), Fasciclin III (green) and counterstained for DNA (blue). Scale = 50µm. A: *D. melanogaster* w1118 testis hub immunostained with anti-N-Cad (red, hub and GSC), anti-FasIII (green, hub) and DNA (red). Hub is indicated with white arrow. A': Single channel anti-N-Cad. A'': Single channel anti-FasIII. A''': Single channel Hoechst 33258. B: *D. pseudoobscura* SLOB3 WT testis hub immunostained with anti-N-Cad (red), anti-FasIII (green, hub) and DNA (red). B': Single channel anti-N-Cad. B'': Single channel anti-FasIII. B''': Single channel Hoechst 33258.

6.2 DNA Staining of *D. pseudoobscura* Seminal Vesicle and Ejaculatory Duct

Individualisation of spermatids in *D. melanogaster* takes place prior to release from the cyst, such that individual mature sperm are visible in the seminal vesicle with DNA staining (Fuller 1993; Fuller 1998; Civetta 1999). Observations of seminal vesicles counterstained for DNA in immunofluorescence studies (previously described) suggested that the *D. pseudoobscura* spermatids do not individualise in the same process as *D. melanogaster*, remaining grouped after release from the cyst and transfer into the seminal vesicle. To observe the structure of sperm within the seminal vesicle, nuclei were stained with Hoechst 33258 DNA dye.

Hoechst staining for DNA showed the presence of mature sperm within the seminal vesicles. *D. melanogaster* seminal vesicles contained mature, individualised sperm (Figure 6.13; A). *D. pseudoobscura* seminal vesicles contained mature sperm, but the nuclei did not appear evenly dispersed as observed in *D. melanogaster* (Figure 6.13; B). The nuclei appear in bundles within the seminal vesicle. It is not clear whether the spermatids fully individualise prior to release from the spermatid cysts, and therefore the presence of bundled sperm nuclei in the seminal vesicle may indicate the mature sperm have in fact not fully individualised. Alternatively, there may be individualisation prior to release from the cyst, but the spermatids remain partly in contact in the seminal vesicle. This could be due to electrochemical forces between the individual sperm cells, or because mature sperm are not released from the cyst cells at the base of the testis but are maintained within the cyst in the seminal vesicle.

Dissection of the seminal vesicles showed that mature sperm removed from the seminal vesicle into testis buffer were not encapsulated in cysts, and were separated into individual sperm cells. This suggests that bundling of sperm nuclei observed by Hoechst 33258 staining of the intact seminal vesicle was not due to cyst encapsulation, and forces maintaining bundling of sperm in the seminal vesicle are weak.

DNA staining of the ejaculatory duct revealed that the sperm were fully individualised (Figure 6.13; C and D). This indicates that the process of sperm transfer from the seminal vesicle to the ejaculatory duct results in the separation of mature sperm, potentially as a result of other components of the ejaculate, produced by the accessory glands.

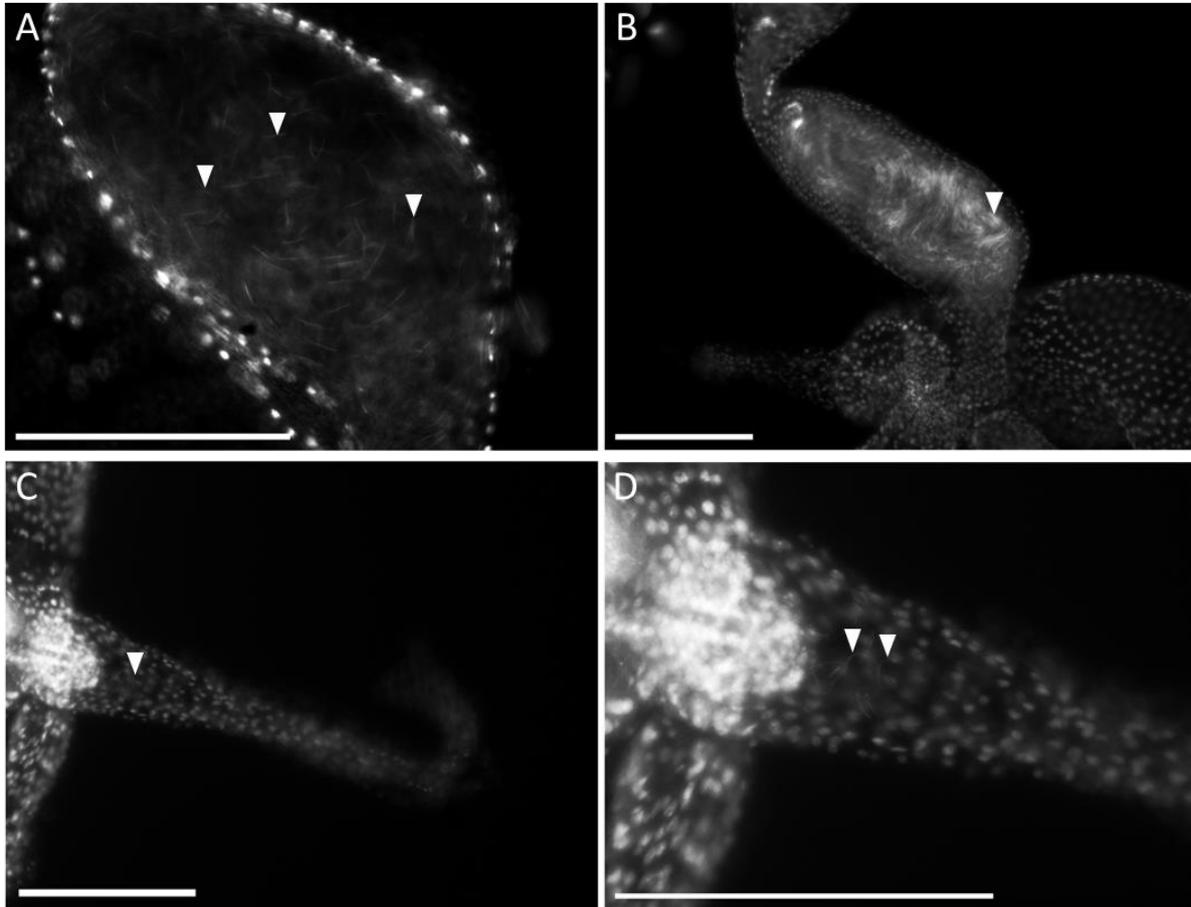


Figure 6.13: A-B: *D. melanogaster* and *D. pseudoobscura* seminal vesicle stained for DNA (Hoechst 33258). White arrows indicate individual mature sperm nuclei in *D. melanogaster* seminal vesicle, and bundle of mature sperm nuclei in *D. pseudoobscura* seminal vesicle. C-D: *D. pseudoobscura* ejaculatory duct. White arrows indicate individual mature sperm nuclei. Scale = 100 μ m.

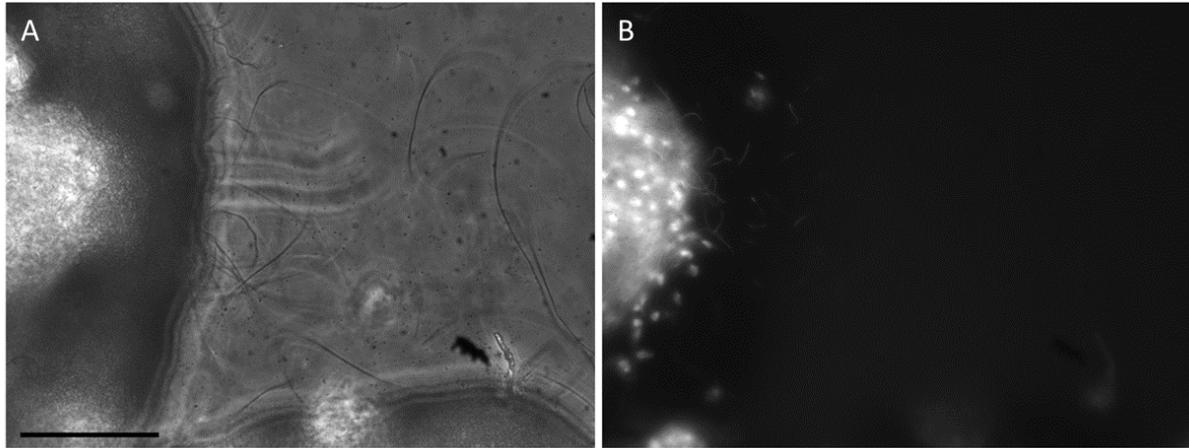


Figure 6.14: Sperm released from *D. pseudoobscura* seminal vesicle dissected in testis buffer with $2\mu\text{g}/\text{mL}$ Hoechst 33342. Sperm released from cyst were individualised and did not show evidence of bundling. Scale = $50\mu\text{m}$. A: Phase contrast image. Left shows seminal vesicle tissue. Right shows individual mature sperm cells. B: Hoechst 33342 staining DNA.

6.3 Summary

The structure of the hub in *D. pseudoobscura* testes differs substantially from that of *D. melanogaster*. The *D. melanogaster* hub is a small domed structure of epithelial cells, to which germline stem cells adhere. The *D. pseudoobscura* hub is a much larger structure, and appears as a large thin disc of epithelial cells, to which GSCs are adhered.

The *D. pseudoobscura* hub does not have a distinct boundary, as *D. melanogaster* hub does. Hub markers were enriched in the centre of the hub, and more diffuse at the edges of the hub. The *D. pseudoobscura* hub appeared to contain sub-structures, as indicated by Esg and DE-Cad. There was no clear cytological evidence of multiple populations of stem cells within the hub, pertaining to the multiple sperm morphs.

Sperm in the seminal vesicle appear to be bundled in *D. pseudoobscura*, which was not observed in *D. melanogaster*. Removal from the seminal vesicle showed individual spermatozoa, indicating that forces maintaining bundles in the seminal vesicles are weak.

7 Discussion

7.1 Project Overview: Prior State of Knowledge

7.1.1 Functions of heteromorphic sperm in *D. pseudoobscura*

The species of the *obscura* group of *Drosophila* are sperm heteromorphic, producing at least two distinct size classes of mature spermatozoa, a form of sperm heteromorphism termed ‘polymegaly’ (Beatty and Sidhu 1969). *D. pseudoobscura* have three distinct sperm size classes; a long eusperm, short parasperm 1, and medium parasperm 2 (Beatty and Sidhu 1969; Alpern *et al.* 2019). Eusperm and parasperm have differing functions. Eusperm are fertilisation competent, whereas parasperm are not (Snook and Karr 1998). Parasperm protect the eusperm from female-mediated spermicides present in the reproductive tract (Holman and Snook 2008; Alpern *et al.* 2019). The presence of parasperm reduces eusperm cell death, thus increasing the likelihood of eusperm entering into the female sperm storage organs (spermathecae) and increasing the likelihood of fertilisation.

7.1.2 Sperm development in *Drosophila*

Sperm development in *Drosophila* follows a general pattern, which has been well characterised in *D. melanogaster* (Lindsley and Tokuyasu 1980; Fuller 1993). Development of mature spermatozoa from germline stem cells takes place in testes, in *D. melanogaster* a blind-ended coiled tube, in *D. pseudoobscura* a blind-ended ellipsoid or balloon shaped structure (Baker 1935; Stern and Hadorn 1939; Hardy *et al.* 1979).

The apical blind end contains the testis niche, consisting of the testis hub and stem cells (Hardy *et al.* 1979). The hub controls germline and somatic stem cell maintenance and proliferation in the testes. Germline stem cells are attached to the hub and receive signalling molecules secreted by the hub (Gonczy *et al.* 1997; Tulina and Matunis 2001). GSC division produces a GSC and a gonialblast, the latter of which is displaced away from the hub. Displacement away from the hub results in decreased exposure to hub signalling molecules, promoting differentiation of the gonialblast (Tulina and Matunis 2001; de Cuevas and Matunis 2011; Matunis *et al.* 2012). Differentiating spermatogonia are encapsulated by two differentiating somatic cyst cells (Fuller 1993). The cyst cells encapsulate the developing germline cells for the remainder of spermatogenesis, until the mature spermatozoa are released at the base of the testis (Tokuyasu *et al.* 1972b; Lindsley and Tokuyasu 1980; Fuller 1993).

Within the cyst, the germline gonialblast begins several rounds of mitosis, known as the transit amplifying stage, producing a bundle of interconnected spermatogonia in a spermatogonial cyst (Lindsley and Tokuyasu 1980; Fuller 1993). Completion of the final mitotic division is followed by the

pre-meiotic S-phase, producing the spermatocyte cyst (Cross and Shellenbarger 1979; Fuller 1993). The spermatocyte stage is characterised by extensive growth and transcription (Olivieri and Olivieri 1965; Tates 1971; Gould-Somero and Holland 1974; Schafer *et al.* 1995). The spermatocytes are the most transcriptionally active cells in adult *Drosophila* (Li *et al.* 2022), and the majority of transcripts required during the later stages of sperm development are expressed during the spermatocyte stage (Gould-Somero and Holland 1974; Schafer *et al.* 1995).

Meiotic division of the spermatocytes results in a cyst of haploid spermatids, 64 cells in *D. melanogaster* and 128 cells in *D. pseudoobscura* (Cross and Shellenbarger 1979; Fuller 1993; Scharer *et al.* 2008). The spermatids undergo a dramatic change in shape, elongating to the full length of the mature spermatozoa, assembly of the flagellum, and condensation and elongation of the nucleus to form the needle shaped sperm 'head' (Tates 1971; Tokuyasu *et al.* 1972b, a; Tokuyasu 1974b, a, 1975b, a; Lindsley and Tokuyasu 1980; Fuller 1993; Fabian and Brill 2012).

The structure of the testis and sperm development processes in the *obscura* species group are less well characterised than those of *D. melanogaster*. *D. pseudoobscura* spermatogenesis follows the same general pattern of germline differentiation and division (Scharer *et al.* 2008; Njogu *et al.* 2010; Moore *et al.* 2013), however little is known about the specific development of the eusperm and parasperm morphs. Spermatid cysts contain a single morph (Beatty and Sidhu 1969; Hauschteck-Jungen and Maurer 1976; Bircher *et al.* 1995). This could suggest that the processes of morph differentiation begin early in development, during GSC division or in the gonialblast. Alternatively, control of morph-specific differentiation may occur later in development, but relies on signalling between the interconnected germline cells within the cyst, or between the somatic cyst cells and germline cells.

Furthermore, the cytological and morphological processes involved in the development of heteromorphic sperm are not well understood. Eusperm are probably energetically costlier and take longer to produce than parasperm (Bircher *et al.* 1995; Snook 1998b; Alpern *et al.* 2019). Parasperm 2 also have a coiled tail structure, which does not appear to present in eusperm or parasperm 1 (Alpern *et al.* 2019).

7.1.3 Transcription in *D. melanogaster* spermatogenesis

Transcriptional control in *Drosophila* spermatogenesis is essential for the regulation of germ cell differentiation. This is evident in the spermatocyte stage, during which transcription of genes whose products are required in the spermatocyte, meiotic and post-meiotic stages take place (Olivieri and Olivieri 1965; Fuller 1993; Schafer *et al.* 1995; Li *et al.* 2022). The transcription factors tMAC and testis-specific subunits of TFIID are required for the transcription of over 1000 and 350 genes respectively in

the spermatocytes (Beall *et al.* 2007; White-Cooper 2010; Laktionov *et al.* 2018). Knockdown of tMAC components and tTAFs results in arrest at the meiotic stage (Lin *et al.* 1996; White-Cooper *et al.* 1998; White-Cooper *et al.* 2000; Ayyar *et al.* 2003; Jiang and White-Cooper 2003; Jiang *et al.* 2007; White-Cooper 2010; Doggett *et al.* 2011).

Given the importance of spermatocyte transcription in sperm development in *D. melanogaster*, it was of interest to examine the potential role of spermatocyte transcription in *D. pseudoobscura*. Transcripts expressed in the spermatocytes are translated in later stages, notably the spermatid stage, during which elongation occurs and the size classes of the *D. pseudoobscura* sperm morphs become apparent. I hypothesised that there would be transcriptome variation between subsets of the spermatocyte cysts in *D. pseudoobscura*, reflecting the transcriptional control of morph-specific differentiation. I further hypothesised that the perdurance of transcripts from the spermatocyte stage into the spermatid stage would also reflect the differential development of the eusperm and parasperm spermatids.

7.1.4 Questions raised and project aims

The current understanding of sperm development in *D. pseudoobscura* raises several questions, which this project aimed to answer.

1. How early in development do germline cells exhibit morph-specific differentiation?
2. Are there subsets of spermatocyte cysts, between which transcriptome variation can be detected?
3. Is there a role for differential gene expression in the spermatocyte cysts in later development stages?
4. Does tMAC, a transcription factor complex required for spermatogenesis and spermiogenesis in *D. melanogaster*, also have a role in *D. pseudoobscura* spermatogenesis?
5. Does tMAC have a role in differential gene expression between spermatocyte cysts of different morphs?
6. How does the stem cell niche of *D. pseudoobscura* differ from that of *D. melanogaster*, and is there a potential role of *D. pseudoobscura* testis hub structure in the development of multiple sperm morphs?
7. Can genetic tools such as CRISPR, currently in use in *D. melanogaster*, be applied to *D. pseudoobscura*?

7.2 Genetic Tools Are Lacking in *D. pseudoobscura*

A major issue with the use of *D. pseudoobscura* as an experimental system is the lack of genetic tools available, compared to *D. melanogaster*, such as UAS-Gal4 expression drivers, RNAi lines, standard mutant lines such as *yellow*, and balancer chromosomes. An early aim of this project was to develop the CRISPR/Cas9 gene editing system for use in *D. pseudoobscura*. It was hoped that the development of *D. pseudoobscura* lines expressing Cas9 would facilitate functional investigation of spermatogenesis genes of interest in this work.

Various methods for use of Cas9 gene editing have previously been described in *D. melanogaster* (Bassett and Liu 2014; Port *et al.* 2014). Early examples of CRISPR in *D. melanogaster* relied on injection into embryos of a plasmid containing the *cas9* sequence under the *hsp70* promoter, and a second plasmid for the expression of a single guide RNA for the target sequence (Gratz *et al.* 2013). A second early method involved co-injection of *cas9* mRNA and sgRNA, and was reported have higher rates of mutagenesis (Bassett *et al.* 2013). Alternatively, embryos of transgenic lines expressing Cas9 in the germline can be injected with a plasmid encoding the sgRNA. In this approach, *cas9* is integrated into the genome under a germline promoter, such as *vasa* or *nanos* (Ren *et al.* 2013; Sebo *et al.* 2014). The benefit of this approach is that the mutagenesis in the somatic cells is limited, since the Cas9 is expressed only in the germline. It has also been found to have high efficiency in generation of mutations transmitted to progeny – above 50-70% in some *D. melanogaster* lines– and is both cost and time effective (Bassett *et al.* 2013; Gratz *et al.* 2014; Port *et al.* 2014; Gratz *et al.* 2015).

7.2.1 Development of *D. pseudoobscura* Cas9 lines

To use CRISPR/Cas9 gene editing in *D. pseudoobscura*, I aimed to develop transgenic lines expressing the Cas9 protein in the germline, by integrating the *cas9* gene into the genome under the *D. pseudoobscura nanos* promoter. I injected *D. pseudoobscura* embryos with a piggyBac plasmid containing the *nos-cas9* sequence and AmCyan under an eye expression promoter, *3xP3*, to enable screening for mutants.

Progeny of F0 *D. pseudoobscura* SLOB3 wild type injected with the piggyBac-3xP3-AmCyan-nos-Cas9 plasmid were screened for the AmCyan eye marker phenotype. A total of 10 transgenic F1 were collected and used to establish stable Cas9 lines. The AmCyan phenotype was less detectable in *D. pseudoobscura* compared to the phenotype in *white* background *D. melanogaster*. Cyan fluorescence detection was limited to the pseudopupil and the ocelli, probably due to interference from the red eye pigment present in wild-type *D. pseudoobscura*, as has previously been observed in other *Drosophila* transformed with the piggyBac-3xP3-Amcyan vector (Schetelig and Handler 2013).

PCR of the ovary cDNA of stable lines expressing the AmCyan eye marker was used to confirm the expression of *cas9*. PCR of each line used in subsequent testing amplified fragments corresponding in size to the *cas9* transcript sequence. To test the function of the Cas9 protein in transgenic lines, guide RNA expression vectors were injected into embryos of Cas9 lines, targeting the *white* and *yellow* gene orthologues in *D. pseudoobscura*. These genes were selected as mutants are easily screened by their modified pigmentation phenotype.

7.2.2 Mutagenesis was unsuccessful in *D. pseudoobscura* Cas9 lines

I was unable to identify any mutants from injection of the *white* or *yellow* sgRNA expression constructs. It is not clear why CRISPR/Cas9 of *D. pseudoobscura white* and *yellow* was unsuccessful, but possible explanations include:

1. The *cas9* mRNA was not localised to the germline.
2. The Cas9 protein was not translated.
3. The Cas9 protein was not functional.
4. Cas9 was functional, but guide RNAs were not efficient in targeting the *white* and *yellow* genes.
5. Cas9 was functional, but guide RNAs were not transcribed.
6. Cas9 was functional, but mutations were synonymous or did not alter phenotypes.
7. Mutations were lethal, reduced survival or reduced fertility.
8. The system was functional, but efficiency was very low.

The presence of *cas9* mRNA was confirmed by PCR of ovary cDNA, indicating that it was expressed in the target tissue. *In situ* hybridisation for *cas9* in ovaries would further confirm the presence of *cas9* mRNA in the germline, and would confirm whether *cas9* is localised within the ovaries to the posterior pole of the oocyte (Forrest and Gavis 2003). However, the cDNA PCR evidence suggests that the *cas9* mRNA is present in the ovaries, and therefore should be available for translation in the germline. *cas9* transcription appeared to be variable between lines, and may have influenced the efficiency of mutagenesis.

It is possible that the lack of mutants was due to Cas9 protein not being translated in the germline, or that the Cas9 protein was not functional. It is less likely that the Cas9 protein itself was not functional, as the *cas9* gene was obtained by restriction digest from a construct which has previously been used in Cas9 gene editing in *D. melanogaster* (Bassett *et al.* 2013). It is possible that the process of ligation and transformation to insert the *nos* promoter and UTR sequences resulted in mutations in the *cas9* sequence, in turn resulting in a non-functional protein. Lack of translation in the germline is possible, for example due to mutations in the *nos* promoter and UTRs. Sequencing during construct assembly

did not identify mutations in the 5' and 3' UTRs, although sequencing was used to confirm the cloning junctions, rather than the whole construct. Translation of the Cas9 protein could be confirmed by immunostaining.

It is unlikely that the lack of mutants was due to all guide RNAs not efficiently binding to the target gene regions. Of the guide RNAs injected, Y1, Y2 and W2 were previously untested, and so may not be efficient in mutagenesis. However, the PLR013, 14 and 15 guide RNAs have been used previously in *D. pseudoobscura* to generate *white* mutants, by injection of embryos with the Cas9-gRNA nucleoprotein complex (Phadnis Lab, referenced in Schroeder *et al.* 2020; Dean Castillo, pers. comm. 2021). This suggests that the lack of mutants was not due to the guide RNA designs. Similarly, injected flies may have contained mutations that were not detectable by screening for *white* and *yellow* mutant phenotypes, for example if the mutations were synonymous or did not result in non-functioning protein. This is also unlikely as the PLR013-15 guide RNAs have been used successfully in the past. It is possible that the guide RNA was not expressed from the pCFD3-U6:3 expression vector, or had very low expression, as the expression vector is optimised for use in *D. melanogaster* and had not been previously tested in *D. pseudoobscura*.

There is a possibility that Cas9 mutagenesis was successful, but that mutations in the *white* and *yellow* genes were lethal or sterile, or reduced viability or fertility of mutant flies. Survival rates were lower than expected for guide RNA injected embryos, approximately 1%, in comparison to previous injections with piggyBac constructs, for which survival was approximately 10%. There were also individuals in the F1 and F2 progeny of F0 injection survivors which lacked testes, or had not developed full size testes connected to the seminal vesicles, although it not possible to determine whether this was a result of Cas9 induced mutation, or the result of random aberrations in development. Again, it is unlikely that all mutations in the *white* and *yellow* genes resulted in death or infertility, as these *D. pseudoobscura* genes have been successfully targeted by others. Lack of viable or fertile adults may have arisen from toxicity of the Cas9 protein (Port *et al.* 2014).

It is unclear why no *white* and *yellow* mutants were identified after guide RNA injection of *D. pseudoobscura* expressing endogenous *cas9* mRNA, but is most likely due to either a lack of functional localised Cas9 protein in the germline, or is a result of low viability or fertility of Cas9 lines and/or mutants. It would be possible to further explore the underlying causes for the failure of gRNA injection to generate mutants, but it may be more useful to explore alternative methods to apply CRISPR/Cas9 gene editing in *D. pseudoobscura*, which do not rely on endogenous expression of Cas9.

7.2.3 Alternative approaches for CRISPR/Cas9 gene editing in *D. pseudoobscura*

Early applications of CRISPR/Cas9 in *D. melanogaster* utilised expression vectors to inject *cas9* DNA or *in vitro* transcribed *cas9* mRNA directly into the posterior region of the embryo (Bassett *et al.* 2013; Gratz *et al.* 2013). These approaches could be used in *D. pseudoobscura* as an alternative to the endogenous Cas9 approach used here. This would be a quicker method to test, as it would not rely on injection, screening and the multiple generations of crosses required to generate stable Cas9 lines in *D. pseudoobscura*. It would also have the benefit of decreasing exposure to the potentially toxic Cas9 in the progeny of injected F0, thus reducing the risk of Cas9 decreasing viability and fertility in the subsequent generations. An alternative method would be to inject the pre-formed ribonucleoprotein complex of Cas9 and transcribed guide RNA directly into the embryos (Lee *et al.* 2014). This method is costlier, but also has the benefit of reducing the deleterious effects of transient Cas9 expression, and has been used by others in *D. pseudoobscura* (Dean Castillo, pers. comm. 2021) and other non-model *Drosophila* species (Kalajdzic and Schetelig 2017).

The process of generating stable lines expressing endogenous Cas9 has demonstrated the use of the piggyBac transposon system in *D. pseudoobscura*. The piggyBac transposon vector system has been utilised by others to generate tagged proteins of interest in *D. pseudoobscura* (Schroeder *et al.* 2020), for mutant rescue in *D. subobscura* (Tanaka *et al.* 2016), and for insertion of site-specific transgene integration systems such as Cre-*loxP*, PhiC31-*att* and Flp-*FRT* in *Drosophila* and non-*Drosophila* insect species (Horn and Handler 2005; Nimmo *et al.* 2006; Labbe *et al.* 2010; Meredith *et al.* 2011; Kudo *et al.* 2018). The piggyBac vector system has been found here to be an efficient method of insertion into the *D. pseudoobscura* genome. Future work could use this system to generate *D. pseudoobscura* lines with site-specific recombination systems currently available in *D. melanogaster*, and would greatly improve the options for researchers investigating gene function in *D. pseudoobscura*. In combination with CRISPR/Cas9 gene editing and RNAi to examine knockout phenotypes, characterised site-specific recombination lines could then be used for mutant recovery (St Johnston 2013), as well as in development of the wider range of genetic tools available in *D. melanogaster*.

7.3 Characterisation of Gene Expression in *D. pseudoobscura*

Spermatocytes and Spermatids

Prior to this project, there was little available knowledge regarding gene expression in *D. pseudoobscura* spermatocytes and spermatids. To identify candidate genes which may have functions in heteromorphic differentiation during sperm development, I first used a single cyst sequencing approach to characterise the transcriptomes of spermatocyte and spermatid cysts in *D. pseudoobscura*. By using this approach, I was able to identify clusters based on the pre-meiotic cyst

transcriptome data, and from this, identify candidate genes for further validation by RNA *in situ* hybridisation.

7.3.1 RNA-seq of primary spermatocyte and spermatid cysts

7.3.1.1 Two alternative methods for library preparation for low-input RNA-seq

RNA-seq of spermatocyte and spermatid cysts was performed in two rounds. In the first round, ten spermatocyte cysts were dissected and libraries synthesised using the SMART-Seq® v4 Ultra Low Input RNA Kit (Takara Bio) and Nextera XT DNA Library Preparation Kit (Illumina). In the second round, twenty spermatocyte cysts from two males and twelve spermatid cysts from six males were dissected, and libraries constructed with the QIAseq FX Single Cell RNA Library Kit (Qiagen). Both methods were optimised for low input, ideal for spermatocyte and spermatid cysts, which contain a total of 34 and 130 cells respectively.

The SMART-Seq® v4 Ultra Low Input RNA Kit has a cDNA amplification step which ensures successful library preparation from small amounts of starting material. In initial RNA-seq of ten spermatocyte cysts, this conferred a benefit as it ensured that libraries could be prepared from the starting material. In the subsequent RNA-seq study of twenty spermatocyte cysts and twelve spermatid cysts, it was evident that low RNA input could generate useful data, and so a PCR-free approach was used with the QIAseq FX Single Cell RNA Library Kit. The use of a PCR-free approach is beneficial for library diversity, as it negates the potential for PCR bias and PCR duplicates in library synthesis. Both methods are still limited by biases resulting from priming with random hexamers, cDNA synthesis, adaptor ligation and sequencing (Ozsolak and Milos 2011).

7.3.1.2 RNA-seq of primary spermatocyte cysts identified transcriptome variation prior to meiosis

RNA-seq of primary spermatocyte cysts showed transcriptional variation between subsets of cysts, suggesting morph-specific spermatocyte cysts are transcriptionally distinct, prior to meiosis and prior to development of morphological differences between spermatid cysts. Previous research has shown the presence of single morphs in spermatid cysts, but no morph-specific characteristics have been described in earlier cyst stages (Beatty and Sidhu 1969; Hauschteck-Jungen and Maurer 1976; Bircher *et al.* 1995). It has previously been suggested that morph-specific differentiation may occur early in sperm development, during GSC differentiation to spermatogonia or early in the transit amplification stage, or that multiple GSC populations are present in the testis niche (Moore *et al.* 2013). RNA-seq of primary spermatocyte cysts provides evidence for morph-specific differentiation in earlier stages than previously demonstrated.

Analysis of the RNA-seq datasets clustered cysts according to transcriptome. Clustering was based on the assumption of either two or three cyst clusters, based on the number of known mature sperm morphs (Beatty and Sidhu 1969; Alpern *et al.* 2019). While early research identified three sperm morphs in *D. pseudoobscura*, it was unclear whether there was a distinct 'medium' class, or whether a single class of parasperm simply had higher variation in length than eusperm (Joly *et al.* 1989; Joly and Lachaise 1994; Snook *et al.* 1994; Snook 1997). The presence of a distinct 'medium' class, parasperm 2, was re-confirmed in 2019, after much of the analysis described here had been completed based on two morphs. Clustering and data were reanalysed based on the assumption of three morphs.

Clusters were used as a basis for differential gene expression analysis, allowing examination of differentially expressed genes present in the spermatocyte cysts. RNA-seq analysis of ten spermatocyte cysts identified 1399 genes as differentially expressed when the cysts were sorted into two clusters. When the data was sorted into three clusters, 1528 genes were identified as differentially expressed, either between all three clusters, or a subset of the clusters. 867 genes were differentially expressed in DGE analysis based on both two and three clusters. This analysis suggests that there is substantial variation between cyst morphs.

Expanded RNA-seq analysis of twenty spermatocyte cysts from two males also showed substantial variation between clusters, indicating variation between cyst morphs. 1613 genes were found to be significantly differentially expressed when the cysts were sorted into two clusters, 747 when sorted into three clusters, and 433 genes were significant in both two- and three-cluster analysis. Furthermore, analysis of the RNA-seq data of twenty spermatocyte cysts indicated that cluster analysis was robust, as cyst clustering was consistent between individual analysis of cysts from each male and the combination of the two datasets. This indicates that the differences between cysts was greater between cyst morphs than between individuals, providing further evidence that spermatocyte cyst morphs are transcriptionally distinct.

Differential gene expression analysis identified genes of interest in both spermatocyte cyst datasets. Given the large number of genes identified as differentially expressed in single cyst RNA-seq, prioritisation of those for further analysis was required. This was initially performed by cross-referencing to the extensive literature available for *D. melanogaster*. Genes of interest were those with known *D. melanogaster* orthologues which have functions in spermatogenesis, spermiogenesis, regulation of transcription or translation, axoneme assembly or cell cycle. Identification of the *D. pseudoobscura* orthologues of these genes as differentially expressed in primary spermatocytes indicated a potential role in morph-specific sperm development, and as candidates for further validation and exploration. A subset of these genes identified was selected for validation by RNA *in*

situ hybridisation. Genes were selected covering a range of functions in spermatocytes and spermatids, to develop a broad picture of the potential mechanisms involved in differential spermatogenesis.

7.3.1.3 RNA-seq of elongating spermatid cysts shows transcriptome variation relating to spermiogenesis and post-meiotic transcription

RNA-seq of spermatid cysts produced transcriptome data for long and short spermatid cysts. Long and short cysts were assumed to represent eusperm and parasperm cysts respectively, although partially elongated eusperm cysts may have been mistaken for parasperm cysts, and parasperm 1 and 2 cysts were not identified at the time. Despite this, differential gene expression analysis still identified variation in the transcriptomes of long and short spermatid cysts. 1401 genes were identified by DGE analysis, of which 900 were higher in short cysts, and 501 were higher in long cysts. Of these, 24 genes with orthologues in *D. melanogaster* functioning in spermiogenesis were identified. Genes functioning in spermiogenesis are of particular interest in relation to heteromorphic sperm development; as the three sperm morphs differ substantially in size, it would be expected that spermatid elongation will differ between the three morphs, and this would be reflected in the presence of transcripts whose products contribute to elongation.

Four orthologues of genes showing post-meiotic expression and RNA localisation in *D. melanogaster* were also identified. The cup and comet classes of genes are expressed in spermatids, and show localisation of the RNA to the distal tail end of the elongating spermatid (Barreau *et al.* 2008a; Barreau *et al.* 2008b). Two cup gene orthologues, *presidents-cup* (*p-cup*) and *stanley-cup* (*s-cup*) were higher in long cysts, with *walker-cup* (*wa-cup*) and *sungrazer* (*sunz*) higher in short cysts. In *D. melanogaster*, the cup and comet genes are expressed and show RNA-localisation in the spermatid cyst. The functions of these genes are largely unknown, but they do not appear to encode components of the sperm, and may function in the elongation and individualisation processes of spermiogenesis (Dorus *et al.* 2006; Barreau *et al.* 2008a). The finding that some of these genes appear to be differentially expressed between the spermatid morphs in *D. pseudoobscura* suggests that these four cup and comet genes may contribute to morph-specific elongation and control of individualisation processes.

Notably, a greater number of genes identified by DGE analysis of spermatid cysts were found to be higher in short spermatids compared to long spermatids. There are multiple possible reasons for this:

1. The short spermatids represent both parasperm and 'early' elongating eusperm spermatids. Throughout the process of spermiogenesis, mRNAs transcribed in spermatogenesis are translated, and the transcripts degraded (Schafer *et al.* 1995). Early eusperm spermatids have translated, then degraded, fewer mRNAs than later stage spermatids, resulting in the

transcripts being recognised as ‘differentially expressed’ by DGE analysis, and higher transcript counts in short spermatid cysts.

2. Short spermatid cysts are parasperm cysts. Parasperm spermatocyte cysts have greater expression of the subset of genes identified, and these transcripts persist into the spermatid stage, during which they are translated and their products have morph-specific spermiogenesis functions.
3. Short spermatid cysts are parasperm cysts. Parasperm spermatocyte cysts have greater expression of the subset of genes identified, and these transcripts persist into the spermatid stage. Transcripts do not function in morph-specific spermiogenesis, representing non-specific or ‘junk’ expression.

There is some evolutionary argument to support the third hypothesis – that parasperm cysts have higher ‘junk’ expression. Parasperm are not required to be fertilisation competent, and therefore are under relaxed selection (Snook *et al.* 1994; Snook and Karr 1998; Moore *et al.* 2013). Non-fertile sperm have been found to show greater intra-morph size variation than fertile sperm morphs in sperm heteromorphic species (Snook 1997; Bernasconi and Hellriegel 2005). The same might be expected for transcriptional control; relaxed selection pressures result in less stringent transcription regulation, observed as a greater diversity of transcripts represented in an RNA-seq dataset and a greater number of ‘DE’ genes more highly expressed in parasperm.

Of the genes identified as significantly different, some were found to be higher in elongated eusperm spermatids. These gene products may have some function or require enrichment in morph-specific development.

7.3.1.4 Limitations of single cyst RNA-seq analysis

The single cyst RNA-seq approach allowed broad characterisation of the transcriptomes of *D. pseudoobscura* spermatocyte and spermatid cysts, and identified candidate genes which may function in differential development of heteromorphic sperm. However, there are limitations to this analysis, such that conclusions based on this analysis would require further validation.

The greatest limitation of spermatocyte cyst RNA-seq was the lack of clarity over cyst identity. Prior to this work, there were no known markers for any sperm morph, other than morphological variation in the late spermatid cyst stage. To combat this, I used cluster analysis to separate the cyst samples into groups, based on their transcriptomes (Suzuki and Shimodaira 2015). As a result, subsequent DGE analysis would necessarily find differences between the groups. In the case of spermatocyte cyst RNA-seq, DGE analysis must therefore be treated as a method to identify the genes most likely to be contributing to clustering, rather than a direct measure of significant differences in transcript

abundance between morph cysts. Similar issues arise with more typical single-cyst seq approaches, where larger numbers of cells are sequenced, but gene expression analysis must still be based on clustering (Wu *et al.* 2017; Kharchenko 2021).

The second RNA-seq study, RNA-seq of twenty spermatocyte cysts and twelve spermatid cysts, showed substantial reduction in read depth for all samples after alignment to the reference genome. Reduction in read counts occurred when the alignments were filtered for transcripts which had not mapped to the reference genome, or were optical duplicates. This may have been due to contamination of the cyst samples with non-*D. pseudoobscura* material. Despite this, the analysis should still hold some validity, given that clustering did not appear to be linked to read counts, and cluster analysis of M and R spermatocyte cyst datasets was consistent when performed on separate and merged data. There was also overlap between genes found to be differentially expressed between RNA-seq 1 and RNA-seq 2. As with the limitations of this analysis resulting from unknown cyst identity, further validation was required to draw conclusions based on this data.

7.3.1.5 Improvements to RNA-seq analysis based on the developments in this work

As discussed above, RNA-seq analysis was limited by the lack of knowledge of cyst identity prior to sequencing. New analysis could also make use of marker genes, such as *kmg*, *Caf1A*, *achi/vis*, *mil*, *piwi*, *twe*, *bol* and *GA23025*, identified by RNA-seq and subsequently validated by *in situ* hybridisation, to identify cyst clusters prior to DGE analysis. Future repetition of this RNA-seq analysis could utilise the Kmg-GFP tag lines developed as part of this work, to sort spermatocyte or early spermatid cysts according to cyst morph, negating the requirement for cluster analysis prior to DGE analysis.

More recently, I have adapted a spatial RNA-seq analysis for use in *Drosophila* testes and individual cysts, using the 10X Genomics Spatial Gene Expression sequencing platform. In this approach, embedded and sectioned tissue is placed on a slide containing a capture area of thousands of barcoded spots, each containing millions of capture oligonucleotides with unique spatial barcodes. Fixation and staining on the slides allows visualisation of the tissue sections. The tissue sections are permeabilised to release mRNA, which is captured by the spatially barcoded oligonucleotides, and from which reverse transcription produces cDNA which is processed to generate library for sequencing. Sequence data can then be analysed with reference to imaging, to generate cell-specific transcriptome profiles in the tissue of interest. The adapted, optimised method used to generate *D. pseudoobscura* and *D. melanogaster* testis and cyst spatial-seq libraries is available in Appendix 11. Spatial sequencing will allow analysis of the cyst transcriptomes in the context of whole testes, and may also shed further light on transcript localisation within elongated spermatid cysts.

To summarise, RNA-seq analysis of individual cysts demonstrated broad scale variation between subsets of primary spermatocyte cysts, and between long and short spermatid cysts. This variation may be due to transcriptional differences between morphs at these stages. To confirm the presence of transcriptionally distinct morphs, particularly in the pre-meiotic spermatocyte stage, further validation was performed by RNA *in situ* hybridisation for a sample of the genes identified by RNA-seq.

7.3.2 Validation of RNA-seq data by RNA *in situ* hybridisation

To validate RNA-seq analysis of spermatocyte and spermatid cysts, a subset of genes identified as significantly different in abundance in these datasets was studied by RNA *in situ* hybridisation. The *in situ* hybridisation method has some benefits over RNA-seq, in that localisation of transcripts within the testes can be identified, as well as inferring information regarding relative abundances between cysts of differing morphs and development stages.

The dynamics of gene expression in the testes are such that the presence and absence of transcript-specific staining by this method can indicate the stage at which the transcript is used (Morris *et al.* 2009). The vast majority of transcription of genes whose products are required in meiosis and in post-meiotic spermiogenesis occurs during the spermatocyte stage (Olivieri and Olivieri 1965; Fuller 1993; White-Cooper 2010). The perdurance of transcripts from the spermatocyte stage into the spermatid stage indicates that the gene products are not yet required. Translation exposes the mRNA to RNA degradation machinery, which is observed as the disappearance of detectable staining for the transcript by *in situ* hybridisation in later stage cysts (Schafer *et al.* 1995; Barreau *et al.* 2008a; Morris *et al.* 2009).

RNA *in situ* hybridisation has limitations. It is not quantitative: comparisons between probe staining cannot be used to infer relative or absolute transcript abundances, although this method can indicate relative abundances between cell types and stages of a single probe within a tissue sample (Morris *et al.* 2009). There are also issues with limits of detection by this method: low abundances of transcripts, or low probe concentration or efficiency, can reduce staining intensity, thereby reducing transcript detection. There is also a subjective decision about when a given staining reaction is stopped. For example, where some cyst staining is dark, due to a high abundance of the target transcript, the staining reaction may be stopped so that details in the stained cysts are still clear. Other cysts may have the transcript at much lower levels, but staining is not detectable after the length of staining reaction required for dark staining of high transcript abundance cysts. If left for longer, staining may then be detectable in low abundance cysts. Hence, subjectivity can result in interpretation of an on/off expression pattern, rather than a high/low expression pattern.

7.3.2.1 Genes identified by RNA-seq were expressed in the testes

Of the 31 genes for which RNA *in situ* hybridisation staining is presented in this work, nine showed some evidence of morph-specific or morph-enriched expression. In pre-meiotic cysts, patchy staining indicating that the transcript was present in a subset of spermatocyte cysts. In post-meiotic cysts, staining was localised to a single area of the testis. Of the 31 genes for which staining is presented, 30 showed evidence of testis expression, as indicated by the presence of detectable staining.

Genes found to be differentially expressed by DGE analysis of RNA-seq did not all show the patchy expression indicated higher abundance in subsets of primary spermatocyte cysts, or localisation to one spermatid morph. Genes which were identified as patchy by *in situ* hybridisation showed fold changes in RNA-seq analysis ranging from 1.5 to 89.7, likewise genes not identified as patchy by *in situ* hybridisation showed RNA-seq fold changes ranging from 1.5 to 179.7. This may indicate false positive identification of genes by RNA-seq analysis, or may be due to the higher sensitivity of RNA-seq and limits of detection in *in situ* hybridisation staining.

7.3.2.2 Identification of genes of interest in heteromorphic sperm development; a role for tMAC and tMAC regulation

7.3.2.2.1 *D. pseudoobscura kmg*, GA18735, is differentially expressed in spermatocyte cysts

Of the genes identified by RNA-seq and RNA *in situ* hybridisation, one of the most striking examples of differential gene expression was GA18735, the *D. pseudoobscura* orthologue of *kumgang* (*kmg*). *D. pseudoobscura kmg* was identified as differentially transcribed between spermatocyte cysts by RNA-seq. *In situ* hybridisation for *kmg* showed strong patchy staining in the regions of the testis containing spermatocyte cysts, with no detectable staining in earlier or later stages. This suggests that *kmg* is expressed in early through late spermatocytes, and that *kmg* translation has probably occurred prior to spermatid elongation. *D. pseudoobscura Kmg* may therefore have a role in spermatocytes, meiosis or early spermatids, based on this RNA expression. The differential mRNA accumulation also suggests that *Kmg* has some morph-specific or morph-enriched functions in *D. pseudoobscura* cysts at these stages.

D. melanogaster Kmg is a transcription regulator, involved in the regulation of expression by tMAC. *Kmg* prevents the binding of tMAC to cryptic promoters of genes normally only expressed in somatic cells. This prevents mis-expression in spermatocytes, and maintains germline-specific expression (Kim *et al.* 2017). Given the important role *Kmg* has in transcription regulation in *D. melanogaster*, and the strong evidence of differential expression in *D. pseudoobscura* spermatocyte cysts, *D. pseudoobscura kmg* was selected for further investigation (discussed in 7.4).

7.3.2.2.2 *GA21345*, the *D. pseudoobscura* orthologue of the *D. melanogaster achi* and *vis* paralogues, shows isoform dependent differential expression in spermatocyte cysts

GA21345 is the *D. pseudoobscura* orthologue of the *D. melanogaster* genes *achintya* (*achi*) and *vismay* (*vis*), encoding paralogous DNA binding proteins which co-immunopurify with tMAC components *aly* and *comr* (Ayyar *et al.* 2003; Wang and Mann 2003; Beall *et al.* 2007). *D. pseudoobscura achi/vis* was identified as differentially expressed in RNA-seq of spermatocyte cysts. *In situ* hybridisation showed patchy expression in spermatocyte cysts, suggesting higher transcript abundance in a subset of cyst morphs. *achi/vis* has multiple isoforms, of which two, *-RA* and *-RC*, contain a third exon only expressed in the testes, both in *D. pseudoobscura* and *D. melanogaster* (Ayyar *et al.* 2003; Gramates *et al.* 2017; Yang *et al.* 2018).

In situ hybridisation for *D. pseudoobscura achi/vis* with a probe targeting all three isoforms showed strong staining in the early spermatocytes, with weaker staining in earlier and later cells, suggesting that *achi/vis* is transcribed in early spermatocytes, and translated in the spermatocyte stage. Probe targeting exon 3, present only in the *-RA* and *-RC* isoforms, showed patchy staining in the later stage spermatocytes. *achi/vis-RA* and/or *-RC* isoforms may be translated earlier in a subset of cysts, or *achi/vis* may be transcribed for longer in a subset of spermatocyte cysts, and the transcript splicing includes the testis-specific third exon.

Staining with a probe spanning the exon 2-exon 4 splice junction did not show patchy staining in the spermatocyte cysts. This probe was designed to recognise the *achi/vis-RB* isoform, which lacked the third exon, however it likely also targeted regions within the *-RA* and *-RC* isoforms, and therefore *-RB* specific staining could not be fully determined. The staining pattern observed suggests that *achi/vis-RB* is uniformly expressed in all spermatocyte cysts, and is translated in spermatocyte cysts.

Given the DNA-binding and transcription regulation functions of *D. melanogaster* Achi and Vis, it is possible that *D. pseudoobscura* Achi/Vis has a similar function. Furthermore, cyst-specific variation in the perdurance of isoform(s) containing the third exon suggests that those isoforms may encode an isoform of the Achi/Vis protein required in morph-specific development.

7.3.2.2.3 *D. pseudoobscura* orthologues of *Caf1*, *GA18051* and *GA26389*, show stage-specific transcript localisation and morph-specific enrichment in spermatocyte cysts

D. melanogaster Caf1, like *achi/vis*, encodes a component of tMAC. In *D. pseudoobscura*, two orthologues of *Caf1* have been identified (Calvo-Martin *et al.* 2017). The first, *GA18051* or *Caf1A*, is syntenic to *D. melanogaster Caf1*. The second, *GA26389* or *Caf1B*, is a DNA duplicate located on the same chromosome (2). *Caf1A* and *Caf1B* were both found to be expressed in the testes by *in situ* hybridisation. The *Caf1A* paralogue was found to have stronger staining in a subset of spermatocyte

cysts, indicating that there may be higher expression of *Caf1A* in some cyst morphs. *Caf1A* staining was evident up to the apical tip of the testis, and the spermatogonia and spermatocyte regions of the testis, indicating that *Caf1A* transcription occurs early in spermatogenesis. *Caf1A* may be continue to be transcribed throughout the spermatogonia and spermatocyte stages, or transcripts present in the earlier stages may be stored and translated in later stage spermatocytes, with translation occurring earlier in a subset of cysts. *Caf1A* may have a function in morph-specific development.

Caf1B did not show patchy expression, indicating that this paralogue does not have cyst-specific or cyst-enriched expression or perdurance. *Caf1B* staining is localised to the late spermatogonia and early spermatocyte region of the testis, indicating that it is transcribed in late spermatogonia or at the entry to the spermatocyte stage, and is translated in spermatocytes. *Caf1B* may therefore have a function in all spermatocyte cysts.

Caf1A and *Caf1B* functions may share similarities with the orthologous *D. melanogaster* *Caf1*, which encodes a component of tMAC, as well as a paralogous complex dREAM/MybMuvB (Beall *et al.* 2007), and other chromatin remodelling complexes including Nucleosome Remodelling Factor (NURF), Polycomb repressive complex 2 (PRC2), Chromatin assembly factor 1 (CAF-1), and Nucleosome remodelling deacetylase (NuRD) (Ketel *et al.* 2005; Clapier and Cairns 2009; Alkhatib and Landry 2011; Schuettengruber *et al.* 2011; Yu *et al.* 2013). It is interesting therefore, that the *Caf1A* and *B* paralogues in *D. pseudoobscura* appear to be expressed at different stages of spermatogenesis, and show different expression patterns, with *Caf1A* showing differential expression or differential perdurance in the spermatocytes. It is possible that the *Caf1* paralogues in *D. pseudoobscura* each contribute to different versions of tMAC. Alternatively, one may have become specialised to tMAC, and the other has become specialised to other *Caf1*-containing complexes. *D. melanogaster* *Caf1* is a component of multiple complexes involved in chromatin remodelling and regulation of transcription (Bulger *et al.* 1995; Mizuguchi *et al.* 2001). Assuming the *D. pseudoobscura* *Caf1* paralogues have similar functions, *Caf1A* and *B* may also be components of chromatin remodelling complexes, required for access to promoters of genes for germline expression. Higher expression of *Caf1A* in a subset of cysts may indicate a role for regulating morph-specific gene expression or upregulation.

7.3.2.2.4 The Multiple tMACs Hypothesis: do *Caf1* paralogues and *Achi/Vis* isoforms contribute to morph-specific versions of tMAC?

Two genes, for which *D. melanogaster* orthologues are components of tMAC and are required for meiosis and spermiogenesis, were found to show evidence of differential expression or perdurance in *D. pseudoobscura* spermatocyte cysts. tMAC is required for the expression of over 1000 genes in *D. melanogaster* spermatocytes (White-Cooper 2010; White-Cooper and Davidson 2011; Laktionov *et al.*

2014; Laktionov *et al.* 2018). It is therefore of interest what contribution these genes may have in *D. pseudoobscura* spermatogenesis, particularly in that of the multiple sperm morphs.

One hypothesis for the function of these genes is that the Caf1 paralogues and the Achi/Vis isoforms contribute to different versions of tMAC. The versions of tMAC may be morph-specific, or morph-enriched in spermatocytes, contributing to a morph-specific transcriptome in these cells. Transcription in the spermatocytes is then required for morph-specific morphogenesis during the spermatid stage, giving rise to the multiple sperm morphs.

There is some precedent for the Multiple tMAC Hypothesis; co-immunoprecipitation studies have found evidence of two alternate versions of tMAC, one containing Aly, Topi, Tomb, Comr, Mip40 and Caf1A, and the other containing Achi/Vis, but not Mip40 (Wang and Mann 2003; Beall *et al.* 2007). Based on this, up to four *D. pseudoobscura* tMAC versions could be hypothesised; tMAC-Caf1A, tMAC-Caf1B, tMAC-Achi/Vis(A/C) and tMAC-Achi/Vis(B) (Figure 7.1; A). Alternatively, Caf1A and Achi/Vis(A/C) may contribute to a single tMAC, which is highly expressed in one or more spermatocyte morphs. Caf1B and Achi/Vis(B) may contribute to a single tMAC, which is consistently expressed across early spermatocyte cysts (Figure 7.1; B).

To establish the potential role of tMAC in *D. pseudoobscura* morph-specific spermatogenesis, and to further understand the functions of the *Caf1A* and *Caf1B* paralogues, and *achi/vis* isoforms, further work is needed. A similar approach to that described in Chapter 5 could be taken, to establish the patterns of protein localisation of Caf1A, Caf1B and Achi/Vis in *D. pseudoobscura* spermatogenesis (Appendix 12). Further co-immunoprecipitation studies may also aid with further understanding of how tMAC may contribute to the development of multiple morphs in *D. pseudoobscura*, establishing whether multiple forms of tMAC are present in *D. pseudoobscura* testes, which of the meiotic arrest genes contribute to each complex, and identification of target genes by ChIP-seq.

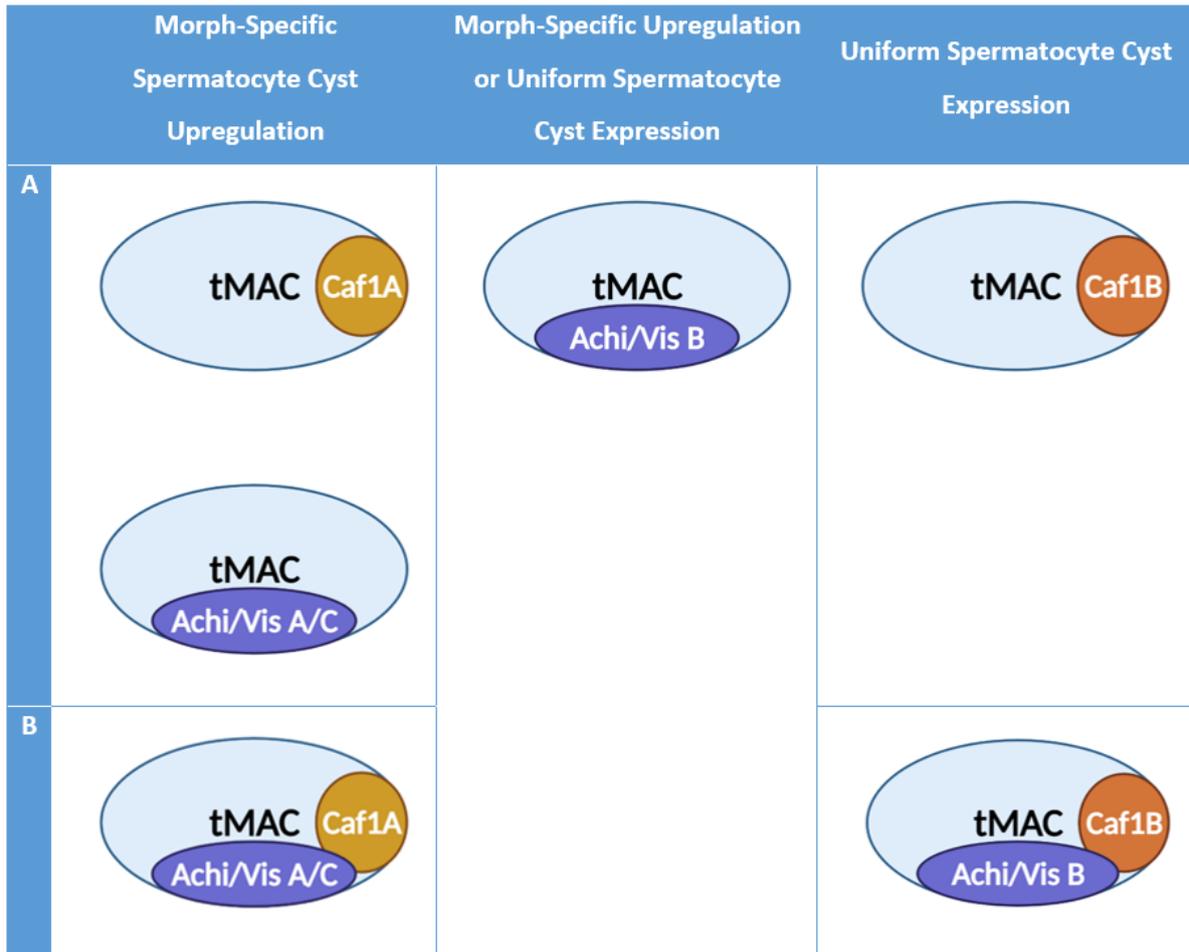


Figure 7.1: The Multiple tMAC Hypothesis: Caf1A and Caf1B paralogues and achi/vis alternate isoforms may contribute to multiple versions of tMAC. A: Up to four tMAC versions, two of which are upregulated in one spermatocyte cyst morph (tMAC-Caf1A and tMAC-Achi/VisA/C), one which is consistently expressed across all spermatocyte cyst morphs (tMAC-Caf1B), and one for which current evidence cannot determine morph-specific expression (tMAC-Achi/VisB). B: Two versions of tMAC, one showing cyst-specific upregulation (tMAC-Caf1A-Achi/VisA/C), one showing consistent expression across morphs (tMAC-Caf1B-Achi/VisB). An alternative tMAC-Caf1A-Achi/VisB could also be differentially expressed between cysts, in hypothesis B. A and B are not mutually exclusive.

7.3.2.3 Meiosis regulators – morph-specific or stage-specific expression?

Two orthologues of *D. melanogaster* regulators of entry into meiosis were found to be differentially expressed by RNA-seq and showed evidence of patchy staining in RNA *in situ* hybridisation; *boule* and *twine*. The *D. pseudoobscura* *boule* (*bol*) orthologue, *GA18412*, was identified in RNA-seq of spermatocyte and spermatid cysts as differentially expressed. Staining for *bol* detected the transcript in spermatogonia (weaker staining), spermatocytes and early spermatids. Staining for the *twine* (*twe*) orthologue, *GA18558*, detected the transcript in a much more localised area of the testis, and appeared to be present in late spermatocyte to early spermatid stage cysts.

D. melanogaster boule is a regulator of *twine*; translocation of Boule from the nucleus is required for translation of Twine, itself required for the initiation of meiosis I (Alphey *et al.* 1992; Courtot *et al.* 1992; Lin *et al.* 1996; Maines and Wasserman 1999). This raises an interesting question – if *boule* and *twine* are required for meiosis, why are they identified as differentially expressed in RNA-seq spermatocyte and spermatid data, and showing stronger staining in subsets of cysts RNA *in situ* hybridisation? There are two obvious hypotheses for this, assuming *twe* and *bol* have similar functions in *D. pseudoobscura* as their *D. melanogaster* orthologues. First, entry into meiosis requires different levels of *bol/twe* accumulation dependent on the cyst morph. Second, that *bol/twe* have stage-specific expression, which is observed in RNA *in situ* hybridisation as patchy staining, similar to that observed in *kmg*, *caf1A* and *achi/vis*. It is possible that *bol* and *twe* have different functions in *D. pseudoobscura*, and that the apparent differential expression between cysts contributed to morph-specific or morph-upregulated functions.

7.3.2.4 Spermiogenesis Genes: eusperm and parasperm-specific morphogenesis?

The *D. melanogaster* genes *dic61B* and *mil* are required during spermiogenesis. Patchy staining was detected by *in situ* hybridisation for the *D. pseudoobscura* orthologues of *dic61B* (GA20060) and *mil* (GA27003), indicating that they may have morph-specific upregulation or expression.

D. pseudoobscura dic61B was identified as differentially expressed in RNA-seq of spermatocytes, and showed staining in the spermatocyte and early spermatid cysts. In *D. melanogaster dic61B* encodes a dynein intermediate chain, a component of the axoneme and required for axoneme assembly in the elongating spermatid (Fatima 2011; Zur Lage *et al.* 2019). *D. pseudoobscura dic61B* appears to be transcribed in the spermatocyte stage, and the transcript persists into the early spermatid stage. Assuming a similar function, *dic61B* would be translated in the early spermatid stage as axoneme assembly begins. Upregulation of *dic61B* expression in a subset of spermatocyte cysts indicates morph-specific upregulation, which hypothetically may be required in elongating eusperm spermatid cysts, where a greater abundance of axoneme components are required for assembly of the longer flagellum.

The orthologue of a second spermiogenesis gene, *mil*, was identified by differential expression analysis in RNA-seq data of spermatocyte cysts. *In situ* hybridisation showed patchy staining in the spermatocyte and early spermatid cysts, similar to that observed in *kmg* staining. The staining pattern suggests that *mil* is transcribed in the spermatocytes, and translated in the elongating spermatids. In *D. melanogaster mil* is testis specific, and required for spermatid nuclear localisation and elongation (Kimura 2013). *D. melanogaster mil* mutants show aberrations in spermatid elongation, resulting in shorter and misshapen cysts, and have scattered nuclei along the length of spermatids within a cyst,

rather than clustered at the basal end of the cyst. Nuclei are also smaller and oval shaped in *D. melanogaster mil* mutants. *D. melanogaster* mutants do not appear to show any defects prior to meiosis, which is consistent with the pattern of transcript staining observed in *D. pseudoobscura. mil* may have a similar function in *D. pseudoobscura*, contributing to nuclear shaping, localisation and spermatid elongation. Given the *D. melanogaster* mutant phenotype, it could be hypothesised that parasperm may have lower abundance of *mil*, resulting in the reduced elongation in the short morphs. Eusperm and parasperm vary in overall length and nucleus length; the parasperm nuclei are shorter than that of the eusperm, and parasperm 1 and 2 nuclei also correlate with length (Figure 7.2) (Pasini *et al.* 1996; Moore *et al.* 2013). It would be expected that *mil* was still expressed at some level in the parasperm, as was found in the *in situ* hybridisation staining, as it is required for complete spermatid morphogenesis.

The *D. pseudoobscura* gene *GA23025* does not have a known *D. melanogaster* orthologue. It is testis-specifically expressed, and appears to share some protein domains (NAP protein domain) with Mil (Kimura 2013; Gramates *et al.* 2017). NAP family proteins have also been implicated in nucleosome assembly and regulation of transcription (Tachiwana *et al.* 2008; Doyen *et al.* 2015). *GA23025* was found to be differentially expressed in spermatocyte and spermatid cyst RNA-seq data, and showed patchy expression in *in situ* hybridisation staining. Staining indicated that *GA23025* is expressed in the spermatocytes, and that the transcript perdures throughout the spermatocyte stage, into the early spermatid cysts, in which *GA23025* is translated. *GA23025* appeared to be more abundant in some spermatocyte cysts, indicating higher expression in a subset of spermatocyte cyst morphs. NAP protein function indicates that *GA23025* may function in up or downregulating genes in a specific morph, via a chromatin reassembly transcription regulation mechanism.

The *D. pseudoobscura* orthologue of *piwi*, *GA19370*, was identified in spermatocyte cyst RNA-seq analysis, but *in situ* hybridisation staining was only detected in spermatids. The incongruous results may have been a result of low probe sensitivity, or due to the higher sensitivity of RNA-seq. Staining appeared stronger in the most elongated cysts, indicating that *piwi* RNA may be more abundant in eusperm spermatid cysts – either in the germline or cyst cells. *D. melanogaster piwi* functions in silencing transposons in spermatogenesis (Ku *et al.* 2016). In *D. pseudoobscura*, *piwi* may have a similar function in eusperm spermatids, or an alternative function.

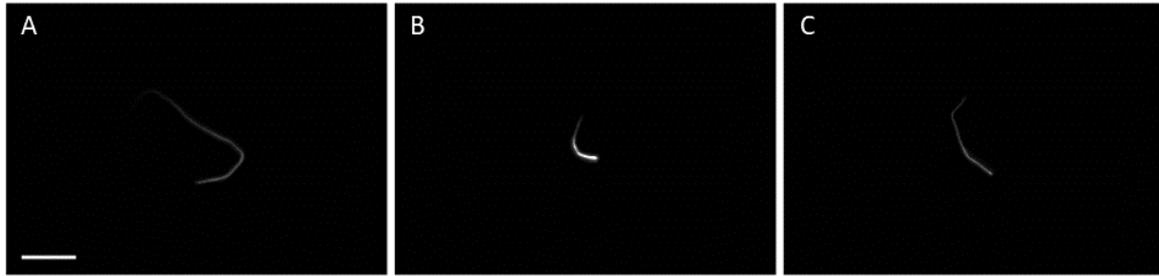


Figure 7.2: Nuclei of mature sperm extracted from *D. pseudoobscura* seminal vesicles, stained for DNA with Hoechst 33258. Scale = 100 μ m. A: Eusperm nucleus. B: Parasperm 1 nucleus. C: Parasperm 2 nucleus.

7.3.2.5 Limitations of inferring functions between species

There is very little information regarding the functions of specific genes and their products in *D. pseudoobscura*. Some functional information can be inferred by predicting protein domains based on gene sequences. Inferring function from orthologous genes in *D. melanogaster*, for which there is much more data available, can help in formation of hypotheses regarding function, but does require further experimental testing.

Gene function could be further examined in *D. pseudoobscura* by piggyBac transposon insertion of fluorescent tag constructs, as was used here with *kmg*. This can be used to examine protein localisation, in particular, to confirm the differential presence of protein in spermatocyte cysts pertaining to specific morphs. An experimental method to test the combined expression of multiple differentially expressed genes and their protein products would be to use a fluorescent *in situ* hybridisation method to target multiple transcripts within a single testis (Jandura *et al.* 2017; Yang *et al.* 2017). This could be combined with analysis of testes expressing multiple tagged proteins of interest, each tagged with a different fluorescent marker, to examine differential or co-localisation between or within cyst morphs.

The lack of available genetic tools in *D. pseudoobscura* has limited examination of mutant phenotypes for genes of interest in this project. Development of RNAi, or further development of CRISPR/Cas9 gene editing, will enable generation of new mutant lines for the genes of interest identified here.

7.3.2.6 Examples of genes which may be post-meiotically expressed in *D. pseudoobscura*

The majority of transcription in sperm development takes place in the spermatocytes (Olivieri and Olivieri 1965; Gould-Somero and Holland 1974; Schafer *et al.* 1995; Li *et al.* 2022). A small number of genes have been found to have post-meiotic expression in the spermatids (Barreau *et al.* 2008a; Barreau *et al.* 2008b; Vbranovski *et al.* 2010). RNA *in situ* hybridisation showed staining which indicated post-meiotic expression of five genes, *wa-cup*, *piwi*, *mip40B*, *Sf3b5* and *dila*. *wa-cup* was

previously identified as post-meiotically expressed in *D. melanogaster* (Barreau *et al.* 2008a). Post-meiotic expression of the genes identified as potentially expressed in spermatids requires further validation by cyst qPCR. Alternatively, a single molecule fluorescent *in situ* hybridisation method could be used to visualise nascent transcripts in the nuclei of spermatids, as has been demonstrated in *D. melanogaster* (Yamashita Lab, Jacklyn Fingerhut, pers. comm.). Further study of these genes in *D. pseudoobscura* may aid in the understanding of the roles of post-meiotically expressed genes throughout the *Drosophila* species group.

7.4 Further Investigation of *D. pseudoobscura kmg* by piggyBac Insertion of GFP-Tagged Kmg for Endogenous Expression

7.4.1 Analysis of GFP-tagged Kmg expression in spermatocyte cysts confirms differential transcription and translation of *D. pseudoobscura kmg*

As previously discussed, the *D. pseudoobscura* orthologue of *kmg* was identified by differential gene expression analysis of spermatocyte cyst RNA-seq. Subsequent *in situ* hybridisation for *kmg* showed a pattern of strong patchy staining, indicating that *kmg* mRNA abundance was higher in a subset of cysts. The pattern of *kmg* staining in testis suggested that *kmg* is expressed in early spermatocytes, and that translation occurs prior to meiosis. Based on *in situ* hybridisation, it was hypothesised that either transcription in primary spermatocytes continues up to the transition to meiosis, with translation taking place simultaneously, or transcripts expressed in early primary spermatocytes are stored until translation in later primary spermatocytes. Analysis of GFP-tagged Kmg indicates the former, that *kmg* is transcribed and translated throughout the primary spermatocyte stage, and that the protein persists into the early spermatids. Kmg-GFP was also detected in earlier cysts, in late stage spermatogonia, suggesting that *kmg* expression may be initiated in the spermatogonia, with some translation prior to the onset of the pre-meiotic S-phase. Comparisons between cysts showed that Kmg-GFP was much lower in the spermatogonia cysts, indicating that the majority of *kmg* translation, in line with transcription, takes place in the spermatocyte cysts.

GFP fluorescence was much higher in a subset of spermatocyte cysts, indicating that Kmg protein is much more abundant in these cysts. Spermatogonia, spermatocyte and early spermatid cysts imaged by fluorescence microscopy with a GFP filter were all heteromorphic for Kmg-GFP abundance. This is consistent with the hypothesis that there are multiple spermatocyte cyst morphs, corresponding to each mature sperm morph. The results presented in this work support a hypothesis that Kmg contributes to heteromorphic spermatogenesis in *D. pseudoobscura*. Imaging of whole testes by Lattice Light Sheet microscopy, which has higher sensitivity than our conventional epifluorescence

microscope, indicated that Kmg-GFP was present in all spermatocyte cysts, with much higher fluorescence in a subset of cysts. This suggests that Kmg may have a dose-dependent contribution to heteromorphic sperm development, with at least one cyst morph requiring higher levels of Kmg than others.

There were at least two levels of Kmg-GFP fluorescence observed in spermatocyte cysts. The high GFP and very low to undetectable GFP were easily distinguished. An intermediate GFP fluorescence intensity was observed in some cysts, suggesting that there may be three discrete Kmg expression levels in spermatocyte cysts, corresponding to the three sperm morphs. This was also observed in spermatocyte cyst staining by *in situ* hybridisation for *kmg*. In RNA-seq data, analysis based on three clusters of M and R spermatocyte cysts also showed three levels, low (group 1), medium (group 2.1) and high (group 2.2). The 'medium expression' cysts were not easily distinguishable from the 'high expression' cysts in *in situ* hybridisation and Kmg-GFP cyst imaging by visual inspection. Further quantification by single cyst qPCR of a range of cysts from a single testis would be required to confirm the 'medium' *kmg* expression spermatocyte cyst class.

GFP-tagged Kmg was localised to the nuclei of spermatocytes and early spermatids. Notably, fluorescence was not detected in the nuclei of cyst cells, suggesting that Kmg has a germline specific expression and function in the testes. *D. melanogaster* Kmg is also localised to the spermatocyte nuclei, similar to the localisation observed in *D. pseudoobscura* (Kim *et al.* 2017). *D. melanogaster* Kmg shows some enrichment within the nucleus to the chromosomes. The same localisation was not clearly observed in *D. pseudoobscura*, although some nuclear sub-localisation was evident in Lattice Light Sheet and some fluorescence microscopy images. Further imaging of Kmg-GFP testes stained for DNA, by Light Sheet or confocal microscopy, may give greater clarity regarding Kmg-GFP localisation within the testes of *D. pseudoobscura*.

Kmg-GFP was detectable in the nuclei of early spermatids, but was not detectable in elongating spermatid cysts. It was therefore not possible to distinguish whether Kmg is more abundant in eusperm, parasperm 1, or parasperm 2 cysts. Establishing downstream targets of Kmg may indicate which morph(s) have higher abundance of Kmg, and its functions in heteromorphic sperm development.

7.4.2 *D. melanogaster* and *D. pseudoobscura* Kmg alignments show similar protein structure

Both *D. melanogaster* and *D. pseudoobscura* Kmg contain five C2H2 zinc finger DNA binding regions indicating functions in transcription (Chung *et al.* 2002; Kim *et al.* 2017). Alignment of the protein

sequences showed 58% similarity between *D. melanogaster* and *D. pseudoobscura* Kmg. Alignment suggests that the *D. pseudoobscura* Kmg has some functional similarities to *D. melanogaster* Kmg.

It is unclear whether *D. pseudoobscura* Kmg regulates transcription of a similar subset of genes as *D. melanogaster* Kmg. Likewise, it remains unclear whether *D. pseudoobscura* Kmg regulated transcription by preventing tMAC binding to promoters. A similar target set of genes would indicate a similar mechanism of function.

7.4.3 Hypotheses for Kmg functions in *D. pseudoobscura* heteromorphic spermatogenesis

The evidence presented in this work suggests that *kmg* functions in heteromorphic sperm development in *D. pseudoobscura*. *kmg* appears to have morph-enriched transcription in spermatocyte cysts, which is also reflected in cyst morph-enriched protein abundance. Kmg was detected in all cysts, suggesting that it may be essential in spermatogenesis, as has been demonstrated for the *D. melanogaster* Kmg orthologue (Kim *et al.* 2017). Kmg in *D. melanogaster* blocks transcription of somatic genes by blocking tMAC binding to promoters. *D. pseudoobscura* Kmg appears to have similar DNA binding motifs, however target genes are currently unknown.

Hypotheses of Kmg function in *D. pseudoobscura* spermatogenesis can be considered in terms of potential function in eusperm or parasperm morphs:

1. Kmg is more abundant in eusperm cysts, and A) blocks transcription of somatic or parasperm-specific germline genes, and/or B) promotes transcription of eusperm-upregulated genes.
2. Kmg is more abundant in parasperm cysts, and A) blocks transcription of somatic or eusperm-specific germline genes, and/or B) promotes transcription of parasperm-upregulated genes.

7.4.3.1 Hypothesis 1A: Kmg is more abundant in eusperm, and blocks transcription.

In this hypothesis, Kmg is more abundant in eusperm spermatocyte cysts, and prevents transcription of genes not required or detrimental to eusperm development by blocking transcription factors binding to promoters. Higher Kmg results in greater reduction in expression of somatic or parasperm-upregulated genes, resulting in a more tightly controlled transcriptional profile. This is the most similar hypothesis of function to that of *D. melanogaster* Kmg.

There is some evidence that may support this hypothesis in RNA-seq data presented in this work, as more differentially expressed genes were found to be upregulated in short spermatid cysts, compared to elongated eusperm cysts, suggesting more transcriptional regulation in eusperm cysts. This is not conclusive as 'short' cysts may represent early elongating eusperm as well as parasperm cysts.

Based on this hypothesis, experimental predictions would be that a knockdown phenotype may reduce the number of eusperm cysts, compared to parasperm cysts. Over-expression may increase the proportion of eusperm cysts. Analysis of Kmg binding, for example by ChIP-seq, would be predicted to have an over-representation of somatic genes, or over-representation of promoters of parasperm-specific sequences.

7.4.3.2 Hypothesis 1B: Kmg is more abundant in eusperm, and promotes transcription.

Kmg may be more abundant in eusperm spermatocyte cysts, but rather than block transcription as in *D. melanogaster*, it functions in increasing transcription of eusperm-upregulated genes. In a *kmg* knockdown phenotype, fewer eusperm cysts may be observed. ChIP-seq would be predicted to show over-representation of germline expressed genes, and over-representation of eusperm-upregulated genes.

7.4.3.3 Hypothesis 2A: Kmg is more abundant in parasperm, and blocks transcription.

Kmg may be more abundant in parasperm spermatocyte cysts, compared to eusperm spermatocyte cysts. In this hypothesis, Kmg may be blocking transcription of somatic genes, or transcription of eusperm-upregulated genes. Where the effects of eusperm-upregulated genes are dose-dependent, Kmg may be required to maintain lower transcription of genes for which eusperm require greater transcript abundance. For example, eusperm likely require greater transcription of genes contributing to axoneme assembly and flagellum components.

Under this hypothesis, knockdown may increase the number of eusperm cysts. ChIP-seq would be predicted to show over-representation of somatic genes, or over-representation of eusperm-upregulated germline genes.

7.4.3.4 Hypothesis 2B: Kmg is more abundant in parasperm, and promotes transcription.

Kmg may be more abundant in parasperm cysts, and bind to parasperm-upregulated genes to promote transcription. Knockdown of *kmg* may decrease the proportion of parasperm cysts. ChIP-seq would be predicted to show over-representation of germline genes upregulated in parasperm.

7.4.3.5 *D. pseudoobscura* Kmg and the Multiple tMACs Hypothesis

D. melanogaster Kmg has been characterised as interacting with the chromatin re-modeller dMi2 to block Aly binding to cryptic promoters of somatic genes (Kim *et al.* 2017). Under hypothesis 1A and 2A, outlined above, *D. pseudoobscura* Kmg function is predicted to be similar to that of *D. melanogaster*, preventing somatic expression in by blocking Aly, a component of tMAC, binding to promoters. These hypotheses can also be considered in the context of the Multiple tMAC Hypothesis, also outlined above. *D. pseudoobscura* Kmg may function in the regulation of one or more of the

predicted *D. pseudoobscura* tMACs, resulting in downregulation of somatic genes or germline genes upregulated in other morphs.

Upregulation of Kmg in one or two cyst morphs may prevent over-expression of somatic genes from upregulated morph-specific tMAC(s). Alternatively, it may prevent tMAC-mediated transcription in one or two cyst morphs by a tMAC version with uniform spermatocyte presence (Figure 7.3). In both of these hypotheses, Kmg functions to maintain a morph-specific transcriptional profile, by preventing tMAC-regulated non-specific expression.

Testing of these hypotheses would require more work to identify and characterise the tMACs present in *D. pseudoobscura*. Co-immunoprecipitation of *D. pseudoobscura* tMAC components and comparisons between ChIP-seq data for Kmg and DNA-binding components of tMAC would provide evidence to support or refute these hypotheses. Co-immunoprecipitation of Kmg would also provide insight into its mechanism and function, by identification of other components with which it interacts. To establish which morph(s) express higher Kmg, I propose a method of using the *kmg* promotor to drive expression of a fluorescently tagged stable protein, such as a protamine, which could be detected in the elongated spermatids. The spermatid cyst morph(s) exhibiting higher fluorescence intensity would most likely be the morph(s) in which *kmg* expression is higher.

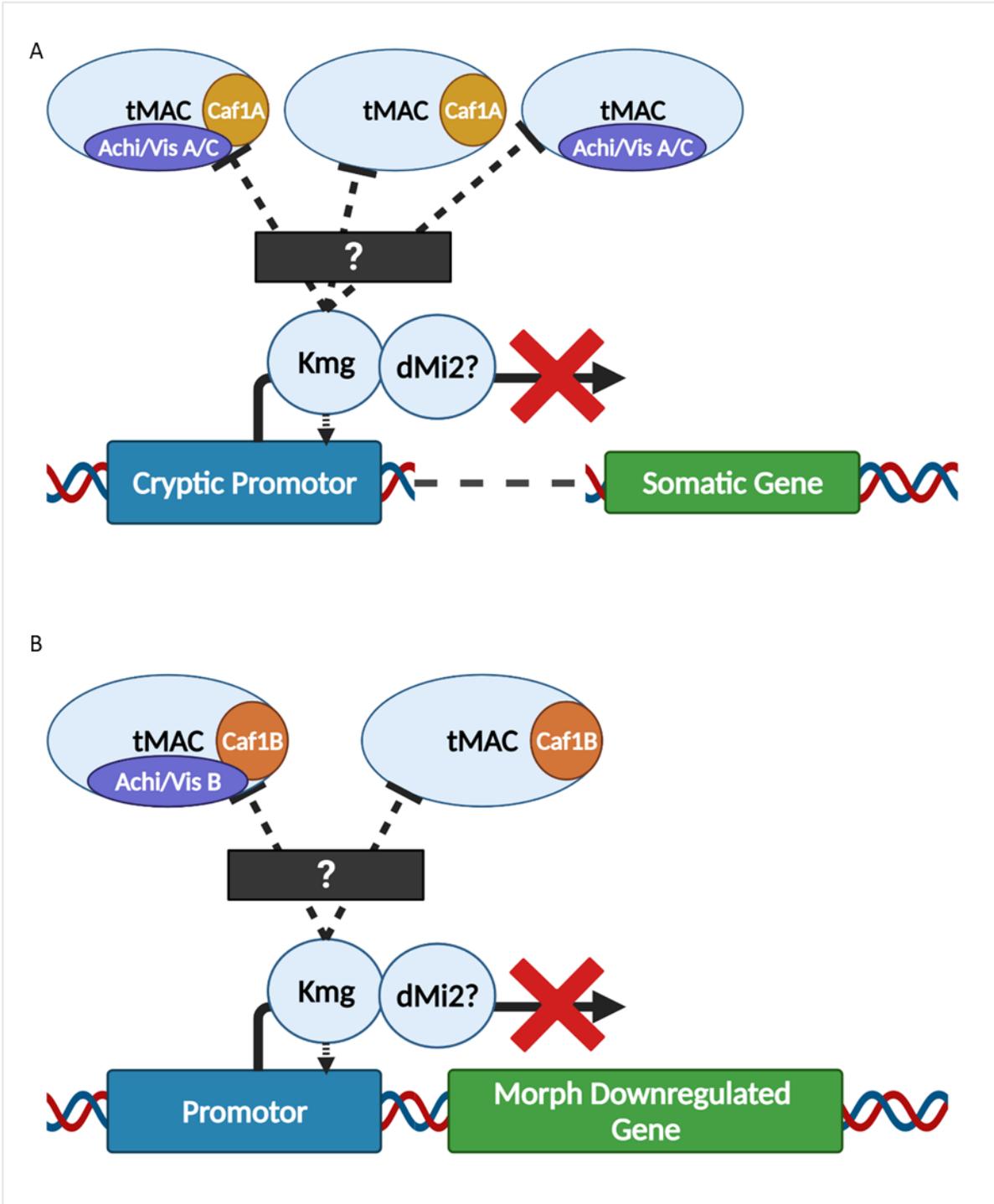


Figure 7.3: Models for potential interactions between Kmg and tMAC in *D. pseudoobscura*, under a 'Multiple tMAC' model. A: Kmg blocks over-expression in cysts with upregulated tMAC components. B: Kmg blocks expression of morph-downregulated genes.

7.5 Comparison of Hub Morphology between *D. pseudoobscura* and *D. melanogaster*

Previous work from the Klaus Lab has found that the hub region of *D. pseudoobscura* is larger in size than that of *D. melanogaster*, as shown by immunofluorescence staining of testes with the hub marker Fasciclin III (Beaury 2012; Mena 2012; Cardaci 2014). The work presented here has confirmed this finding, and presents new evidence of the size and structure of the hub and testis niche in *D. pseudoobscura*.

7.5.1 The ‘General Hubby Area’ model of the *D. pseudoobscura* testis niche

Hub markers FasIII, Neural Cadherin and Escargot supported the model of a large *D. pseudoobscura* hub area, considerably larger in size than that of *D. melanogaster*. Immunofluorescence staining for these markers in *D. melanogaster* showed a compact cluster of cells – the hub – with a clear differentiation between the peripheral boundary of the hub and the surrounding germline stem cells. In *D. pseudoobscura*, this was less clear, with FasIII and N-Cad immunofluorescence staining higher in intensity at the centre of the hub, and more diffuse at the periphery of the hub. Esg immunofluorescence staining showed a ring-like structure, possibly differentiated cells, cyst stem cells, or an unknown hub-structure. Staining for Hts, a fusome marker, showed the presence of unbranched spectrosomes, present in the GSCs, throughout the hub region.

The more diffuse adhesion molecule staining and GSC dispersion suggests that unlike the *D. melanogaster* hub, the *D. pseudoobscura* hub does not form a ‘cone’ shape. Instead, the hub appears to be more of a flat disc structure, with GSCs and CySCs dispersed across the ‘surface’ of the disc, as indicated by Hts. GSCs and CySCs differentiate away from the hub into the testis lumen. This updated model of the *D. pseudoobscura* hub is presented in Figure 7.4. This model of the more diffuse ‘flat disc’ hub, covered with GSCs interspersed with CySCs has been termed the ‘General Hubby Area’ model.

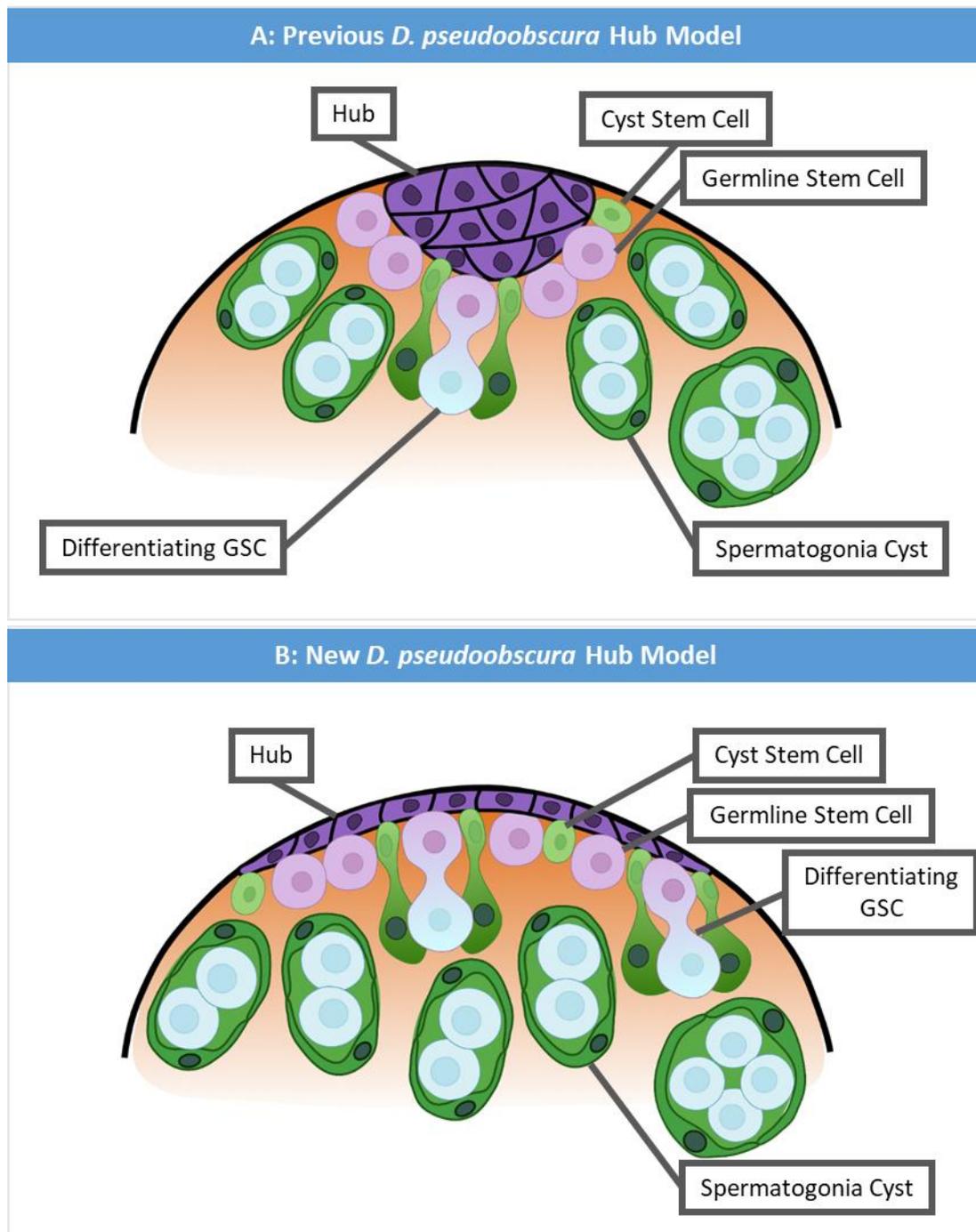


Figure 7.4: An updated model of the *D. pseudoobscura* testis hub structure. Hub = dark purple. GSCs = light purple. Spermatogonia = blue. CySCs = light green. Cyst cells = dark green. A: The previous model based on that of the *D. melanogaster* hub, in which the apical hub cells form a cone structure, to which germline and cyst stem cells adhere by nanotube projections and secretion of adhesion molecules (Tulina and Matunis 2001; Inaba et al. 2015). B: General Hubby Area model of the *D. pseudoobscura* hub. The hub is a flat disc of epithelial cells, which secretes signalling and adhesion molecules. The GSCs and CySCs are distributed across the surface of the hub. Differentiating cells are displaced away from the hub, into the testis lumen.

7.5.2 Sub-Hubs, multiple GSCs, and cyst cell contributions to heteromorphic sperm development?

It is unclear how early in development the processes leading to heteromorphic sperm development are initiated. Some have suggested that eusperm and parasperm morph differentiation begins during GSC or gonialblast differentiation (Moore *et al.* 2013).

Immunofluorescence for DE-Cadherin and Esg indicated that there may be sub-structures within the hub. DE-Cadherin is an adhesion molecule secreted by the hub cells and is essential for GSC maintenance (Yamashita *et al.* 2003; Wang *et al.* 2006). Localisation of DE-cadherin within the *D. pseudoobscura* general hubby area may indicate that there are specialised hub cells within the hub, secreting DE-cadherin. This may or may not be associated with multiple *D. melanogaster*-like hubs within the general hubby area. DE-cadherin indicated the presence of two or three structures within the hub.

The current evidence does not extend to supporting a model of multiple stem cell populations within the general hubby area, each contributing to a single morph, and maintained by a single sub-hub. A more conclusive test of this hypothesis would be a clonal analysis, where a FLP/FRT marker is induced at low frequencies so that testes contain a single clonal GSC, then determining whether the marker is present in one or more spermatocyte morphs (Chou and Perrimon 1992; Xu and Rubin 1993; Chou and Perrimon 1996; Decotto and Spradling 2005; Matunis *et al.* 2012; Davies *et al.* 2013). Additionally, further work to establish the signalling pathways involved in GSC maintenance and differentiation should expand the current understanding of the early stages of heteromorphic sperm development.

7.6 Summary: How does *D. pseudoobscura* control production of multiple sperm morphs?

In summary, this work has shown that transcriptional variation is present between subsets of spermatocyte cysts in *D. pseudoobscura*. Transcriptional variation in spermatocyte cysts is predicted to contribute to the development of multiple sperm morphs, by transcribing products required for morph-specific and morph-upregulated processes during spermiogenesis. This work presents the first evidence of multiple primary spermatocyte cyst morphs, demonstrating morph-specific differentiation prior to meiosis and the spermatid cyst stage, during which cyst length indicates morph. A short-list of candidate genes which show strong evidence of differential gene expression during spermatogenesis was identified. These genes may contribute to the development of heteromorphic sperm.

Similar to *D. melanogaster*, the testis meiotic arrest complex appears to have a function in *D. pseudoobscura* spermatocyte transcription. Some tMAC components appear to be differentially expressed in the spermatocytes and are strong candidates for a function in heteromorphic sperm development. Furthermore, multiple tMACs may be present in *D. pseudoobscura*, and may contribute to morph-specific transcription.

kmg was found to be differentially expressed between the spermatocyte cyst morphs. Kmg protein was also differentially expressed, indicating a function in heteromorphic sperm development. Kmg was detected in all spermatocyte cysts, indicating it may have an essential function in spermatogenesis, similar to that of *D. melanogaster*. Kmg may block transcription of genes to maintain morph-specific transcription in *D. pseudoobscura* spermatocyte cysts.

7.7 Future Directions

The development of more genetic tools in *D. pseudoobscura* is paramount for gaining greater understanding of the genetic and molecular processes underlying heteromorphic spermatogenesis. The piggyBac transposon vector was successfully used for genomic insertion of a GFP fusion. The same system could be adapted for development of UAS-Gal4, RNAi and site-specific recombination lines, as well as fluorescently tagged protein marker lines. Further development of CRISPR/Cas9 gene editing in *D. pseudoobscura* would greatly benefit research in this species, particularly in generating mutants and endogenous fluorescent tags.

There are multiple potential directions for future research. Further examination of the testis niche and early germline proliferation would benefit understanding of the earliest stages of morph differentiation. Expansion of the number of fluorescence-tagged candidate genes, in particular the tMAC components, would be beneficial in understanding the interactions between multiple genes and their products in heteromorphic sperm development.

Identification of *D. pseudoobscura kmg* as differentially expressed in spermatocyte cysts has given insight into some of the mechanisms which may contribute to transcriptional control in heteromorphic sperm development. To further understand the function of Kmg in *D. pseudoobscura*, identification of the target genes by ChIP-seq and of co-localising proteins by co-immunoprecipitation would be valuable. Generation of *kmg* knockdown or mutants would also be of use in identification of upregulation and/or downregulation of target genes, and the contribution of upregulated *kmg* expression to morph-specific development. Finally, it may also be interesting to compare *kmg* expression between the *obscura* group species, to establish whether *kmg* contributes to morph-specific development across the sperm heteromorphic *obscura* group.

Bibliography

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

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J. *et al.* (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**:W3-W10.

Alkhatib, S. G. and Landry, J. W. (2011). The Nucleosome Remodeling Factor. *FEBS Lett* **585**:3197-3207.

Alpern, J. H. M., Asselin, M. M. and Moehring, A. J. (2019). Identification of a novel sperm class and its role in fertilization in *Drosophila*. *J Evol Biol* **32**:259-266.

Alphey, L., Jimenez, J., White-Cooper, H., Dawson, I., Nurse, P. and Glover, D. M. (1992). *twine*, a *cdc25* homolog that functions in the male and female germline of *Drosophila*. *Cell* **69**:977-988.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-3402.

Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schaffer, A. A. and Yu, Y. K. (2005). Protein database searches using compositionally adjusted substitution matrices. *FEBS J* **272**:5101-5109.

Anders, S., Pyl, P. T. and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166-169.

Andrews, J., Bouffard, G. G., Cheadle, C., Lü, J., Becker, K. G. and Oliver, B. (2000). Gene Discovery Using Computational and Microarray Analysis of Transcription in the *Drosophila melanogaster* Testis. *Genome Research* **10**:2030-2043.

Andrews, S. (2010). FastQC – A Quality Control tool for High Throughput Sequence Data. [Online]. Babraham Bioinformatics. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [Accessed: November 20].

Arbouzova, N. I. and Zeidler, M. P. (2006). JAK/STAT signalling in *Drosophila*: insights into conserved regulatory and cellular functions. *Development* **133**:2605-2616.

Assis, R. and Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in *Drosophila*. *Proc Natl Acad Sci USA* **110**:17409-17414.

Attisano, L. and Wrana, J. L. (2000). Smads as transcriptional co-modulators. *Current Opinion in Cell Biology* **12**:235-243.

Ayyar, S., Jiang, J., Collu, A., White-Cooper, H. and White, R. A. (2003). *Drosophila* TGIF is essential for developmentally regulated transcription in spermatogenesis. *Development* **130**:2841-2852.

Bai, Y., Casola, C., Feschotte, C. and Betran, E. (2007). Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biol* **8**:R11.

Baker, H. S. (1935). Influence of Temperature on Testis Size in *Drosophila pseudoobscura*. *The American Naturalist* **69**:412-416.

Baker, R. R. and Bellis, M. A. (1988). 'Kamikaze' sperm in mammals? *Animal Behaviour* **36**:936-939.

Baker, R. R. and Bellis, M. A. (1989). Elaboration of the Kamikaze Sperm Hypothesis: a reply to Harcourt. *Animal Behaviour* **37**:865-867.

Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F. and White-Cooper, H. (2008a). Post-meiotic transcription in *Drosophila* testes. *Development* **135**:1897-1902.

Barreau, C., Benson, E. and White-Cooper, H. (2008b). Comet and cup genes in *Drosophila* spermatogenesis: the first demonstration of post-meiotic transcription. *Biochem Soc Trans* **36**:540-542.

Bassett, A. R. and Liu, J. L. (2014). CRISPR/Cas9 and genome editing in *Drosophila*. *J Genet Genomics* **41**:7-19.

Bassett, A. R., Tibbit, C., Ponting, C. P. and Liu, J. L. (2013). Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep* **4**:220-228.

Baudier, K. M., Kaschock-Marenda, S. D., Patel, N., Diangelus, K. L., O'Donnell, S. and Marenda, D. R. (2014). Erythritol, a non-nutritive sugar alcohol sweetener and the main component of truvia(R), is a palatable ingested insecticide. *PLoS One* **9**:e98949.

Beall, E. L., Lewis, P. W., Bell, M., Rocha, M., Jones, D. L. and Botchan, M. R. (2007). Discovery of tMAC: a *Drosophila* testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev* **21**:904-919.

Beatty, R. A. and Sidhu, N. S. (1969). Polymegaly of spermatozoan length and its genetic control in *Drosophila* species. *Proceedings of the Royal Society of Edinburgh Section B: Biological Sciences* **71B**:14-29.

Beaury, M. W. (2012). Confocal Characterization of the Apical Testes and Ultrastructural Analysis of the Seminal Visicles of *Drosophila pseudoobscura*. Thesis Master of Science (MS), Seton Hall University.

Bernasconi, G. and Hellriegel, B. (2005). Fertilization competence and sperm size variation in sperm-heteromorphic insects. *Evolutionary Ecology* **19**:45-54.

Bier, E., Harrison, M. M., O'Connor-Giles, K. M. and Wildonger, J. (2018). Advances in Engineering the Fly Genome with the CRISPR-Cas System. *Genetics* **208**:1-18.

Bircher, U. and Hauschteck-Jungen, E. (1997). The length of the sperm nucleus in *Drosophila obscura* group species is depending on the total length of the sperm. *Invertebrate Reproduction & Development* **32**:225-229.

Bircher, U., Jungen, H., Burch, R. and Hauschteck-Jungen, E. (1995). Multiple morphs of sperm were required for the evolution of the Sex Ratio trait in *Drosophila*. *Journal of Evolutionary Biology* **8**:575-588.

Blachon, S., Gopalakrishnan, J., Omori, Y., Polyanovsky, A., Church, A., Nicastro, D., Malicki, J. *et al.* (2008). *Drosophila asterless* and vertebrate *Cep152* are orthologs essential for centriole duplication. *Genetics* **180**:2081-2094.

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G. *et al.* (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49**:D344-D354.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114-2120.

Brawley, C. and Matunis, E. (2004). Regeneration of male germline stem cells by spermatogonial dedifferentiation in vivo. *Science* **304**:1331-1334.

Bressac, C. and Hauschteck-Jungen, E. (1996). *Drosophila subobscura* females preferentially select long sperm for storage and use. *Journal of Insect Physiology* **42**:323-328.

Bressac, C., Joly, D., Devaux, J., Serres, C., Feneux, D. and Lachaise, D. (1991). Comparative kinetics of short and long sperm in sperm dimorphic *Drosophila* species. *Cell Motil Cytoskeleton* **19**:269-274.

Bukhari, H. and Muller, T. (2019). Endogenous Fluorescence Tagging by CRISPR. *Trends Cell Biol* **29**:912-928.

Bulger, M., Ito, T., Kamakaka, R. T. and Kadonaga, J. T. (1995). Assembly of regularly spaced nucleosome arrays by *Drosophila* chromatin assembly factor 1 and a 56-kDa histone-binding protein. *Proc Natl Acad Sci USA* **92**:11726-11730.

Bunt, S. M. and Hime, G. R. (2004). Ectopic activation of Dpp signalling in the male *Drosophila* germline inhibits germ cell differentiation. *Genesis* **39**:84-93.

Calvo-Martin, J. M., Papaceit, M. and Segarra, C. (2017). Evidence of neofunctionalization after the duplication of the highly conserved Polycomb group gene *Caf1-55* in the *obscura* group of *Drosophila*. *Sci Rep* **7**:40536.

Cardaci, P. J. (2014). Apical Testis Structure and the Effects of Cadmium Treatment on Spermatogenesis in *Drosophila*. Thesis Master of Sciences, Seton Hall University.

Celniker, S. E. and Rubin, G. M. (2003). The *Drosophila melanogaster* genome. *Annu Rev Genomics Hum Genet* **4**:89-117.

Chapman, T. (2008). Evolutionary biology: sterile saviours. *Curr Biol* **18**:R261-263.

Chen, B. S. and Hampsey, M. (2002). Transcription activation: unveiling the essential nature of TFIID. *Curr Biol* **12**:R620-622.

Chen, D. and McKearin, D. (2003). Dpp signaling silences bam transcription directly to establish asymmetric divisions of germline stem cells. *Curr Biol* **13**:1786-1791.

Chen, X., Hiller, M., Sancak, Y. and Fuller, M. T. (2005). Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. *Science* **310**:869-872.

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

Chou, T. B. and Perrimon, N. (1992). Use of a yeast site-specific recombinase to produce female germline chimeras in *Drosophila*. *Genetics* **131**:643-653.

Chou, T. B. and Perrimon, N. (1996). The autosomal FLP-DFS technique for generating germline mosaics in *Drosophila melanogaster*. *Genetics* **144**:1673-1679.

Chung, H. R., Schafer, U., Jackle, H. and Bohm, S. (2002). Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in *Drosophila*. *EMBO Rep* **3**:1158-1162.

Church, K. and Lin, H. P. (1985). Kinetochore microtubules and chromosome movement during prometaphase in *Drosophila melanogaster* spermatocytes studied in life and with the electron microscope. *Chromosoma* **92**:273-282.

Civetta, A. (1999). Direct visualization of sperm competition and sperm storage in *Drosophila*. *Curr Biol* **9**:841-844.

Clapier, C. R. and Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annu Rev Biochem* **78**:273-304.

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C. *et al.* (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**:203-218.

Colozza, G., Montembault, E., Quenerch'du, E., Riparbelli, M. G., D'Avino, P. P. and Callaini, G. (2011). *Drosophila* nucleoporin Nup154 controls cell viability, proliferation and nuclear accumulation of Mad transcription factor. *Tissue Cell* **43**:254-261.

Concordet, J. P. and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**:W242-W245.

Cook, P. A. and Wedell, N. (1999). Non-fertile sperm delay female remating. *Nature* **397**.

Courtot, C., Fankhauser, C., Simanis, V. and Lehner, C. F. (1992). The *Drosophila cdc25* homolog *twine* is required for meiosis. *Development* **116**:405-416.

Cross, D. P. and Shellenbarger, D. L. (1979). The dynamics of *Drosophila melanogaster* spermatogenesis in in vitro cultures. *J Embryol Exp Morphol* **53**:345-351.

Crowley, T. E. and Hazelrigg, T. (1995). A male-specific 3'-UTR regulates the steady-state level of the *exuperantia* mRNA during spermatogenesis in *Drosophila*. *Mol Gen Genet* **248**:370-374.

Crudginton, H. S., Fellows, S., Badcock, N. S. and Snook, R. R. (2009). Experimental manipulation of sexual selection promotes greater male mating capacity but does not alter sperm investment. *Evolution* **63**:926-938.

Crysnanto, D. and Obbard, D. J. (2019). Widespread gene duplication and adaptive evolution in the RNA interference pathways of the *Drosophila obscura* group. *BMC Evol Biol* **19**:99.

Davies, E. L., Lim, J. G., Joo, W. J., Tam, C. H. and Fuller, M. T. (2013). The transcriptional regulator *lola* is required for stem cell maintenance and germ cell differentiation in the *Drosophila* testis. *Dev Biol* **373**:310-321.

de Cuevas, M. and Matunis, E. L. (2011). The stem cell niche: lessons from the *Drosophila* testis. *Development* **138**:2861-2869.

de Souza, H. V. and Itoyama, M. M. (2010). Comparative Study of Spermatogenesis and Nucleolar Behavior in Testicular Lobes of *Euschistus heros* (Heteroptera: Pentatomidae). *Psyche: A Journal of Entomology* **2010**:1-10.

Decotto, E. and Spradling, A. C. (2005). The *Drosophila* ovarian and testis stem cell niches: similar somatic stem cells and signals. *Dev Cell* **9**:501-510.

Demarco, R. S., Eikenes, A. H., Haglund, K. and Jones, D. L. (2014). Investigating spermatogenesis in *Drosophila melanogaster*. *Methods* **68**:218-227.

Dhole, S. and Pfennig, K. S. (2014). Age-dependent male mating investment in *Drosophila pseudoobscura*. *PLoS One* **9**:e88700.

Dobzhansky, T. (1934). Studies on hybrid sterility. *Z Zellforsch Mikrosk Anat* **21**:169-223.

Doggett, K., Jiang, J., Aleti, G. and White-Cooper, H. (2011). Wake-up-call, a *lin-52* paralogue, and Always early, a *lin-9* homologue physically interact, but have opposing functions in regulating testis-specific gene expression. *Dev Biol* **355**:381-393.

Dorus, S., Busby, S. A., Gerike, U., Shabanowitz, J., Hunt, D. F. and Karr, T. L. (2006). Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat Genet* **38**:1440-1445.

dos Santos, G., Schroeder, A. J., Goodman, J. L., Strelets, V. B., Crosby, M. A., Thurmond, J., Emmert, D. B. *et al.* (2015). FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res* **43**:D690-697.

Doyen, C. M., Chalkley, G. E., Voets, O., Bezstarosti, K., Demmers, J. A., Moshkin, Y. M. and Verrijzer, C. P. (2015). A Testis-Specific Chaperone and the Chromatin Remodeler ISWI Mediate Repackaging of the Paternal Genome. *Cell Rep* **13**:1310-1318.

Eberhart, C. G., Maines, J. Z. and Wasserman, S. A. (1996). Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. *Nature* **381**:783-785.

Epstein, Y., Perry, N., Volin, M., Zohar-Fux, M., Braun, R., Porat-Kuperstein, L. and Toledano, H. (2017). *miR-9a* modulates maintenance and ageing of *Drosophila* germline stem cells by limiting N-cadherin expression. *Nat Commun* **8**:600.

Evans Lab. (2014). Single Fly Genomic Prep [Online]. <http://www.evansflylab.com/files>: Available at: [http://www.evansflylab.com/files/Protocol File/Single fly genomic DNA prep.pdf](http://www.evansflylab.com/files/Protocol%20File/Single%20fly%20genomic%20DNA%20prep.pdf) [Accessed.

Fabian, L. and Brill, J. A. (2012). *Drosophila* spermiogenesis: Big things come from little packages. *Spermatogenesis* **2**:197-212.

Fairchild, M. J., Smendziuk, C. M. and Tanentzapf, G. (2015). A somatic permeability barrier around the germline is essential for *Drosophila* spermatogenesis. *Development* **142**:268-281.

Fatima, R. (2011). *Drosophila* Dynein intermediate chain gene, *Dic61B*, is required for spermatogenesis. *PLoS One* **6**:e27822.

Foe, V. E., Odell, G. M. and Edgar, B. A. (1993). Mitosis and Morphogenesis in the *Drosophila* Embryo. In: Bate, M. and Martinez Arias, A. (eds.) *The Development of Drosophila melanogaster*. Vol. 1. New York, USA: Cold Spring Harbor Laboratory Press, pp. 149-300.

Forrest, K. M., Clark, I. E., Jain, R. A. and Gavis, E. R. (2004). Temporal complexity within a translational control element in the *nanos* mRNA. *Development* **131**:5849-5857.

Forrest, K. M. and Gavis, E. R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for *nanos* mRNA localization in *Drosophila*. *Curr Biol* **13**:1159-1168.

Friedländer, M. (1997). Control of the eupyrene-apyrene sperm dimorphism in Lepidoptera. *J Insect Physiol* **43**:1085-1092.

Friedländer, M. and Gitay, H. (1972). The fate of the normal-enucleated spermatozoa in inseminated females of the silkworm *Bombyx mori*. *J Morphol* **138**:121-129.

Friedländer, M., Seth, R. K. and Reynolds, S. E. (2005). Eupyrene and Apyrene Sperm: Dichotomous Spermatogenesis in Lepidoptera. In: Simpson, S.J. (ed.) *Advances in Insect Physiology Volume 32*. Vol. 32. Academic Press, pp. 206-308.

- Frolova, S. and Astaurov, B. (1929). *Drosophila pseudoobscura*. *Z Zellforsch Mikrosk Anat* **10**.
- Fuller, M. (1993). Spermatogenesis. In: Bate, M. and Martinez Arias, A. (eds.) *Development of Drosophila*. Vol. 1. Cold Spring Harbor Laboratory Press, pp. 71-147.
- Fuller, M. T. (1998). Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis. *Semin Cell Dev Biol* **9**:433-444.
- Galletta, B. J., Jacobs, K. C., Fagerstrom, C. J. and Rusan, N. M. (2016). Asterless is required for centriole length control and sperm development. *J Cell Biol* **213**:435-450.
- Gigliotti, S., Callaini, G., Andone, S., Riparbelli, M. G., Pernas-Alonso, R., Hoffmann, G., Graziani, F. *et al.* (1998). *Nup154*, a new *Drosophila* gene essential for male and female gametogenesis is related to the *Nup155* vertebrate nucleoporin gene. *J Cell Biol* **142**:1195-1207.
- Gonczy, P., Matunis, E. and DiNardo, S. (1997). *bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development* **124**:4361-4371.
- Gould-Somero, M. and Holland, L. (1974). The timing of RNA synthesis for spermiogenesis in organ cultures of *Drosophila melanogaster* testes. *Wilhelm Roux Arch Entwickl Mech Org* **174**:133-148.
- Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., Rey, A. J. *et al.* (2017). FlyBase at 25: looking to the future. *Nucleic Acids Res* **45**:D663-D671.
- Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. *et al.* (2013). Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**:1029-1035.

Gratz, S. J., Rubinstein, C. D., Harrison, M. M., Wildonger, J. and O'Connor-Giles, K. M. (2015). CRISPR-Cas9 Genome Editing in *Drosophila*. *Curr Protoc Mol Biol* **111**:31 32 31-31 32 20.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. and O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* **196**:961-971.

Greeff, J. M. and Parker, G. A. (2000). Spermicide by females: what should males do? *Proc Biol Sci* **267**:1759-1763.

Gromko, M. H., Gilbert, D. G. and Richmond, R. C. (1984). Sperm Transfer and Use in the Multiple Mating System of *Drosophila*. In: Smith, R.L. (ed.) *Sperm Competition and the Evolution of Animal Mating systems*. 1 ed. London: Academic Press.

Hamaratoglu, F., Affolter, M. and Pyrowolakis, G. (2014). Dpp/BMP signaling in flies: from molecules to biology. *Semin Cell Dev Biol* **32**:128-136.

Handler, A. M., McCombs, S. D., Fraser, M. J. and Saul, S. H. (1998). The lepidopteran transposon vector, *piggyBac*, mediates germ-line transformation in the Mediterranean fruit fly. *Proc Natl Acad Sci U S A* **95**:7520-7525.

Harcourt, A. H. (1989). Deformed sperm are probably not adaptive. *Animal Behaviour* **37**:863-865.

Harcourt, A. H. (1991). Sperm Competition and the Evolution of Nonfertilizing Sperm in Mammals. *Evolution* **45**:314-328.

Hardy, R. W., Tokuyasu, K. T., Lindsley, D. L. and Garavito, M. (1979). The germinal proliferation center in the testis of *Drosophila melanogaster*. *J Ultrastruct Res* **69**:180-190.

Harris, R. E. and Ashe, H. L. (2011). Cease and desist: modulating short-range Dpp signalling in the stem-cell niche. *EMBO Rep* **12**:519-526.

Hauschteck-Jungen, E. and Maurer, B. (1976). Sperm dysfunction in sex ratio males of *Drosophila subobscura*. *Genetica* **46**:459-477.

Hayakawa, Y. (2007). Parasperm: morphological and functional studies on nonfertile sperm. *Ichthyological Research* **54**:111-130.

Hayashi, S., Hirose, S., Metcalfe, T. and Shirras, A. D. (1993). Control of imaginal cell development by the *escargot* gene of *Drosophila*. *Development* **118**:105-115.

Haynes, S. R., Cooper, M. T., Pype, S. and Stolow, D. T. (1997). Involvement of a tissue-specific RNA recognition motif protein in *Drosophila* spermatogenesis. *Mol Cell Biol* **17**:2708-2715.

Herrera, S. C. and Bach, E. A. (2018). JNK signaling triggers spermatogonial dedifferentiation during chronic stress to maintain the germline stem cell pool in the *Drosophila* testis. *eLife* **7**.

Herrera, S. C. and Bach, E. A. (2019). JAK/STAT signaling in stem cells and regeneration: from *Drosophila* to vertebrates. *Development* **146**.

Hiller, M., Chen, X., Pringle, M. J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B. *et al.* (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* **131**:5297-5308.

Hiller, M. A., Lin, T. Y., Wood, C. and Fuller, M. T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev* **15**:1021-1030.

Holman, L., Freckleton, R. P. and Snook, R. R. (2008). What use is an infertile sperm? A comparative study of sperm-heteromorphic *Drosophila*. *Evolution* **62**:374-385.

Holman, L. and Snook, R. R. (2006). Spermicide, cryptic female choice and the evolution of sperm form and function. *J Evol Biol* **19**:1660-1670.

Holman, L. and Snook, R. R. (2008). A sterile sperm caste protects brother fertile sperm from female-mediated death in *Drosophila pseudoobscura*. *Curr Biol* **18**:292-296.

Holtzman, S., Miller, D., Eisman, R., Kuwayama, H., Niimi, T. and Kaufman, T. (2010). Transgenic tools for members of the genus *Drosophila* with sequenced genomes. *Fly (Austin)* **4**:349-362.

Hombria, J. C. and Brown, S. (2002). The fertile field of *Drosophila* Jak/STAT signalling. *Curr Biol* **12**:R569-575.

Horn, C. and Handler, A. M. (2005). Site-specific genomic targeting in *Drosophila*. *Proc Natl Acad Sci USA* **102**:12483-12488.

Horn, C. and Wimmer, E. A. (2000). A versatile vector set for animal transgenesis. *Dev Genes Evol* **210**:630-637.

Housden, B. E., Lin, S. and Perrimon, N. (2014). Cas9-based genome editing in *Drosophila*. *Methods Enzymol* **546**:415-439.

Housden, B. E. and Perrimon, N. (2016). Detection of Indel Mutations in *Drosophila* by High-Resolution Melt Analysis (HRMA). *Cold Spring Harb Protoc* **2016**:pdb. prot090795.

Inaba, M., Buszczak, M. and Yamashita, Y. M. (2015). Nanotubes mediate niche-stem-cell signalling in the *Drosophila* testis. *Nature* **523**:329-332.

Insko, M. L., Leon, A., Tam, C. H., McKearin, D. M. and Fuller, M. T. (2009). Accumulation of a differentiation regulator specifies transit amplifying division number in an adult stem cell lineage. *Proc Natl Acad Sci USA* **106**:22311-22316.

Jandura, A., Hu, J., Wilk, R. and Krause, H. M. (2017). High Resolution Fluorescent In Situ Hybridization in *Drosophila* Embryos and Tissues Using Tyramide Signal Amplification. *J Vis Exp*.

Jayaramaiah Raja, S. and Renkawitz-Pohl, R. (2005). Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus. *Mol Cell Biol* **25**:6165-6177.

Jiang, J., Benson, E., Bausek, N., Doggett, K. and White-Cooper, H. (2007). Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the *Drosophila* male germline and physically interacts with always early. *Development* **134**:1549-1559.

Jiang, J. and White-Cooper, H. (2003). Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of *aly* and a novel meiotic arrest gene *cookie monster*. *Development* **130**:563-573.

Joly, D., Cariou, M. and Lachaise, D. (1991). Can sperm competition explain sperm polymorphism in *Drosophila teissieri*? *Evol. Biol.* **5**:25-44.

Joly, D., Cariou, M. L., Lachaise, D. and David, J. R. (1989). Variation of sperm length and heteromorphism in drosophilid species. *Genet Sel Evol* **21**:283-293.

Joly, D. and Lachaise, D. (1994). Polymorphism in the sperm heteromorphic species of the *Drosophila obscura* group. *J Insect Physiol* **40**:933-938.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res* **20**:1313-1326.

Kaessmann, H., Vinckenbosch, N. and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet* **10**:19-31.

Kahney, E. W., Snedeker, J. C. and Chen, X. (2019). Regulation of *Drosophila* germline stem cells. *Curr Opin Cell Biol* **60**:27-35.

Kalajdzic, P. and Schetelig, M. F. (2017). CRISPR/Cas-mediated gene editing using purified protein in *Drosophila suzukii*. *Entomologia Experimentalis et Applicata* **164**:350-362.

Karr, T. L. (1991). Intracellular sperm/egg interactions in *Drosophila*: a three-dimensional structural analysis of a paternal product in the developing egg. *Mech Dev* **34**:101-111.

Kawase, E., Wong, M. D., Ding, B. C. and Xie, T. (2004). Gbb/Bmp signaling is essential for maintaining germline stem cells and for repressing *bam* transcription in the *Drosophila* testis. *Development* **131**:1365-1375.

Ketel, C. S., Andersen, E. F., Vargas, M. L., Suh, J., Strome, S. and Simon, J. A. (2005). Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. *Mol Cell Biol* **25**:6857-6868.

Kharchenko, P. V. (2021). The triumphs and limitations of computational methods for scRNA-seq. *Nat Methods* **18**:723-732.

Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. and Fuller, M. T. (2001). Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294**:2542-2545.

Kiger, A. A., White-Cooper, H. and Fuller, M. T. (2000). Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**:750-754.

Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A., Comeault, A. A. *et al.* (2021). Highly contiguous assemblies of 101 drosophilid genomes. *eLife* **10**:e66405.

Kim, D., Langmead, B. and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**:357-360.

Kim, J., Lu, C., Srinivasan, S., Awe, S., Brehm, A. and Fuller, M. T. (2017). Blocking promiscuous activation at cryptic promoters directs cell type-specific gene expression. *Science* **356**:717-721.

Kimura, S. (2013). The Nap family proteins, CG5017/Hanabi and Nap1, are essential for *Drosophila* spermiogenesis. *FEBS Lett* **587**:922-929.

Kleene, K. C. (2005). Sexual selection, genetic conflict, selfish genes, and the atypical patterns of gene expression in spermatogenic cells. *Dev Biol* **277**:16-26.

Klowden, M. J. and Chambers, G. M. (2004). Production of polymorphic sperm by anopheline mosquitoes and their fate within the female genital tract. *J Insect Physiol* **50**:1163-1170.

Kondo, S. and Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics* **195**:715-721.

Kondo, S., Vedanayagam, J., Mohammed, J., Eizadshenass, S., Kan, L., Pang, N., Aradhya, R. *et al.* (2017). New genes often acquire male-specific functions but rarely become essential in *Drosophila*. *Genes Dev* **31**:1841-1846.

Korzelius, J., Naumann, S. K., Loza-Coll, M. A., Chan, J. S., Dutta, D., Oberheim, J., Glasser, C. *et al.* (2014). Escargot maintains stemness and suppresses differentiation in *Drosophila* intestinal stem cells. *EMBO J* **33**:2967-2982.

Ku, H. Y., Gangaraju, V. K., Qi, H., Liu, N. and Lin, H. (2016). Tudor-SN Interacts with Piwi Antagonistically in Regulating Spermatogenesis but Synergistically in Silencing Transposons in *Drosophila*. *PLoS Genet* **12**:e1005813.

Kudo, A., Awasaki, T., Ishikawa, Y. and Matsuo, T. (2018). *piggyBac*- and phiC31 integrase-mediated transgenesis in *Drosophila prolongata*. *Genes Genet Syst* **92**:277-285.

Kuhn, R., Schafer, U. and Schafer, M. (1988). Cis-acting regions sufficient for spermatocyte-specific transcriptional and spermatid-specific translational control of the *Drosophila melanogaster* gene *mst(3)gl-9*. *EMBO J* **7**:447-454.

Kurokawa, H. and Hihara, F. (1976). Number of first spermatocytes in relation to phylogeny of *Drosophila* (Diptera: Drosophilidae). *Int. J. Insect. Morphol. & Embryol.* **5**:51-63.

Labbe, G. M., Nimmo, D. D. and Alphey, L. (2010). *piggybac*- and PhiC31-mediated genetic transformation of the Asian tiger mosquito, *Aedes albopictus* (Skuse). *PLoS Negl Trop Dis* **4**:e788.

Laktionov, P. P., Maksimov, D. A., Romanov, S. E., Antoshina, P. A., Posukh, O. V., White-Cooper, H., Koryakov, D. E. *et al.* (2018). Genome-wide analysis of gene regulation mechanisms during *Drosophila* spermatogenesis. *Epigenetics Chromatin* **11**:14.

Laktionov, P. P., White-Cooper, H., Maksimov, D. A. and Belyakin, S. N. (2014). Transcription factor Comr acts as a direct activator in the genetic program controlling spermatogenesis in *D. melanogaster*. *Molecular Biology* **48**:130-140.

Lambkin, C. L., Sinclair, B. J., Pape, T., Courtney, G. W., Skevington, J. H., Meier, R., Yeates, D. K. *et al.* (2013). The phylogenetic relationships among infraorders and superfamilies of Diptera based on morphological evidence. *Systematic Entomology* **38**:164-179.

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., Dos Santos, G., Garapati, P. V., Goodman, J. L. *et al.* (2021). FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res* **49**:D899-D907.

Leatherman, J. L. and Dinardo, S. (2008). Zfh-1 controls somatic stem cell self-renewal in the *Drosophila* testis and nonautonomously influences germline stem cell self-renewal. *Cell Stem Cell* **3**:44-54.

Leatherman, J. L. and Dinardo, S. (2010). Germline self-renewal requires cyst stem cells and *stat* regulates niche adhesion in *Drosophila* testes. *Nat Cell Biol* **12**:806-811.

Lee, J. S., Kwak, S. J., Kim, J., Kim, A. K., Noh, H. M., Kim, J. S. and Yu, K. (2014). RNA-guided genome editing in *Drosophila* with the purified Cas9 protein. *G3 (Bethesda)* **4**:1291-1295.

Lee, P. E. and Wilkes, A. (1965). Polymorphic Spermatozoa in the Hymenopterous Wasp *Dahlbominus*. *Science* **147**:1445-1446.

Lewis, S. H., Webster, C. L., Salmela, H. and Obbard, D. J. (2016). Repeated Duplication of *Argonaute2* is Associated with Strong Selection and Testis Specialization in *Drosophila*. *Genetics* **204**:757-769.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. *et al.* (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**:2078-2079.

Li, H., Janssens, J., De Waegeneer, M., Kolluru, S. S., Davie, K., Gardeux, V., Saelens, W. *et al.* (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**:eabk2432.

Lim, C., Gandhi, S., Biniossek, M. L., Feng, L., Schilling, O., Urban, S. and Chen, X. (2015). An Aminopeptidase in the *Drosophila* Testicular Niche Acts in Germline Stem Cell Maintenance and Spermatogonial Dedifferentiation. *Cell Rep* **13**:315-325.

Lim, C., Tarayrah, L. and Chen, X. (2012). Transcriptional regulation during *Drosophila* spermatogenesis. *Spermatogenesis* **2**:158-166.

Lin, H. (2002). The stem-cell niche theory: lessons from flies. *Nat Rev Genet* **3**:931-940.

Lin, T. Y., Pringle, M. J. and Fuller, M. T. (2000). Regulation of meiosis and spermatid differentiation in *Drosophila* primary spermatocytes. In: Goldberg, E. (ed.) *The Testis*. New York: Springer, pp. 120-132.

Lin, T. Y., Viswanathan, S., Wood, C., Wilson, P. G., Wolf, N. and Fuller, M. T. (1996). Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development* **122**:1331-1341.

Lindsley, D. L. and Tokuyasu, K. T. (1980). Spermatogenesis. In: Ashburner, M. and F., W.T.R. (eds.) *The Genetics and Biology of Drosophila*. Vol. 2. London: Academic Press, pp. 225-294.

Liu, Y., Bao, H., Wang, W. and Lim, H. Y. (2019). Cardiac Snail family of transcription factors directs systemic lipid metabolism in *Drosophila*. *PLoS Genet* **15**:e1008487.

Loza-Coll, M. A., Southall, T. D., Sandall, S. L., Brand, A. H. and Jones, D. L. (2014). Regulation of *Drosophila* intestinal stem cell maintenance and differentiation by the transcription factor Escargot. *EMBO J* **33**:2983-2996.

Ma, L. and Jarman, A. P. (2011). Dilatory is a *Drosophila* protein related to AZI1 (CEP131) that is located at the ciliary base and required for cilium formation. *J Cell Sci* **124**:2622-2630.

Maimon, I., Popliker, M. and Gilboa, L. (2014). Without children is required for Stat-mediated *zfh1* transcription and for germline stem cell differentiation. *Development* **141**:2602-2610.

Maines, J. Z. and Wasserman, S. A. (1999). Post-transcriptional regulation of the meiotic Cdc25 protein Twine by the Dazl orthologue Boule. *Nat Cell Biol* **1**:171-174.

Matangkasombut, O., Auty, R. and Buratowski, S. (2004). Structure and function of the TFIID complex. *Adv Protein Chem* **67**:67-92.

Matunis, E. L., Stine, R. R. and de Cuevas, M. (2012). Recent advances in *Drosophila* male germline stem cell biology. *Spermatogenesis* **2**:137-144.

McCarthy, D. J., Chen, Y. and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**:4288-4297.

Meisel, R. P. (2009a). Evolutionary dynamics of recently duplicated genes: Selective constraints on diverging paralogs in the *Drosophila pseudoobscura* genome. *J Mol Evol* **69**:81-93.

Meisel, R. P. (2009b). Repeat mediated gene duplication in the *Drosophila pseudoobscura* genome. *Gene* **438**:1-7.

Meisel, R. P., Hilldorfer, B. B., Koch, J. L., Lockton, S. and Schaeffer, S. W. (2010). Adaptive evolution of genes duplicated from the *Drosophila pseudoobscura* neo-X chromosome. *Mol Biol Evol* **27**:1963-1978.

Mena, V. J. (2012). Characterization of the Stem Cell Niche in *Drosophila* Testes. Thesis MS Biology, Seton Hall University.

Meredith, J. M., Basu, S., Nimmo, D. D., Larget-Thierry, I., Warr, E. L., Underhill, A., McArthur, C. C. *et al.* (2011). Site-specific integration and expression of an anti-malarial gene in transgenic *Anopheles gambiae* significantly reduces *Plasmodium* infections. *PLoS One* **6**:e14587.

Metcalf, C. E. and Wassarman, D. A. (2007). Nucleolar colocalization of TAF1 and testis-specific TAFs during *Drosophila* spermatogenesis. *Dev Dyn* **236**:2836-2843.

Mizuguchi, G., Vassilev, A., Tsukiyama, T., Nakatani, Y. and Wu, C. (2001). ATP-dependent nucleosome remodeling and histone hyperacetylation synergistically facilitate transcription of chromatin. *J Biol Chem* **276**:14773-14783.

Monk, A. C., Siddall, N. A., Volk, T., Fraser, B., Quinn, L. M., McLaughlin, E. A. and Hime, G. R. (2010). HOW is required for stem cell maintenance in the *Drosophila* testis and for the onset of transit-amplifying divisions. *Cell Stem Cell* **6**:348-360.

Moore, A. J., Bacigalupe, L. D. and Snook, R. R. (2013). Integrated and independent evolution of heteromorphic sperm types. *Proc Biol Sci* **280**:20131647.

Morris, C. A., Benson, E. and White-Cooper, H. (2009). Determination of gene expression patterns using in situ hybridization to *Drosophila* testes. *Nat Protoc* **4**:1807-1819.

Mossman, J. A., Pearson, J. T., Moore, H. D. and Pacey, A. A. (2013). Variation in mean human sperm length is linked with semen characteristics. *Hum Reprod* **28**:22-32.

Nimmo, D. D., Alphey, L., Meredith, J. M. and Eggleston, P. (2006). High efficiency site-specific genetic engineering of the mosquito genome. *Insect Mol Biol* **15**:129-136.

Njogu, M., Ricketts, P. G. and Klaus, A. V. (2010). Spermatogenic cyst and organ culture in *Drosophila pseudoobscura*. *Cell Tissue Res* **341**:453-464.

Olivieri, G. and Olivieri, A. (1965). Autoradiographic study of nucleic acid synthesis during spermatogenesis in *Drosophila melanogaster*. *Mutat Res* **2**:366-380.

Ozsolak, F. and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* **12**:87-98.

Papagiannouli, F. and Mechler, B. M. (2009). *discs large* regulates somatic cyst cell survival and expansion in *Drosophila* testis. *Cell Res* **19**:1139-1149.

Parisi, M., Nuttall, R., Edwards, P., Minor, J., Naiman, D., Lu, J., Doctolero, M. *et al.* (2004). A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol* **5**:R40.

Parker, L., Gross, S. and Alphey, L. (2001). Vectors for the expression of tagged proteins in *Drosophila*. *Biotechniques* **31**:1280-1282, 1284, 1286.

Pasini, M. E., Redi, C. A., Caviglia, O. and Perotti, M. E. (1996). Ultrastructural and cytochemical analysis of sperm dimorphism in *Drosophila subobscura*. *Tissue Cell* **28**:165-175.

Pasyukova, E. G., Roshina, N. V. and Mackay, T. F. (2004). Shuttle craft: a candidate quantitative trait gene for *Drosophila* lifespan. *Aging Cell* **3**:297-307.

Pek, J. W., Lim, A. K. and Kai, T. (2009). *Drosophila maelstrom* ensures proper germline stem cell lineage differentiation by repressing *microRNA-7*. *Dev Cell* **17**:417-424.

Perezgasga, L., Jiang, J., Bolival, B., Jr., Hiller, M., Benson, E., Fuller, M. T. and White-Cooper, H. (2004). Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein *matotopetli*. *Development* **131**:1691-1702.

Pflanz, R., Voigt, A., Yakulov, T. and Jackle, H. (2015). *Drosophila* gene *tao-1* encodes proteins with and without a Ste20 kinase domain that affect cytoskeletal architecture and cell migration differently. *Open Biol* **5**:140161.

Pitnick, S., Hosken, D. J. and Birkhead, T. R. (2009). Sperm morphological diversity. In: Birkhead, T.R. and Hosken, D.J. and Pitnick, S. (eds.) *Sperm Biology*. London: Academic Press, pp. 69-149.

Pitnick, S., Marrow, T. and Spicer, G. S. (1999). Evolution of Multiple Kinds of Female Sperm-Storage Organs in *Drosophila*. *Evolution* **53**:1804-1822.

Pitnick, S., Wolfner, M. F. and Dorus, S. (2020). Post-ejaculatory modifications to sperm (PEMS). *Biol Rev Camb Philos Soc* **95**:365-392.

Port, F., Chen, H. M., Lee, T. and Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci USA* **111**:E2967-2976.

Presgraves, D. C., Baker, R. H. and Wilkinson, G. S. (1999). Coevolution of sperm and female reproductive tract morphology in stalk-eyed flies. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**:1041-1047.

Price, T. A., Lize, A., Marcello, M. and Bretman, A. (2012). Experience of mating rivals causes males to modulate sperm transfer in the fly *Drosophila pseudoobscura*. *J Insect Physiol* **58**:1669-1675.

Pulecio, J., Verma, N., Mejia-Ramirez, E., Huangfu, D. and Raya, A. (2017). CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* **21**:431-447.

Quicke, D., Ingram, S., Baillie, H. S. and Gaitens, P. V. (1992). Sperm structure and ultrastructure in the Hymenoptera (Insecta). *Zoologica Scripta* **21**:381-402.

R Core Team. (2017). R: A language and environment for statistical computing [Online]. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/> [Accessed: November 30].

Rathke, C., Baarends, W. M., Jayaramaiah-Raja, S., Bartkuhn, M., Renkawitz, R. and Renkawitz-Pohl, R. (2007). Transition from a nucleosome-based to a protamine-based chromatin configuration during spermiogenesis in *Drosophila*. *J Cell Sci* **120**:1689-1700.

Redhouse, J. L., Mozziconacci, J. and White, R. A. (2011). Co-transcriptional architecture in a Y loop in *Drosophila melanogaster*. *Chromosoma* **120**:399-407.

Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., Liu, L. P. *et al.* (2013). Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. *Proc Natl Acad Sci U S A* **110**:19012-19017.

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S. J. *et al.* (2014). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. *Cell Rep* **9**:1151-1162.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139-140.

Rockwell, A. L. and Hongay, C. F. (2020). Dm Ime4 depletion affects permeability barrier and Chic function in *Drosophila* spermatogenesis. *Mech Dev* **164**:103650.

Sanger, W. G. and Miller, D. D. (1973). Spermatozoan Length in Species of the *Drosophila affinis* Subgroup. *American Midland Naturalist* **90**:485-489.

Sasakawa, K. (2009). Marked sperm dimorphism in the ground beetle *Scarites terricola*: a novel type of insect sperm polymorphism. *Physiological Entomology* **34**:387-390.

Schafer, M., Nayernia, K., Engel, W. and Schafer, U. (1995). Translational control in spermatogenesis. *Dev Biol* **172**:344-352.

Scharer, L., Da Lage, J. L. and Joly, D. (2008). Evolution of testicular architecture in the Drosophilidae: a role for sperm length. *BMC Evol Biol* **8**:143.

Schetelig, M. F. and Handler, A. M. (2013). Germline transformation of the spotted wing drosophilid, *Drosophila suzukii*, with a *piggyBac* transposon vector. *Genetica* **141**:189-193.

Schrader, F. (1960). Evolutionary Aspects of Aberrant Meiosis in Some Pentatominae (Heteroptera). *Evolution* **14**:498-508.

Schroeder, C. M., Valenzuela, J. R., Mejia Natividad, I., Hocky, G. M. and Malik, H. S. (2020). A Burst of Genetic Innovation in *Drosophila* Actin-Related Proteins for Testis-Specific Function. *Mol Biol Evol* **37**:757-772.

Schuettengruber, B., Martinez, A. M., Iovino, N. and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**:799-814.

Sebo, Z. L., Lee, H. B., Peng, Y. and Guo, Y. (2014). A simplified and efficient germline-specific CRISPR/Cas9 system for *Drosophila* genomic engineering. *Fly (Austin)* **8**:52-57.

Sgromo, A., Raisch, T., Backhaus, C., Keskeny, C., Alva, V., Weichenrieder, O. and Izaurralde, E. (2018). *Drosophila* Bag-of-marbles directly interacts with the CAF40 subunit of the CCR4-NOT complex to elicit repression of mRNA targets. *RNA* **24**:381-395.

Shi, Z., Lim, C., Tran, V., Cui, K., Zhao, K. and Chen, X. (2020). Single-cyst transcriptome analysis of *Drosophila* male germline stem cell lineage. *Development* **147**.

Shields, A. R., Spence, A. C., Yamashita, Y. M., Davies, E. L. and Fuller, M. T. (2014). The actin-binding protein profilin is required for germline stem cell maintenance and germ cell enclosure by somatic cyst cells. *Development* **141**:73-82.

Silberglie, R. E., Shepherd, J. G. and Dickinson, J. L. (1984). Eunuchs: The Role of Apyrene Sperm in Lepidoptera? *The American Naturalist* **123**:255-265.

Sinden, D., Badgett, M., Fry, J., Jones, T., Palmen, R., Sheng, X., Simmons, A. *et al.* (2012). Jak-STAT regulation of cyst stem cell development in the *Drosophila* testis. *Dev Biol* **372**:5-16.

Singh, S. R. and Hou, S. X. (2008). Immunohistological Techniques for Studying the *Drosophila* Male Germline Stem Cell. In: Hou, S.X. and Singh, S.R. (eds.) *Germline Stem Cells*. Totowa, NJ: Humana Press, pp. 45-59.

Snook, R. R. (1997). Is the production of multiple sperm types adaptive? *Evolution* **51**:797-808.

Snook, R. R. (1998a). The risk of sperm competition and the evolution of sperm heteromorphism. *Anim Behav* **56**:1497-1507.

Snook, R. R. (1998b). Sperm Production and Sterility in Hybrids between Two Subspecies of *Drosophila pseudoobscura*. *Evolution* **52**:266-269.

Snook, R. R. and Karr, T. L. (1998). Only long sperm are fertilization-competent in six sperm-heteromorphic *Drosophila* species. *Curr Biol* **8**:291-294.

Snook, R. R. and Markow, T. A. (1996). Possible role of nonfertilizing sperm as a nutrient source for female *Drosophila pseudoobscura* Frolova (Diptera: Drosophilidae). *Pan-Pacific Entomologist* **72**:121-129.

Snook, R. R. and Markow, T. A. (2002). Efficiency of gamete usage in nature: sperm storage, fertilization and polyspermy. *Proc Biol Sci* **269**:467-473.

Snook, R. R., Markow, T. A. and Karr, T. L. (1994). Functional nonequivalence of sperm in *Drosophila pseudoobscura*. *Proc Natl Acad Sci USA* **91**:11222-11226.

Southern, H. M., Berger, M. A., Young, P. G. and Snook, R. R. (2018). Sperm morphology and the evolution of intracellular sperm-egg interactions. *Ecol Evol* **8**:5047-5058.

Spradling, A., Fuller, M. T., Braun, R. E. and Yoshida, S. (2011). Germline stem cells. *Cold Spring Harb Perspect Biol* **3**:a002642.

St Johnston, D. (2013). Using mutants, knockdowns, and transgenesis to investigate gene function in *Drosophila*. *Wiley Interdiscip Rev Dev Biol* **2**:587-613.

Stern, C. and Hadorn, E. (1939). The relation between the color of testes and vasa efferentia in *Drosophila*. *Genetics* **24**:162-179.

Stroumbakis, N. D., Li, Z. and Tolia, P. P. (1996). A homolog of human transcription factor NF-X1 encoded by the *Drosophila shuttle craft* gene is required in the embryonic central nervous system. *Mol Cell Biol* **16**:192-201.

Sun, D., Guo, Z., Liu, Y. and Zhang, Y. (2017). Progress and Prospects of CRISPR/Cas Systems in Insects and Other Arthropods. *Front Physiol* **8**:608.

Suzuki, R. and Shimodaira, H. (2015). pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 2.0-0. [Online]. Available at: <https://CRAN.R-project.org/package=pvclust> [Accessed: December 10].

Swallow, J. G. and Wilkinson, G. S. (2002). The long and short of sperm polymorphisms in insects. *Biol Rev Camb Philos Soc* **77**:153-182.

Tachiwana, H., Osakabe, A., Kimura, H. and Kurumizaka, H. (2008). Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins in vitro. *Nucleic Acids Res* **36**:2208-2218.

Takami, Y. and Sota, T. (2007). Sperm competition promotes diversity of sperm bundles in *Ohomopterus* ground beetles. *Naturwissenschaften* **94**:543-550.

Takara Bio USA. (2018). SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing-Single-Cell Transcriptome Analysis with Ultimate Sensitivity [Online]. Clontech. Available at: http://www.clontech.com/GQ/Products/cDNA_Synthesis_and_Library_Construction/Next_Gen_Sequencing_Kits/Single_cell_RNA_Seq_Kits_for_mRNA_seq/Single_Cell_RNA_Seq_v4 [Accessed: February 26].

Tanaka-Matakatsu, M., Uemura, T., Oda, H., Takeichi, M. and Hayashi, S. (1996). Cadherin-mediated cell adhesion and cell motility in *Drosophila* trachea regulated by the transcription factor Escargot. *Development* **122**:3697-3705.

Tanaka, R., Murakami, H., Ote, M. and Yamamoto, D. (2016). Clustered regulatory interspaced short palindromic repeats (CRISPR)-mediated mutagenesis and phenotype rescue by *piggyBac* transgenesis in a nonmodel *Drosophila* species. *Insect Mol Biol* **25**:355-361.

Tates, A. D. (1971). Cytodifferentiation during spermatogenesis in *Drosophila melanogaster*: an electron microscope study. Thesis Ph. D., Rijksuniversiteit.

Theofel, I., Bartkuhn, M., Boettger, T., Gartner, S. M. K., Kreher, J., Brehm, A. and Rathke, C. (2017). tBRD-1 and tBRD-2 regulate expression of genes necessary for spermatid differentiation. *Biol Open* **6**:439-448.

Theofel, I., Bartkuhn, M., Hundertmark, T., Boettger, T., Gartner, S. M., Leser, K., Awe, S. *et al.* (2014). tBRD-1 selectively controls gene activity in the *Drosophila* testis and interacts with two new members of the bromodomain and extra-terminal (BET) family. *PLoS One* **9**:e108267.

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., Matthews, B. B. *et al.* (2019). FlyBase 2.0: the next generation. *Nucleic Acids Res* **47**:D759-D765.

Till-Bottraud, I., Joly, D., Lachaise, D. and Snook, R. R. (2005). Pollen and sperm heteromorphism: convergence across kingdoms? *J Evol Biol* **18**:1-18.

Tokuyasu, K. T. (1974a). Dynamics of spermiogenesis in *Drosophila melanogaster*. III Relation between axoneme and mitochondrial derivatives. *Experimental Cell Research* **84**:239-250.

Tokuyasu, K. T. (1974b). Dynamics of spermiogenesis in *Drosophila melanogaster*. IV. Nuclear transformation. *J Ultrastruct Res* **48**:284-303.

Tokuyasu, K. T. (1975a). Dynamics of spermiogenesis in *Drosophila melanogaster*. V. Head-tail alignment. *J Ultrastruct Res* **50**:117-129.

Tokuyasu, K. T. (1975b). Dynamics of spermiogenesis in *Drosophila melanogaster*. VI. Significance of "onion" nebenkern formation. *J Ultrastruct Res* **53**:93-112.

Tokuyasu, K. T., Peacock, W. J. and Hardy, R. W. (1972a). Dynamics of spermiogenesis in *Drosophila melanogaster*. I. Individualization process. *Z Zellforsch Mikrosk Anat* **124**:479-506.

Tokuyasu, K. T., Peacock, W. J. and Hardy, R. W. (1972b). Dynamics of spermiogenesis in *Drosophila melanogaster*. II. Coiling process. *Z Zellforsch Mikrosk Anat* **127**:492-525.

Tu, R., Duan, B., Song, X. and Xie, T. (2020). Dlp-mediated Hh and Wnt signaling interdependence is critical in the niche for germline stem cell progeny differentiation. *Sci Adv* **6**:eaaz0480.

Tulina, N. and Matunis, E. (2001). Control of stem cell self-renewal in *Drosophila* spermatogenesis by JAK-STAT signaling. *Science* **294**:2546-2549.

Ueishi, S., Shimizu, H. and Inoue, Y. H. (2009). Male germline stem cell division and spermatocyte growth require insulin signaling in *Drosophila*. *Cell Struct Funct* **34**:61-69.

van der Horst, G. and Marea, L. (2014). Sperm form and function in the absence of sperm competition. *Mol Reprod Dev* **81**:204-216.

Vedelek, V., Bodai, L., Grezal, G., Kovacs, B., Boros, I. M., Laurinyecz, B. and Sinka, R. (2018). Analysis of *Drosophila melanogaster* testis transcriptome. *BMC Genomics* **19**:697.

Verrotti, A. C. and Wharton, R. P. (2000). Nanos interacts with cup in the female germline of *Drosophila*. *Development* **127**:5225-5232.

Vibrantovski, M. D., Chalopin, D. S., Lopes, H. F., Long, M. and Karr, T. L. (2010). Direct evidence for postmeiotic transcription during *Drosophila melanogaster* spermatogenesis. *Genetics* **186**:431-433.

Vidaurre, V. and Chen, X. (2021). Epigenetic regulation of *Drosophila* germline stem cell maintenance and differentiation. *Dev Biol* **473**:105-118.

Vieillard, J., Paschaki, M., Duteyrat, J. L., Augiere, C., Cortier, E., Lapart, J. A., Thomas, J. *et al.* (2016). Transition zone assembly and its contribution to axoneme formation in *Drosophila* male germ cells. *J Cell Biol* **214**:875-889.

Vinckenbosch, N., Dupanloup, I. and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci USA* **103**:3220-3225.

Voog, J., D'Alterio, C. and Jones, D. L. (2008). Multipotent somatic stem cells contribute to the stem cell niche in the *Drosophila* testis. *Nature* **454**:1132-1136.

Voog, J., Sandall, S. L., Hime, G. R., Resende, L. P., Loza-Coll, M., Aslanian, A., Yates, J. R., 3rd *et al.* (2014). Escargot restricts niche cell to stem cell conversion in the *Drosophila* testis. *Cell Rep* **7**:722-734.

Walker, A. K., Rothman, J. H., Shi, Y. and Blackwell, T. K. (2001). Distinct requirements for *C. elegans* TAF(II)s in early embryonic transcription. *EMBO J* **20**:5269-5279.

Wang, H., Singh, S. R., Zheng, Z., Oh, S. W., Chen, X., Edwards, K. and Hou, S. X. (2006). Rap-GEF signaling controls stem cell anchoring to their niche through regulating DE-cadherin-mediated cell adhesion in the *Drosophila* testis. *Dev Cell* **10**:117-126.

Wang, J., Duncan, D., Shi, Z. and Zhang, B. (2012). WEB-based GENE SeT ANALYSIS Toolkit (WebGestalt): update 2013 [Online]. Available at: <http://nar.oxfordjournals.org/content/41/W1/W77.full> [Accessed: February 26].

Wang, Z. and Mann, R. S. (2003). Requirement for two nearly identical TGIF-related homeobox genes in *Drosophila* spermatogenesis. *Development* **130**:2853-2865.

White-Cooper, H. (2004). Spermatogenesis: Analysis of Meiosis and Morphogenesis. In: Henderson, D.S. (ed.) *Drosophila Cytogenetics Protocols*. Vol. 247. Totowa, NJ: Humana Press, pp. 45-75.

White-Cooper, H. (2009). Studying how flies make sperm--investigating gene function in *Drosophila* testes. *Mol Cell Endocrinol* **306**:66-74.

White-Cooper, H. (2010). Molecular mechanisms of gene regulation during *Drosophila* spermatogenesis. *Reproduction* **139**:11-21.

White-Cooper, H. and Davidson, I. (2011). Unique aspects of transcription regulation in male germ cells. *Cold Spring Harb Perspect Biol* **3**.

White-Cooper, H., Leroy, D., MacQueen, A. and Fuller, M. T. (2000). Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development* **127**:5463-5473.

White-Cooper, H., Schafer, M. A., Alphey, L. S. and Fuller, M. T. (1998). Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. *Development* **125**:125-134.

Wigby, S., Brown, N. C., Allen, S. E., Misra, S., Sitnik, J. L., Sepil, I., Clark, A. G. *et al.* (2020). The *Drosophila* seminal proteome and its role in postcopulatory sexual selection. *Philos Trans R Soc Lond B Biol Sci* **375**:20200072.

Wright, A. V., Nunez, J. K. and Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**:29-44.

Wu, A. R., Wang, J., Streets, A. M. and Huang, Y. (2017). Single-Cell Transcriptional Analysis. *Annu Rev Anal Chem (Palo Alto Calif)* **10**:439-462.

Wu, H., Sun, L., Wen, Y., Liu, Y., Yu, J., Mao, F., Wang, Y. *et al.* (2016). Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans. *Proc Natl Acad Sci USA* **113**:4134-4139.

Xu, T. and Rubin, G. M. (1993). Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**:1223-1237.

Yamashita, Y. M., Jones, D. L. and Fuller, M. T. (2003). Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. *Science* **301**:1547-1550.

Yang, H., Jaime, M., Polihronakis, M., Kanegawa, K., Markow, T., Kaneshiro, K. and Oliver, B. (2018). Re-annotation of eight *Drosophila* genomes. *Life Sci Alliance* **1**:e201800156.

Yang, L., Titlow, J., Ennis, D., Smith, C., Mitchell, J., Young, F. L., Waddell, S. *et al.* (2017). Single molecule fluorescence in situ hybridisation for quantitating post-transcriptional regulation in *Drosophila* brains. *Methods* **126**:166-176.

Yu, Z., Wu, H., Chen, H., Wang, R., Liang, X., Liu, J., Li, C. *et al.* (2013). CAF-1 promotes Notch signaling through epigenetic control of target gene expression during *Drosophila* development. *Development* **140**:3635-3644.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K. *et al.* (2018). Ensembl 2018. *Nucleic Acids Res* **46**:D754-D761.

Zhang, B., Kirov, S. A. and Snoddy, J. R. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts [Online]. WebGestalt. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&list_uids=15980575&dopt=Abstract [Accessed: February 26].

Zur Lage, P., Newton, F. G. and Jarman, A. P. (2019). Survey of the Ciliary Motility Machinery of *Drosophila* Sperm and Ciliated Mechanosensory Neurons Reveals Unexpected Cell-Type Specific Variations: A Model for Motile Ciliopathies. *Front Genet* **10**:24.