



Novel Insights into the Non-Neuronal Expression
& Function of the RNA-Binding Protein Alan
Shepard (Shep) during *D. melanogaster*
Oogenesis

Abdulla Omar Almoalem

September 2023

This thesis is submitted for the Degree of
Doctor of Philosophy

*To my Mum, Dad,
& Myself*

Acknowledgments

First and foremost, I would like to extend my heartfelt gratitude to my principal supervisor, Sonia López de Quinto, whose invaluable guidance, patience, and unwavering support paved the way for the completion of this PhD thesis. Your expertise and mentorship have been instrumental in shaping my academic journey. I am deeply indebted to her for the constructive feedback, insightful critiques, and encouragement. Your collective wisdom has been a beacon that guided my research pursuits.

I also want to thank my supervisory team: Dr. Nicholas Kent (2nd supervisor) and Dr. Simon Scofield (assessor). Their support and constructive feedback during our annual progress monitoring events throughout my PhD degree programme.

My sincere appreciation goes out to the Fly Group based at Cardiff University, including the labs of Prof. Helen White-Cooper, Dr. Fisun Hamaratoglu Dion, Dr. Wynand van der Goes Van Naters, and Dr. Michael Taylor. I would also like to extend my gratitude to fly researchers based at the UK Dementia Research Institute (DRI), specifically in the labs of Dr. Gaynor Smith and Dr. Owen Peters. Their collaborative spirit, in-depth discussions, and relentless pursuit of knowledge created an intellectually stimulating environment that was not only conducive to my research but also pivotal in shaping my academic journey. Engaging with such a passionate and dedicated community has enriched my understanding and perspective, and I am immensely grateful for the opportunities and insights they have provided.

I would also like to express my preemptive gratitude to my distinguished examiners, Professor Michael Lodomery and Professor Helen White-Cooper, for dedicating their time and expertise to review this thesis. The commitment and discernment involved in critically evaluating an academic manuscript are substantial, and for that, I deeply appreciate your willingness to undertake this responsibility. I am confident that your expertise and feedback will serve to significantly enrich the quality of my research immeasurably, and I look forward to learning from your insights.

Additionally, a heartfelt thank you goes to my former personal tutor, Dr. Walter Dewitte, for chairing my Viva. Your continued involvement in my academic journey, even beyond our initial interactions, is a testament to your dedication, and I am genuinely grateful for your support.

I would also like to acknowledge Elissa Lei, Dahong Chen, and Kathleen Beckingham, who generously shared their reagents, specifically the anti-Shep antibodies, with me. Your willingness to share and facilitate research beyond your own labs highlights the collaborative spirit of the scientific community. This invaluable gesture not only catalyzed my research progress but also underscored the importance of sharing and cooperation in advancing scientific knowledge.

Special thanks are reserved for my fellow researchers and close lab mates. To Silvia Ziliotto, Menna El Gammal, Hoi Ping Weeks, Dana Jackson, Jack Bruton, and Irene Castellan: I am deeply grateful for everything. Engaging with each one of you in both academic and personal capacities has been one of the great privileges of this journey. I consider myself truly fortunate to be part of such a vibrant, supportive, and talented group. Your camaraderie has been the linchpin that made the highs of this process even more rewarding and the challenges more bearable. The strength and beauty of our collective spirit is something I deeply cherish, and it has undoubtedly made this endeavour all the more memorable.

On a personal note, heartfelt gratitude is extended to my entire family. Special mention must be made to my parents, my big brother, little sister, and beloved grandmothers. Your unconditional love, endless support, and unshakable faith have been my stronghold. It's your unwavering belief in my potential and aspirations that has truly been the cornerstone of my journey, propelling me forward even in the most challenging times.

∞ Abstract ∞

Abstract:

In *Drosophila*, the major body axes of the adult fly are established during oogenesis. This is achieved within the oocyte cytoplasm through the asymmetric distribution and local translation of key maternal mRNAs, such as *oskar* and *gurken*. The spatiotemporal regulation of these mRNAs is tightly controlled at the post-transcriptional level by the dynamic remodelling of their ribonucleoprotein complexes.

The *D. melanogaster alan shepard (shep)* gene encodes a highly conserved RNA-binding protein with two RNA-recognition motifs. Although *shep* has been predominantly studied in the nervous system, much remains unknown about its molecular function in other tissues. In a yeast two-hybrid screen, our group identified Shep as a potential interacting partner of the RNA-binding protein PTB, a known regulator of *oskar* and *gurken* mRNAs. Therefore, we decided to characterize the expression patterns and potential roles of the *Drosophila* Shep during oogenesis.

Our *in vivo* characterization revealed that Shep localizes to the oocyte cortex during mid-oogenesis with enrichment at the posterior and dorsal-anterior corners, where *oskar* and *gurken* mRNAs are restricted to define posterior and dorsal fate, respectively. Furthermore, we demonstrated that Shep interacts with *oskar* and *gurken* mRNAs *in vivo* and *in vitro*. Moreover, overexpression of Shep reduced Oskar and Gurken levels, which compromised body axis formation. Conversely, *shep* loss-of-function led to the ectopic and premature translation of *oskar* mRNA within the oocyte cytoplasm. Our work reveals a novel non-neuronal role for *shep* in *Drosophila* body axis patterning by regulating *oskar* and *gurken* mRNAs.

Here, we show that *shep* is transcribed into various ovarian mRNA transcripts leading to the expression of several isoforms with distinct localization patterns. We also demonstrate that the conserved Delta-Notch signalling pathway regulates *shep* expression within ovarian somatic cells. Moreover, we provide *in vivo* and *in vitro* evidence suggesting that Shep and PTB are part of similar RNP complexes in the ovary and may work together to regulate a variety of mRNAs at the post-transcriptional level.

Keywords:

Drosophila melanogaster, Oogenesis, RNA-binding proteins, Alan Shepard (Shep), Notch-Delta signalling pathway, *oskar*, *gurken*, Dorsal appendages Cytoplasmic post-transcriptional regulation, Translational control, Polypyrimidine tract-binding protein (PTB), Messenger ribonucleoprotein (mRNP) complexes.

Table of Contents

<i>Abbreviations</i>	1
<i>List of Figures</i>	7
<i>List of Tables</i>	10
1 Chapter I – Introduction:	13
1.1 Gene Expression Regulation	13
1.1.1 RNA – An Essential Molecule in the Regulation of Gene Expression	14
1.1.2 The Role of Untranslated Regions in Defining mRNA Fate	17
1.1.3 Role of RNA-Binding Proteins in Gene Expression Regulation	20
1.2 <i>Drosophila</i> Fruit Fly as an Animal Model System for Development	25
1.2.1 Benefits of Using <i>Drosophila</i> as a Model Organism	25
1.2.2 Overview of <i>D. melanogaster</i> Development	26
1.2.3 Ovaries are a Powerful Model Organ to Study the Function of RNA-Binding Proteins.....	28
1.3 The <i>Drosophila</i> Toolkit for Genetic Manipulation	28
1.3.1 The UAS-Gal4 System.....	29
1.3.2 Protein-Trap as a Tool for the Characterization of Gene Expression & Protein Distribution	34
1.4 Female Reproductive Organ & Oocyte Development in <i>Drosophila</i>	36
1.4.1 Development of <i>Drosophila</i> Ovaries.....	37
1.4.2 Ovarian Morphology of Adult Female <i>Drosophila</i>	37
1.4.3 The Process of Oogenesis in <i>Drosophila</i>	38
1.4.4 Border Cell Migration	46
1.4.5 Establishment of Major Body Axes During Oogenesis	48
1.4.6 mRNA Localization & Spatial Translational Regulation	49
1.5 The <i>Drosophila</i> Alan Shepard (Shep) Protein	54
1.5.1 The <i>shep</i> Gene in <i>Drosophila</i>	54
1.5.2 Biological Roles of <i>Drosophila</i> Shep & its Non- <i>Drosophila</i> Orthologs.....	57
1.6 Aims & Objectives	60
2 Chapter II – Materials & Methods:	63
2.1 Standard Molecular Biology Techniques	63
2.1.1 Cloning Vectors	63
2.1.2 DNA Templates	64
2.1.3 Polymerase Chain Reaction	68
2.1.4 Oligonucleotide & Primer Design	72
2.1.5 Agarose Gel Electrophoresis.....	74
2.1.6 Restriction Endonuclease Digestion of DNA.....	76
2.1.7 Dephosphorylation of Plasmid DNA	77
2.1.8 DNA Ligation	78
2.1.9 Bacterial Growth Media.....	78
2.1.10 Bacterial Transformation.....	80
2.1.11 Inoculation of Bacterial Overnight Culture	81
2.1.12 Purification of Plasmid DNA.....	82
2.1.13 DNA Sanger Sequencing of Constructs	82
2.1.14 <i>Drosophila</i> Transgenesis – Germline Transformation of Transgenic Constructs	83
2.1.15 Spectrophotometric Quantification of Nucleic Acid.....	84

2.1.16	Quality Assessment of RNA Samples	84
2.1.17	Bacterial Glycerol Stocks	85
2.1.18	Chemical-Based Disinfection & Decontamination of Bacterial Cultures	85
2.2	Molecular Cloning – Generation of New Recombinant Constructs	86
2.2.1	Constructs to Generate Transgenic Fly Stocks	86
2.2.2	pBluescript-II Constructs for <i>in vitro</i> Synthesis of RNA Probes	96
2.2.3	Constructs for Protein Expression in Bacteria	97
2.3	General Biochemical Techniques	101
2.3.1	<i>In vitro</i> Transcription	101
2.3.2	Preparation of <i>Drosophila</i> Ovarian Lysates	103
2.3.3	Immunoblotting – Fluorescent Western Blot Detection Method	105
2.3.4	Bacterial Protein Expression	109
2.4	Biochemical RNA-Protein & Protein-Protein Interaction Assays	111
2.4.1	RNA-Affinity Pulldown Assay	111
2.4.2	Co-Immunoprecipitation (Co-IP) Assay	114
2.4.3	RNA Immunoprecipitation Assay	117
2.5	Fly Husbandry & Routine Fly Work	120
2.5.1	General Fly Maintenance	120
2.5.2	Fly Stocks	123
2.5.3	Mating Scheme & Experimental Crosses	124
2.5.4	Preparation of Female Flies for Ovarian Dissection, Egg Collection, or Fertility Assay	128
2.5.5	Fertility Assay of <i>D. melanogaster</i> Adult Females	129
2.5.6	Dissection of <i>D. melanogaster</i> Adult Ovaries	130
2.5.7	Measuring the Size of <i>D. melanogaster</i> Ovaries	130
2.5.8	Quantification of the Number of Ovarioles Within <i>D. melanogaster</i> Ovaries	131
2.5.9	Oviposition Assays – Collection of <i>D. melanogaster</i> Mature Eggs	132
2.5.10	Whole-Mount Immunostaining of Adult <i>Drosophila</i> Ovaries	136
2.5.11	Microscopy	139
2.5.12	Microscopy Image Construction:	140
2.6	Fruit Fly Database	140
2.7	Computational Analysis	140
2.7.1	Identification of Shep Binding Sites on Target RNA Transcripts	140
2.7.2	Identification of Notch Regulatory Elements in the <i>shep</i> Gene Locus	141
2.7.3	Identification of a Putative Nuclear Localization Signal in Shep Protein	142
2.7.4	Molecular Weight Estimation of Electrophoresed Proteins	143
2.8	Statistics	143
2.9	Experimental Replicas	144
3	Chapter III – Shep Expression Pattern During <i>D. melanogaster</i> Oogenesis:	146
3.1	Preliminary Data	146
3.2	Shep as a Potential Interacting Partner of PTB	147
3.3	Ovaries of Shep GFP Protein-Trap Flies are Small	148
3.4	Expression Pattern of Shep in <i>D. melanogaster</i> Ovaries	152
3.4.1	Characterization of the Shep Isoforms Expressed in <i>Drosophila</i> Ovarian Tissues	152
3.4.2	<i>Drosophila</i> Ovaries Express Multiple <i>shep</i> Transcripts	154
3.4.3	Characterization of the Subcellular Distribution Pattern of Shep Protein in Ovaries	157
3.5	Notch Signalling Regulates Expression of Shep in the Ovary	167
3.5.1	Downregulation of Delta Levels in the Female Germline	168
3.5.2	Knockdown of the Notch Receptor in Ovarian Somatic Cells	172

3.6	Conclusions.....	175
4	Chapter IV – Deciphering the Role of Shep during <i>D. melanogaster</i> Oogenesis:	178
4.1	Loss-of-Function Analysis via RNAi-Mediated Downregulation of Shep	178
4.1.1	RNAi-Mediated Reduction of <i>shep</i> Levels in the Germline Resulted in No Obvious <i>oskar</i> -Related Phenotype.....	178
4.1.2	<i>oskar</i> -Unrelated Phenotypes Upon Germline Reduction of Shep Levels.....	185
4.2	Characterization of <i>shep</i> Deficiencies & P-element Insertions	192
4.3	Shep Gain-of-Function Analysis.....	199
4.3.1	Females Overexpressing Shep in the Germline are Sterile.	201
4.3.2	Overexpression of Shep in the Germline Disrupts Eggshell Patterning.....	203
4.3.3	Germline Overexpression of Shep Disrupts the Expression of Key Maternal mRNAs	205
4.4	Conclusions.....	215
5	Chapter V – Characterizing the Composition of the Ovarian Shep mRNP Complexes:	218
5.1	Characterization of Interaction Between Shep and its mRNA Targets.....	219
5.1.1	Computational Prediction of Shep-Binding Sites within Candidate mRNAs	219
5.1.2	Biochemical Mapping of Shep Binding to mRNA Targets	221
5.2	Nature of the Interaction Between the RNA-Binding Proteins Shep and PTB	225
5.2.1	Does Shep Directly Interact with PTB <i>in vivo</i> ?.....	225
5.2.2	Is the Shep-PTB Interaction RNA-Dependent?	227
5.3	Identification of Ovarian mRNAs Bound to Shep <i>in vivo</i>	229
5.3.1	Optimization of the RIP Protocol	232
5.3.2	Quality Control Check of RIP Samples for Transcriptome Sequencing.....	233
5.4	Conclusions.....	236
6	Chapter VI – Discussion:	239
6.1	Overview of the State of Knowledge Prior to Our Study.....	239
6.2	<i>shep</i> Displays a Complex Gene Expression Pattern in <i>Drosophila</i> Ovaries	240
6.2.1	Transcription of Multiple <i>shep</i> Transcripts Gives Rise to Several Protein Isoforms	240
6.2.2	Shep Expression Pattern is Altered in the CC00236 Protein-Trap Line.....	244
6.2.3	Shep Isoforms Display Different Subcellular Distribution Patterns Reflecting Their Diverse mRNP Complex Associations.....	246
6.2.4	Shep Isoforms Displaying a Posterior Crescent Pattern.....	248
6.3	The Notch Signalling Pathway Regulates <i>shep</i> Expression	249
6.3.1	<i>shep</i> Represents a Novel Target Gene of the Notch Signalling Pathway	252
6.4	Shep Levels in the Germline Affect Border Cell Migration	255
6.5	Shep & PTB are Part of Similar mRNP Complexes in the Ovary	257
6.6	Regulatory Role of Shep as a Component of <i>oskar</i> & <i>gurken</i> mRNP Complexes..	260
6.7	Potential Regulatory Role of Shep as a Component of the <i>nanos</i> mRNP Complex	265
6.8	Shep is Not Nuclear in any of the Ovarian Cell Types	266
6.9	Future Directions	270
7	Chapter VII – Conclusions:	274

7.1	Characterization of <i>shep</i> Gene Expression Pattern in the Ovary	274
7.2	Biological Role of Shep Protein in the Female Germline	275
7.3	Shep Protein Interactions with mRNP Complexes & PTB	276
	References:	278
	Appendices:	313
	Appendix 1: List of Primer Sequences Designed	313
	Appendix 1.1 – Primers design for the identification of <i>shep</i> transcripts using qualitative RT-PCR.	317
	Appendix 2: List of FastDigest® Enzymes	318
	Appendix 3: Maps of Empty Vectors Used During Cloning of Constructs Listed in Appendix 4	319
	Appendix 4: Recombinant Constructs Generated	325
	Appendix 4.1 – Constructs generated for the synthesis of RNA probes.....	325
	Appendix 4.2 – Constructs generated for protein expression	326
	Appendix 4.3 – Constructs generated for fly transgenesis	327
	Appendix 5: Computational Scoring of Predicted Shep Binding Sites	328
	Appendix 5.1 – Shep RNA-Binding Predictions Based on Motif 1.	329
	Appendix 5.2 – Shep RNA-Binding Predictions Based on Motif 2.	329
	Appendix 5.3 – Shep RNA-Binding Predictions Based on Motif 3.	330
	Appendix 5.4 – Shep Binding Putative Sites Plotted Based on All Three RNAcompete Motifs (Raw Data).....	330
	Appendix 5.5 – Lengths of mRNA Candidates Investigated in Motif Analysis.	331
	Appendix 5.6 – The Nucleotide Code: IUPAC Nomenclature for Nucleic Acids.....	332
	Appendix 6: List of Fly Stocks	333
	Appendix 6.1 – Genetic Background of Deficiency Fly Stocks	336
	Appendix 7: Alignment of Shep Isoforms	337

List of Abbreviations

Abbreviation(s)	Stands for
A	Anterior
A	Adenine
aa	Amino acid
APS	Ammonium Persulfate
ARE	AU-rich elements
attB	Bacterial attachment site
attP	Phage attachment site
AU-rich	Adenine- and uridine-rich sequences
AUG	Start/Initiation codon
Bam	Bag of marbles
Bcd	Bicoid
BDGP	Berkeley Drosophila Genome Project
bp	Base pair
BSA	Bovine Serum Albumin
C	Cytosine
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
cDNA	Complementary DNA
CDS	Coding sequence
CoIP	Co-immunoprecipitation
CPE	Cytoplasmic polyadenylation element
Cq	Quantification cycle (commonly known as cycle threshold, CT)
Cyo	Curly O (Genetic marker affecting wing shape)
D	Dorsal
<i>D. magna</i>	<i>Daphnia magna</i>
<i>D. melanogater</i>	<i>Drosophila melanogater</i>
DA	Splice donor sequen
DAPI	4',6-diamidino-2-phenylindole dilactate
DE-Cad	DE-Cadherin (called Shg in Drosophila)

DGRC	Drosophila Genomics Resource Center
dH ₂ O	Distilled water
Dhc	Dynein heavy chain
DHFR	Dihydrofolate reductase (Human metabolic enzyme)
DIG	Digoxigenin
Δ	Delta
DNA	Deoxyribonucleic acid
DPE	Days post eclosion
DPP	Decapentaplegic (Drosophila BMP signalling homolog)
dsRBD	Double-stranded RNA-binding domain
DTT	1,4-Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMBL	European Molecular Biology Laboratory
EMT	Epithelial-to-mesenchymal transition
EST	Expressed sequence tags
F ₀	Parent flies used to set up a cross
F ₁	First generation of a cross
F ₂	Second generation of a cross
FISH	RNA Fluorescence in situ hybridization
G	Guanine
g	Gram
GAL	Galactose
Gal4	Galactose-responsive transcription factor
Gal80	Galactose-responsive transcriptional repressor
gDNA	Genomic DNA (also known as chromosomal DNA)
GFP	Green Fluorescent Protein
Grk	Gurken
GSC	Germline stem cells
GST	Glutathione S-transferase tag
H ₂ O	Water
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine

hnRNP	Heterogeneous nuclear ribonucleoproteins
HPLC	High-performance liquid chromatography
HS	High sensitivity
Hsp70	Heat shock protein 70
Hu	Humeral (Genetic marker affecting number of thoracic lateral bristles)
If	Irregulate facet (Genetic marker affecting eye shape)
IGEPAL	Octylphenyl-polyethylene glycol (chemically indistinguishable from NP-40)
IRES	Internal Ribosome Entry Sites
K10	Female sterile (1) K10
Kb	Kilobase
KCl	Potassium chloride
kDa	Kilodalton
KH	K-homology domain
Khc	Kinesin heavy chain
KOH	Potassium hydroxide
L	Litre
LB	Lysogeny broth (commonly referred to Luria-Bertani Broth)
M	Molar
m7G cap	5'-7-methylguanosine cap
MBP	Maltose binding protein (common protein expression tag)
mCh	mCherry (a bright red monomeric fluorescent protein)
MCS	Multiple cloning site
MgCl ₂	Magnesium chloride
mL	Milliliter
mm	Millimeter
mM	Millimolar
MRE	MicroRNA response elements
mRNA	Messenger mRNA
mRNP	Messenger ribonucleoprotein
MTD	Maternal triple driver
MW	Molecular weight
N	Notch
NaCl	Sodium Chloride
NaOH	Sodium hydroxide

ncRNA	Non-coding RNA
NEB	New England Biolabs
ng	Nanogram
NICD	Notch intra-cellular domain
NLS	Nuclear localization signal
nm	Nanometers (light wavelength measurment)
Nos	Nanos
NP-40	Nonyl phenoxypolyethoxyethanol (detergent that is also commonly known as Nonidet P-40 or IGEPAL)
nt	Nucleotides
NTC	No template control
OD	Optical Density (wavelength absorbance measurement)
ORF	Open reading frame
Osk	Oskar
P	Posterior
PABP	Poly(A) binding protein
PBS	Phosphate buffered saline
PBST	Phosphate-buffered saline with Tween 20 (or Triton X-100)
PBSTB	Phosphate-buffered saline with Tween 20 (or Triton X-100) and BSA
PBSTM	Phosphate-buffered saline with Tween 20 (or Triton X-100) and Milk
PCR	Polymerase chain reaction
pH	Potential of hydrogen
PI	Protease inhibitor
PMSF	Phenylmethylsulfonyl fluoride
Poly(A)	Polyadenylic acid
PSA	Polyadenylation signal
PTB	Polypyrimidine tract-binding protein
qPCR	Quantitative polymerase chain reaction
RBD	RNA-Binding Domain
RBMS1-3	RNA-binding motif, single-stranded-interacting protein 1-3
RBP	RNA-Binding Protein
RIP	RNA Immunoprecipitation
RIP-RT-PCR	RNA Immunoprecipitation followed by Reverse transcription polymerase chain reaction
RIP-seq	RNA Immunoprecipitation followed by RNA sequencing

RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNAi	RNA interference
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RRM	RNA recognition motif domain
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
S10	Soluble cytoplasmic protein extract
SA	Splice acceptor sequen
Sb	Stubble (Genetic marker affecting length of thoracic dorsal bristles)
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
Shep	Alan Shepard
Shg	Shotgun (aka. Drosophila DE-Cadherin)
shRNA	Short hairpin RNA
Slbo	Slow border cells
SLQ	Sonia López de Quinto
SSC	Somatic stem cells
SSC	Saline-sodium citrate
ssDNA	Single-stranded DNA
Stau	Staufen
Su(H)	Suppressor of Hairless
Sup-26	Supressour-26
SV40	Simian virus 40
Sxl	Sex-lethal
T	Thymine
Ta	Annealing temperature
TBE	Tris-borate-EDTA
TBS	Tris-Buffered Saline
TBST	Tris-Buffered Saline with Tween 20 (or Triton X-100)

TBSTM	Tris-Buffered saline with Tween 20 (or Triton X-100) and Milk
TGF α	Transforming growth factor alpha
T _j	Traffic jam
T _m	Melting temperature
TRiP	Transgenic RNAi Project
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TSS	Transcriptional start site
Tsu	Tsunagi (also referred to Y14)
Tub	Tubulin
U	Uridine
U	Unit (standard method for protein measurement)
UAS	Upstream activating sequence
Upd	Unpaired
UTP	Uridine-5'-triphosphate
V	Ventral
WID	Wing Imaginal Disc
Y14	Tsunagi (also referred to tsu)
ZnF	Zinc-finger
α	Alpha
β	Beta
μ g	Microgram
μ L	Microlitre
μ M	Micromolar
$^{\circ}$ C	Degree Celsius
\times g	Gravitational force (also referred to relative centrifugal force, RCF)
2L	Left chromosomal arm of Chromosome 2
3'UTR	3' untranslated region
3D	Three-dimensional
3L	Left chromosomal arm of Chromosome 3
5'UTR	5' untranslated region



List of Figures & Tables



Figure #	Title	Page N ^o
Chapter 1 – Introduction:		
Figure 1.1	Generic features of mature eukaryotic mRNA.	17
Figure 1.2	Overview of post-transcriptional regulatory processes in eukaryotic cells.	21
Figure 1.3	<i>Drosophila melanogaster</i> life cycle.	27
Figure 1.4	The UAS-Gal4 system induces spatiotemporal expression of transgenes in <i>Drosophila</i> .	31
Figure 1.5	Comparison of the pUAS _T and pUAS _{Sp} expression vectors.	33
Figure 1.6	The principle of GFP protein-traps for tagging endogenous proteins.	36
Figure 1.7	Overview of the <i>Drosophila</i> female reproductive system.	38
Figure 1.8	<i>D. melanogaster</i> oogenesis.	40
Figure 1.9	Patterning and morphogenesis of the <i>Drosophila</i> ovarian follicular epithelium.	44
Figure 1.10	Border cell cluster migration within <i>D. melanogaster</i> egg chambers.	46
Figure 1.11	Patterning of the <i>Drosophila</i> body axes.	51
Figure 1.12	<i>Drosophila</i> shep gene orthologs.	55
Figure 1.13	Gene region of <i>shep</i> in <i>Drosophila melanogaster</i> .	56
Figure 1.14	Isoforms of the <i>Drosophila</i> Shep protein.	57
Chapter 2 – Materials & Methods:		
Figure 2.1	Workflow for analyzing transcript I expression in the ovary.	71
Figure 2.2	TRIP hairpin oligonucleotides design.	87
Figure 2.3	Generation of Shep RRM mutants unable to bind RNA.	91
Figure 2.4	Workflow for obtaining Shep-E RRM mutant via Overlap Extension PCR.	95
Figure 2.5	A brief overview of the T7 in vitro transcription workflow.	102
Figure 2.6	Assembly of semi-dry transfer sandwich using the trans-blot turbo system	107

Figure 2.7	Biotinylated RNA affinity pulldown assay for the detection of Protein-RNA interactions.	112
Figure 2.8	Co-Immunoprecipitation assay for the detection of protein complexes.	115
Figure 2.9	A Schematic of the RNA Immunoprecipitation Assay for the Identification of RNA Targets.	118
Figure 2.10	The impact of ambient temperature on the developmental progression of <i>Drosophila melanogaster</i> .	123
Figure 2.11	Layout of placing <i>Drosophila</i> ovaries on glass slide for imaging.	131
Figure 2.12	Quantification of ovarioles within <i>D. melanogaster</i> ovaries.	132
Figure 2.13	Workflow for collecting <i>Drosophila</i> eggs.	134
Chapter 3 – Shep Expression Pattern During <i>D. melanogaster</i> Oogenesis:		
Figure 3.1	Shep enrichment at the posterior pole of the oocyte.	148
Figure 3.2	Genetic manipulation of shep affects female reproductive organs.	150
Figure 3.3	Different Shep protein isoforms are expressed in the <i>Drosophila</i> ovary.	153
Figure 3.4	Analysis of the shep transcripts expressed in <i>Drosophila</i> ovaries.	156
Figure 3.5	Distribution of Shep protein during <i>Drosophila</i> oogenesis.	160
Figure 3.6	Different Shep isoforms display distinct localization patterns in late-stage oocytes.	163
Figure 3.7	mCherry-tagged Shep isoform distribution in various ovarian tissues.	165
Figure 3.8	Shep subcellular localization to the oocyte posterior pole is Kinesin-1 dependent.	167
Figure 3.9	Direction of Notch signalling pathway in the <i>Drosophila</i> egg chamber	169
Figure 3.10	<i>Delta</i> RNAi knockdown in the female germline affects Shep expression in ovarian somatic cells.	171
Figure 3.11	RNAi knockdown of the Notch receptor in the soma affects Shep expression in ovarian somatic cells.	174
Chapter 4 – Deciphering the Role of Shep during <i>D. melanogaster</i> Oogenesis:		
Figure 4.1	Expression of <i>oskar</i> in <i>shep</i> RNAi-mediated knockdowns in the female germline.	181
Figure 4.2	Expression time-window of various germline-specific Gal4 drivers during oogenesis.	183
Figure 4.3	Expression of <i>shep</i> RNAi hairpins efficiently depletes Shep-GFP levels in the female germline.	184
Figure 4.4	Depletion of <i>shep</i> levels in the female germline affects egg chamber architecture.	186
Figure 4.5	Shep levels in the female germline can impact border cell migration.	188

Figure 4.6	Shep levels in the female germline regulate Lamin C expression in stalk cells.	190
Figure 4.7	FasIII expression pattern following germline knockdown of <i>shep</i> .	191
Figure 4.8	Genomic locations of deficiencies and P-element insertions in relation to the <i>shep</i> gene.	194
Figure 4.9	Loss-of-function of the <i>shep</i> gene dysregulates <i>oskar</i> expression in the oocyte.	196
Figure 4.10	<i>oskar</i> expression is dysregulated in different <i>shep</i> mutant backgrounds.	198
Figure 4.11	Schematic of the Shep germline overexpression analysis workflow.	201
Figure 4.12	Germline overexpression of Shep causes female sterility.	202
Figure 4.13	Germline overexpression of Shep disrupts the formation of mature eggs' dorsal appendages.	204
Figure 4.14	Shep colocalizes with Gurken protein at the dorso-anterior corner of late-stage oocytes.	206
Figure 4.15	Germline overexpression of Shep disrupts Gurken expression.	207
Figure 4.16	<i>gurken</i> mRNA levels in ovaries overexpressing Shep protein in the germline.	209
Figure 4.17	Germline overexpression of Shep disrupts <i>oskar</i> expression.	211
Figure 4.18	Germline overexpression of Shep reduces Oskar protein levels.	213
Figure 4.19	<i>oskar</i> mRNA levels in ovaries overexpressing Shep in the germline.	215
Chapter 5 – Characterizing the Composition of the Ovarian Shep mRNP Complexes:		
Figure 5.1	Computational characterization of Shep binding sites within various mRNA targets.	220
Figure 5.2	Mapping the interactions of Shep with its various mRNA targets.	223
Figure 5.3	Endogenous PTB protein co-immunoprecipitates with Shep protein.	226
Figure 5.4	PTB protein interacts with Shep in an RNA-dependent manner.	228
Figure 5.5	Differential enrichment of mRNAs bound to Shep in the ovary.	231
Figure 5.6	Analysis of RNA from the RNA immunoprecipitation experiments.	235
Chapter 6 – Discussion:		
Figure 6.1	modENCOD transcriptome data of the <i>shep</i> gene in the <i>Drosophila</i> ovary.	241
Figure 6.2	Protein expression of Shep isoforms in the <i>Drosophila</i> ovary.	243
Figure 6.3	Protein-trap insertion hints at transcriptional disruption of Shep small isoforms.	245

Figure 6.4	Model of the Delta-Notch signalling pathway regulating <i>shep</i> gene expression in the ovarian soma.	251
Figure 6.5	Single-cell transcriptome profile of <i>shep</i> in adult fly ovaries.	253
Figure 6.6	Potential binding sites for Su(H) in the <i>Drosophila shep</i> gene.	254
Figure 6.7	Translational repression of the <i>oskar</i> mRNP complex by the cytoplasmic Shep.	262
Figure 6.8	<i>shep</i> loss-of-function backgrounds dysregulated <i>oskar</i> expression but not <i>gurken</i> .	264
Figure 6.9	Putative nuclear localization signal sequence in Shep protein.	268

Table #	Title	Page N^o
Chapter 1 – Introduction:		
Table 1.1	RNA sequence motifs recognized by various RBPs containing the RNA-binding domain RRM.	25
Chapter 2 – Materials & Methods:		
Table 2.1	List of recombinant DNA construct used in this project.	65
Table 2.2	<i>Drosophila melanogaster</i> DNA clones purchased from DGRC.	66
Table 2.3	List of <i>D. melanogaster</i> non-ovarian cDNA templates used in this project.	67
Table 2.4	Formulation of Luria-Bertani media for microbial cultivation.	79
Table 2.5	Antibiotic stock & working concentrations for plasmid selection.	79
Table 2.6	attP landing sites used for site-specific integration of transgenic constructs.	83
Table 2.7	Recipe to prepare 15mL of Hypotonic Lysis buffer from stock solutions.	105
Table 2.8	Recipe for preparing resolving and stacking gels.	105
Table 2.9	List of primary & secondary antibodies used for fluorescent immunoblotting.	108
Table 2.10	Recipe to prepare 1L of cornmeal-based fly food.	120
Table 2.11	List of primary & secondary antibodies used for fluorescent immunohistochemistry.	137
Chapter 3 – Shep Expression Pattern During <i>D. melanogaster</i> Oogenesis:		
Chapter 4 – Deciphering the Role of Shep during <i>D. melanogaster</i> Oogenesis:		
Table 4.1	List of genotypes analysed for knockdown of <i>shep</i> in the female germline.	180

Table 4.2	Summary of the genotypes analyzed for the <i>shep</i> complementation test.	193
------------------	---	-----

Chapter 5 – Characterizing the Composition of the Ovarian Shep mRNP Complexes:

Table 5.1	Quality control of RIP samples to check RNA compatibility for transcriptome sequencing.	234
------------------	---	-----

Chapter 6 – Discussion:

⌘ Chapter I ⌘

Introduction

1 Chapter I – Introduction:

In the midst of the 20th century, somatic cell nuclear transfer was first attempted in amphibians (Spemann 1938; Briggs and King 1952; Gurdon and Uehlinger 1966). This process involves the transfer of the nucleus from a somatic donor cell into an enucleated oocyte to produce a viable cloned organism (Wilmut *et al.* 2015). Since then, this technique has been reproduced in other organisms, including mammals. A famous example is Dolly the sheep, who was born in 1996 (Wilmut *et al.* 1997). These studies demonstrate a key principle of eukaryotic multicellular development: genomic equivalence, which implies that all cells within a multicellular organism – with few exceptions – contain the same genetic material (Barresi and Gilbert 2020). Despite carrying the same genetic information, cells within a multicellular organism exhibit high diversity and specialization. Cells acquire their unique characteristics through differential gene expression, which accounts for cellular differences, rather than changes in the deoxyribonucleic acid (DNA) sequence of genes (Alberts *et al.* 2022). The expression of genes in a regulated manner enables cells to synthesize and accumulate different sets of ribonucleic acid (RNA) and protein molecules, thereby resulting in a variety of cellular morphologies and functions (Das *et al.* 2021).

1.1 Gene Expression Regulation

In response to developmental cues or environmental signals, cells need to orchestrate the spatiotemporal expression of proteins, a sophisticated process achieved through the regulation of gene expression (Buccitelli and Selbach 2020). Gene expression control is a complex and multifaceted process, regulated at multiple levels, beginning with transcriptional regulation in the nucleus, which dictates both the timing and quantity of the transcripts produced (Buchberger *et al.* 2019). Although gene transcription is primarily controlled by transcription factors, epigenetic marks, and chromatin topology (Tsai and Cullen 2020; Deng *et al.* 2022), RNA-binding proteins and RNA-based mechanisms have recently emerged as key players in regulating gene expression at the DNA level and also at the RNA level (Johnson and Straight 2017; Xiao *et al.* 2019; Du and Xiao 2020).

Following transcription, the fine-tuning of gene expression is regulated post-transcriptionally, not only in the nucleus but also in the cytoplasm (Glisovic *et al.* 2008). Post-transcriptional gene expression regulation is governed by RNA-binding proteins and non-coding RNAs (Briata and Gherzi 2020; Pisignano and Lodomery 2021; Statello *et al.* 2021). RNA-binding proteins, in particular, dynamically modulate mRNAs during various biological processes, including splicing, capping, and polyadenylation, which occur in the nucleus before mRNAs are exported into the cytoplasm (Dreyfuss *et al.* 2002). Once in the cytoplasm, gene expression control continues through the regulation of mRNA transport, stability, localization, and translation (Glisovic *et al.* 2008).

The asymmetric localization of mRNAs within a cell ensures the local translation of proteins in a specific region, precisely when and where they are needed (Clark *et al.* 2000). Translation is intimately linked to mRNA localization and the regulation of mRNA turnover, acting as coordinated mechanisms controlling the timing and extent of protein expression (Zhao *et al.* 2016; Aloufi *et al.* 2021). When an mRNA is no longer needed, it can be targeted for degradation within the cytoplasm to reduce protein production (Filipowicz *et al.* 2008). Conversely, RNA-stabilizing proteins can bind to mRNAs, preventing their degradation (Suresh Babu *et al.* 2015). Finally, gene expression can also be regulated at the protein level via a complex, dynamic set of processes known as post-translational regulation (Chen and Kashina 2021). Through these processes, proteins undergo extensive modifications that influence their abundance, interactions, and activity.

1.1.1 RNA – An Essential Molecule in the Regulation of Gene Expression

RNA represents a diverse and large family of nucleic acid molecules with a broad range of cellular roles (Liljas 2013; Elliott and Lodomery 2015; Santosh *et al.* 2015). Based on their cellular roles, RNAs can be categorized into two main types: RNAs that are protein-coding (i.e., mRNA) and those that are non-coding (e.g., tRNA, rRNA, ncRNA, etc.; Allison 2021). As the name suggests, mRNA molecules carry genetic instructions encoding a protein. On the other hand, non-coding RNAs do not encode for proteins, but they play structural, catalytic, and regulatory roles that ultimately affect gene

expression. RNA molecules themselves significantly contribute to fine-tuning gene expression regulation. In eukaryotic cells, three distinct nuclear RNA polymerases are responsible for the transcription of diverse functional RNA molecules (Alberts *et al.* 2022). Specifically, RNA polymerase II transcribes mRNAs from protein-coding genes, while RNA polymerases I and III handle the transcription of non-coding RNA molecules.

RNA is chemically similar to DNA at the primary sequence level, in that both are polymers made up of monomer units called nucleotides (Lodish *et al.* 2016). Despite this similarity, the nucleotides making RNA contain a ribose sugar with an additional hydroxyl group, 2'-OH, which makes them more reactive and structurally dynamic than DNA. Therefore, the primary structure of RNA is defined by four distinct fundamental building blocks: adenine (A), guanine (G), cytosine (C), and uridine (U; thymine is used instead in DNA). Nucleotides are further categorized based on the number of rings present in their nitrogen bases: purines have two-ring structures (A and G), while pyrimidines have a single-ring structure (C and U). Additionally, noncanonical bases such as pseudouridine and dihydrouridine are frequently found in RNA and not DNA (Grosjean and Benne 1998; Ge and Yu 2013).

Because of the base-pairing interactions between nucleotides, the linear sequence (i.e., primary structure) of an RNA folds in on itself, giving rise to secondary and tertiary structures. As with DNA, RNA also forms the standard Watson-Crick base pairs, with two hydrogen bonds between A and U bases, and three hydrogen bonds between C and G bases. An important difference is that unconventional, yet fundamental, base pairing can also occur naturally within RNAs, which does not follow the Watson-Crick base pair principle (Varani and McClain 2000). An example of this non-Watson-Crick base pairing is the wobble base pair, which facilitates interactions between G and U bases through two hydrogen bonds (Kuchin 2011). These unique intramolecular interactions within and between RNA molecules give rise to three-dimensional structures, such as loops and helices (Wan *et al.* 2011; Lim and Brown 2018; Duan *et al.* 2020). RNA molecules are much more like proteins than DNA in terms of their complexity and diversity of 3D structures. In fact, RNA can form versatile structures that extend to the quaternary level of organization (Jones and Ferré-D'Amaré 2015). These structural

elements of RNA are recognized by regulatory binding proteins and play a crucial role in the function, interactions, and general activity of RNA molecules (Mignone *et al.* 2002; Elliott and Lodomery 2015). Functionally, RNA extends far beyond merely being a carrier of genetic information and has important cellular roles, including catalytic and scaffolding roles (Routh and Sankaranarayanan 2017; Fernandes and Buchan 2021). The proper and intricate folding of RNA secondary structures is fundamental for both its biological function within cells and its post-transcriptional regulation (Vandivier *et al.* 2016). For example, proper folding of tRNAs is key for their interactions with ribosomes, a process that is essential for the translation of protein-coding RNAs into functional proteins (Vacher *et al.* 1984; Demeshkina *et al.* 2010; Bhaskaran *et al.* 2012; Vandivier *et al.* 2016).

Messenger RNA (mRNA) is a protein-coding RNA molecule that carries a nucleic acid sequence complementary to the DNA sequence that serves as the template for protein synthesis (Feher 2017). They play a key role in the ‘central dogma’ of molecular biology, serving as an intermediary molecule between DNA and proteins. mRNAs convey genetic information from DNA to the cellular translation machinery, thereby facilitating the synthesis of functional proteins (Liljas 2013). Unlike prokaryotic polycistronic mRNAs, eukaryotic mRNAs undergo a series of processing steps in the nucleus. These steps include splicing, the addition of a 5'-7-methylguanosine cap (5' m7G cap), and polyadenylation. Numerous studies have shown that all three forms of modifications that take place in the nucleus influence the fate of mRNA at almost every step of gene expression, from transcription to translation, as detailed in (Decker and Parker 2002; Fuke and Ohno 2008; Jacob and Smith 2017; Yamada *et al.* 2018; Passmore and Coller 2022). The resultant mature mRNA has a tripartite structure: a 5' untranslated region (5'UTR), a coding region comprised of triplet codons (each encoding an amino acid), and a 3' untranslated region (3'UTR; Mignone *et al.* 2002). **Figure 1.1** highlights these regions among other regulatory features.

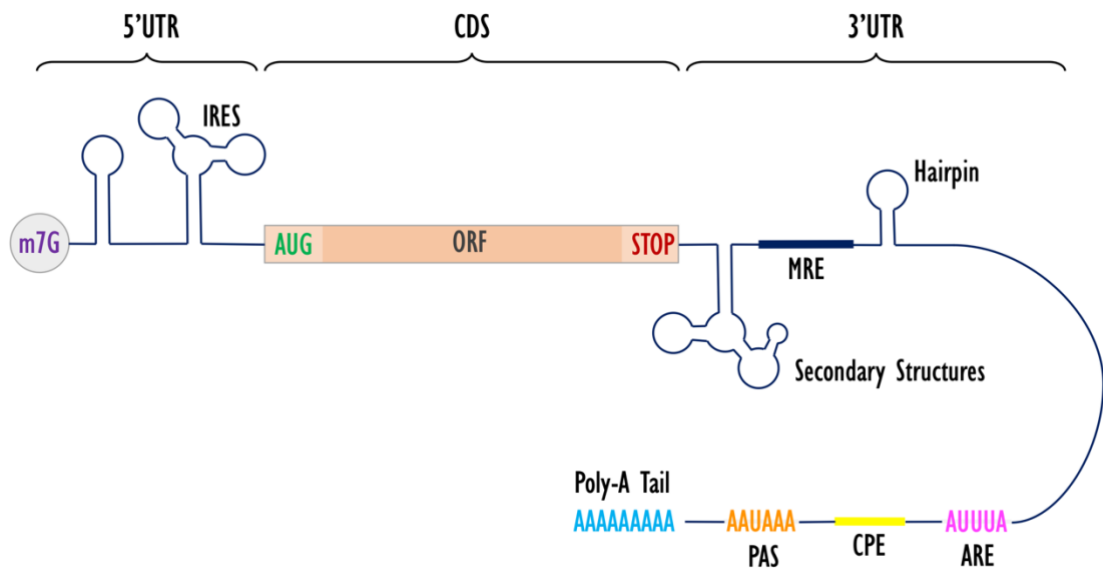


Figure 1.1: Generic features of mature eukaryotic mRNA. A schematic representation of mature eukaryotic mRNA after processing in the nucleus (i.e., after splicing, 5' capping, and polyadenylation). A mature mRNA is organised into three distinct functional regions: the 5' untranslated region (5'UTR), the coding sequence (CDS), and the 3' untranslated region (3'UTR). The diagram highlights some post-transcriptional *cis*-regulatory elements found within mRNAs that affect gene expression regulation. These elements are based on sequence or structural motifs, and in some instances a combination of both. Internal ribosome entry sites (IRES) are among the elements unique to the 5'UTR. The initiation codon (AUG), open reading frame (ORF), and stop codon are unique to the coding sequence (CDS). Meanwhile, the 3'UTR contains several *cis*-regulatory elements that regulate various aspects of mRNA metabolism. These elements include microRNA response elements (MRE), hairpin-like secondary structures, the cytoplasmic polyadenylation element (CPE), AU-rich elements (AUUUA), and polyadenylation signal (PSA, AAUAAA). The ribosome translates mRNA by first scanning the 5'UTR until it locates the AUG start codon of the open reading frame. This process can be facilitated in two ways: through a cap-dependent mechanism, which involves the recognition of a 5'-7-methylguanosine cap (5' m7G), or through an IRES-mediated mechanism. The diagram is adapted from Mignone *et al.* (2002), and Goss and Domashevskiy (2022).

1.1.2 The Role of Untranslated Regions in Defining mRNA Fate

Gene expression is finely regulated at the post-transcriptional level, predominantly through regulatory elements found in the untranslated regions of mRNAs (Mignone *et al.* 2002; Chatterjee and Pal 2009). These elements typically comprise a combination of single-stranded sequences and structural motifs (Taliaferro *et al.* 2016; Dominguez *et al.* 2018). Nucleotide sequences and structural patterns located in both the 5'UTR and 3'UTR serve as beacons for recognition and provide a platform for interaction with specific *trans*-regulatory proteins and RNAs. On average, the 5'UTR is relatively shorter

in length and has a higher GC% content than the 3'UTR (Pesole *et al.* 2002; Mazumder *et al.* 2003; Menendez-Gil and Toledo-Arana 2021). These differences in average length and nucleotide composition imply that the 3'UTR can not only harbour more regulatory elements and motifs for interaction with *trans*-acting factors but also contain more AU-rich regions, both of which influence mRNA fate. The average length difference between the untranslated regions also reflects the complexity of the post-transcriptional regulatory processes that the 3'UTR can coordinate compared with the 5'UTR. Regardless, both untranslated regions serve as a scaffold platform for interaction with regulatory *trans*-acting factors, facilitating differential gene expression at various developmental stages and in different tissues.

Generally, secondary structures formed in the 5'UTR, in particular, are thought to primarily regulate gene expression via translational control. In the cap-dependent mechanism of translation, the presence of very stable secondary structures within the 5'UTR acts as inhibitory elements that hamper translation efficiency (Mignone *et al.* 2002; Di Liegro *et al.* 2014). These structures could stall the scanning of the 40S ribosomal subunit along the mRNA from the 5' m7G-cap to the translation initiation codon of the open reading frame, thereby inhibiting translation. An alternative mechanism for initiating eukaryotic translation, independent of the 5' m7G-cap, is through internal ribosome entry sites (IRES, **Figure 1.1**; Yang and Wang 2019). These regulatory structural elements are located within the 5'UTR, upstream of the translation initiation codon, and allow the recruitment of ribosomes to bind internally and initiate translation (Jackson and Kaminski 1995; Hellen and Sarnow 2001). Among other structures, the presence of stem-loop structures, which are found in both the 5' and 3' untranslated regions of mRNA, serves as a platform for interactions with *trans*-acting factors. These structures can influence various aspects of mRNA metabolism, with specific effects depending on the nature of the interacting *trans*-acting factors and their associated complexes. Studies have shown that RNA structures are dynamically regulated in response to cellular cues to fine-tune RNA functions and expression within living cells, thereby providing cells with a rapid means of gene expression regulation (Al-Hashimi and Walter 2008; Sponer *et al.* 2018; Ganser *et al.* 2019; Chełkowska-Pauszek *et al.* 2021). This aspect further augments the complexity of gene expression

regulation by highlighting the dynamic nature of interactions between *cis*- and *trans*-acting elements. Such dynamism allows for the fine-tuning of gene expression in response to a spectrum of cellular conditions and signals.

On the other hand, the 3'UTR contains a number of *cis*-regulatory elements that control various aspects of mRNA metabolism, including AU-rich elements, microRNA response elements, cytoplasmic polyadenylation elements, stem-loop structures, and others (**Figure 1.1**). Mignone *et al.* (2002) provided a summary of some of the annotated functional *cis*-regulatory elements present in the 3'UTR (for detailed information, refer to table 6 of their paper), which further supports the notion of the complexity of the regulatory processes coordinated by the 3'UTR.

AU-rich elements (ARE) are among the most common elements present in the 3'UTR of mRNAs, influencing mRNA stability and translation in response to different extracellular and intracellular signals (Peng *et al.* 1996). These elements are stretches of sequence motifs, each characterized by a high uracil content often interspersed with adenine (Goss and Domashevskiy 2022). AREs have been experimentally grouped into three classes based on their sequence, repetition, and functional properties (Mignone *et al.* 2002; Matoulkova *et al.* 2012; Mayr 2019). Briefly, Classes I and II each contain a basic ARE motif, represented by the pentamer 'AUUUA'. In Class I, this motif appears three times, while in Class II, it appears multiple times. Class III, on the other hand, is based on a non-AUUUA motif and primarily contains U-rich sequences (Peng *et al.* 1996). Nevertheless, these AU-rich elements are responsible for mRNA destabilization and translational repression. Analysis of chimeric genes has demonstrated a positive correlation between the increased number of ARE repeats within a 3'UTR and mRNA instability (Zubiaga *et al.* 1995). The presence of these motifs is a characteristic feature of short-lived mRNAs and is also associated with mRNAs whose expression is tightly regulated (Otsuka *et al.* 2019; Ripin *et al.* 2019). mRNAs can be stabilized, leading to enhanced gene expression by masking the AREs from destabilizing *trans*-regulatory elements. This stabilization can occur when *trans*-regulatory elements mask AREs, thereby protecting them from interactions that might otherwise destabilize the mRNA molecule. Initially, AU-rich elements were considered

merely mRNA decay elements (Chen and Shyu 1995). However, subsequent research by Krays *et al.* (1989) and others demonstrated that they also play a role in repressing translation (García-Mauriño *et al.* 2017; Otsuka *et al.* 2019).

Cytoplasmic polyadenylation elements (CPEs) are *cis*-regulatory elements in the 3'UTR (Figure 1.1), which offer an alternative means to regulate mRNAs (Moore and von Lindern 2018). CPEs are composed of specific nucleotide sequences that can regulate mRNA translation by controlling the length of the poly(A) tail (Villalba *et al.* 2011). These *cis*-acting elements are involved in cytoplasmic polyadenylation, a process by which dormant and translationally inactive mRNAs become activated via the elongation of their poly(A) tails in the cytoplasm (Rouhana *et al.* 2023). This process is regulated by the conserved cytoplasmic polyadenylation element binding (CPEB) protein family (Chen *et al.* 2016). This mechanism is particularly important in certain biological contexts such as oocyte maturation and neuronal activity (Benoit *et al.* 2008; Kozlov *et al.* 2021), where the translation of specific mRNAs is tightly controlled.

1.1.3 Role of RNA-Binding Proteins in Gene Expression Regulation

RNA-binding proteins (RBPs), which constitute a large class of proteins, sit at the heart of co-transcriptional and post-transcriptional regulation of gene expression (Glisovic *et al.* 2008; Van Assche *et al.* 2015; Prall *et al.* 2019). As *trans*-acting factors, they recognize and interact with *cis*-regulatory elements within mRNA molecules, thereby directly influencing the fate of mRNA by mediating events such as splicing, RNA folding, base modification, transportation, subcellular localization, turnover, and translation efficiency (Figure 1.2; Hentze *et al.* 2018). RNA-binding proteins dynamically interact with RNAs both in the nucleus and cytoplasm, regulating various aspects of RNA metabolism, from affecting the splicing of nascent transcripts to modulating their translation (Masuda *et al.* 2021). In fact, RBPs associate co-transcriptionally with newly transcribed RNAs. In other words, as soon as the RNA is transcribed, RBPs bind to distinct sites of the nascent transcript to assemble the ribonucleoprotein (RNP) complex.

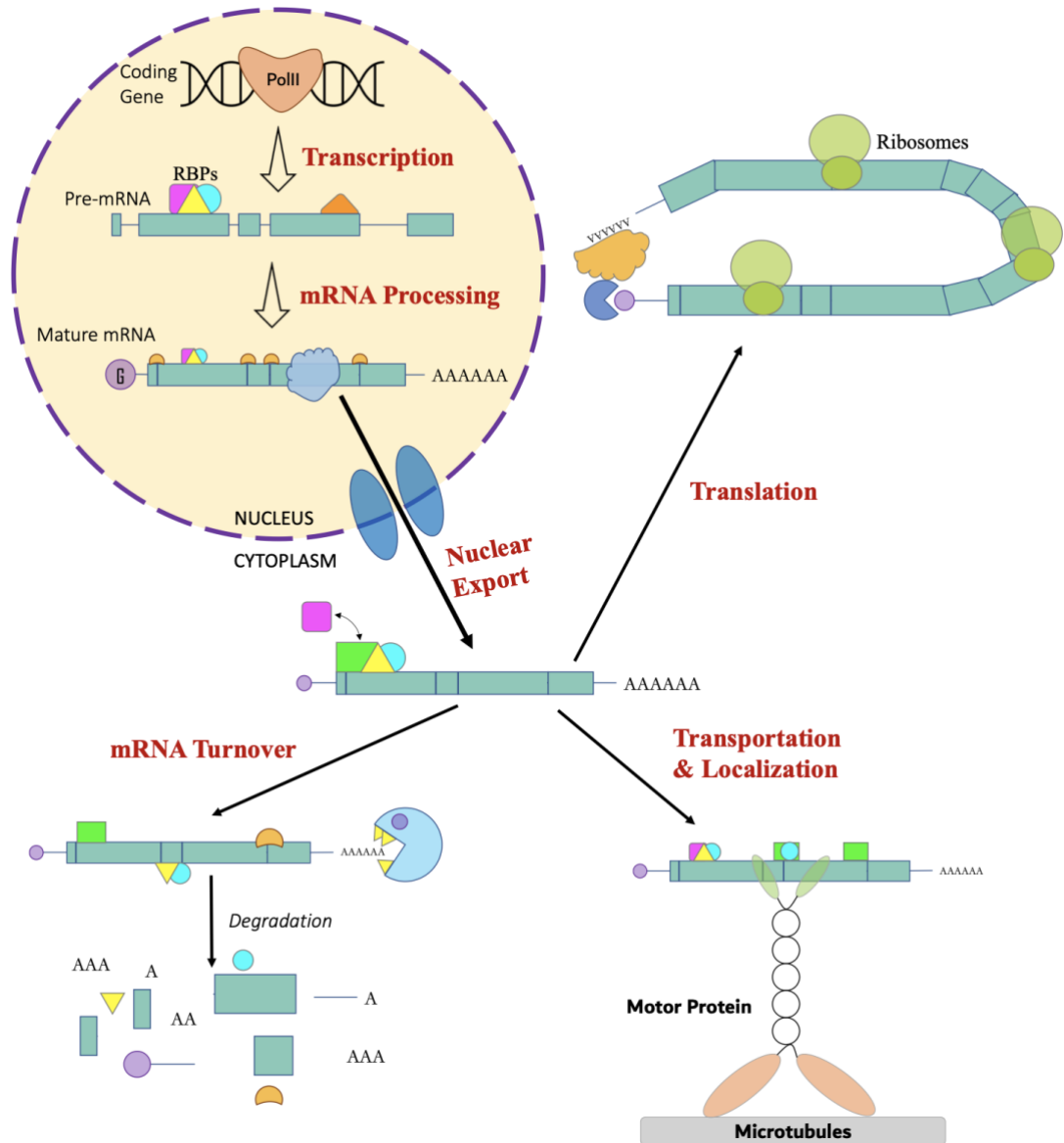


Figure 1.2: Overview of post-transcriptional regulatory processes in eukaryotic cells. While transcription is the most well-studied area of gene expression regulation, various post-transcriptional mechanisms also play an essential role, not only in fine-tuning gene expression but also in controlling the spatiotemporal expression of proteins (Glisovic *et al.* 2008; Ganser *et al.* 2019). Gene expression regulation at the RNA level is mediated by RNA-binding proteins (RBPs) and non-coding RNAs. In particular, RBPs associate with RNA molecules starting from their biogenesis in the nucleus, forming a ribonucleoprotein (RNP) complex unique to each transcript. In the nucleus, the nascent mRNA (i.e., pre-mRNA) produced after transcription undergoes a series of processing steps, including 5' capping, splicing, and 3' polyadenylation, which define its fate in the cytoplasm. After processing in the nucleus, mature mRNA is exported to the cytoplasm, where RBPs within mRNA complexes undergo extensive and dynamic remodelling to influence the cytoplasmic fate of the mRNA. Among the post-transcriptional processes that RBPs regulate are mRNA stability, transportation, localization, and translation.

RNP complexes are often very complex in composition and are dynamic in nature (Gebauer *et al.* 2012). They are assembled by recruiting various proteins and non-

coding RNAs through the *cis*-regulatory elements in the mRNA or by protein-protein interactions (Müller-Mcnicoll and Neugebauer 2013). It is common for different RBPs within the same RNP complex to function redundantly, ensuring proper regulation even in the absence of one or several RBPs. An example of this is the *Drosophila* mRNA *oskar*, whose translational control is regulated by the RNA-binding proteins Bruno, PTB, Hrp48, and others (Kim-Ha *et al.* 1995; Johnstone and Lasko 2001; Yano *et al.* 2004; Besse *et al.* 2009; Bayer *et al.* 2023). Such mechanism of functional redundancy becomes particularly important for mRNAs that are essential for development or those requiring tight regulation.

The composition of RNP complexes consistently undergoes remodelling in response to various cues, reflecting the dynamic nature of RNA processing and regulation. It was traditionally believed that RNA-binding proteins, which control the cytoplasmic localization and translation of mRNAs, only associate with RNP complexes after the mRNA has been exported from the nucleus. However, accumulating evidence has challenge this view, suggesting that RBP interactions or RNP complexes assembled in the nucleus can influence the fate of the mRNA in the cytoplasm, even when the RBP is no longer bound (Matsumoto *et al.* 1998; Farina and Singer 2002; Goler-Baron *et al.* 2008; Trcek and Singer 2010). This gives rise to the phenomenon known as ‘RNP nuclear history’. In the case of *Drosophila* oogenesis, transgenic versions of the *oskar* mRNA lacking one of its three introns or a combination thereof have demonstrated that the intronless *oskar* mRNA is not only mislocalized but also translated less efficiently compared to the wild-type *oskar* mRNA (Hachet and Ephrussi 2004). This shows that splicing can lead to the assembly of different mRNA–protein complexes with diverse cytoplasmic localization patterns and translational efficiency.

Bioinformatic analyses suggest that the eukaryotic genome encodes many predicted RBPs (Liao *et al.* 2020). Unexpectedly, only approximately 8% of yeast and around 2% of *C. elegans*, *D. melanogaster*, and *M. musculus* genomes have been annotated as genes encoding RBPs (Kerner *et al.* 2011). Furthermore, 7.5% of annotated human protein-coding genes have been assigned as RBPs (Weiße *et al.* 2020). Given the complexity of post-transcriptional regulation that eukaryotic mRNAs require, the actual number of RBPs in multicellular eukaryotes is believed to be much higher than that

currently identified, as there may be others yet undiscovered. Interestingly, proteome-wide discovery efforts, combined with other mathematical tools, predict thousands of RBP candidates, many of which lack canonical RNA-binding domains (Jin *et al.* 2023).

Eukaryotic transcripts undergo comprehensive regulation to fine-tune gene expression, involving transcription in the nucleus, intricate processing, alternative splicing, and the necessity for precise temporal and spatial control before their export to the cytoplasm (**Figure 1.2**). Remarkably, over 70% of the expressed transcripts analyzed were subcellularly localized in *Drosophila* embryos, as demonstrated by Lécuyer *et al.* (2007). Additional studies have shown that hundreds to thousands of mRNAs are indeed localized within distinct cellular compartments or regions in the cytoplasm (Medioni *et al.* 2012). Also, such asymmetric subcellular localization pattern has been observed in a wide range of organisms outside the animal kingdom, including plants and fungi (Medioni *et al.* 2012). This prevalent phenomenon highlights the necessity of extensive and tight post-transcriptional regulation required by eukaryotic cells to achieve differential and spatiotemporal expression. Despite their central role in RNA function, the RNA-binding specificity of most RBPs remain elusive or poorly understood.

RNA-binding proteins are modular in structure and acquire their RNA-binding capabilities through the presence of various functional units known as RNA-binding domains (RBDs; Corley *et al.* 2020). Such domains, responsible for binding RNA, include the RNA recognition motifs (RRM), the K-homology (KH) domain, the zinc-finger (ZnF) domain, the double-stranded RNA-binding domain (dsRBD), and many more (Lunde *et al.* 2007). The specificity of these proteins for their diverse substrates arises from the number, type, and relative arrangement of the RNA-binding domains (Dreyfuss *et al.* 2002). RNA-binding proteins, through their RNA-binding domains, have the ability to bind to both single-stranded and double-stranded regions of RNA. Additionally, these proteins can recognize motifs primarily based on the ribonucleic acid sequence, commonly referred to as the primary structure. Alternatively, they can identify specific geometric 3D structures formed due to RNA folding, or possibly a combination of both.

RNA Recognition Motif: Numerous RNA-binding domains have been identified (Anantharaman *et al.* 2002); however, the RNA recognition motif stands out as the best characterised and most common RNA-binding domain, being found in more than 50% of RBPs (Kerner *et al.* 2011; Tsai *et al.* 2014). A typical RRM domain consists of 80–90 amino acids that form four-stranded anti-parallel beta-sheets with two alpha-helices, giving rise to a barrel-like topology (Maris *et al.* 2005). The RRM folds into an alpha-beta sandwich structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (Oubridge *et al.* 1994; Lunde *et al.* 2007). The structural arrangement of the domain creates a surface that allows for RNA interaction. Two highly conserved motifs are found within the RRM that facilitate RNA interaction, specifically the RNP1 (octamer) and RNP2 (hexamer) motifs, which are present in the 3rd and 1st beta-sheets, respectively (Nagai *et al.* 1990; Maris *et al.* 2005; Cléry *et al.* 2008). RNA recognition by the RRM domain often occurs on the surface of beta-sheets through stacking and ionic interactions with the single-stranded regions of the target RNA.

Many RNA-binding proteins contain multiple copies of RRM domains, and it is unclear how each RRM contributes to the binding specificity of the entire protein (SenGupta 2013). However, using multiple RRMs at varying distances between one another can provide additional versatility in sequence recognition and binding dynamics with the target RNA molecule (Singh and Valcárcel 2005; Lunde *et al.* 2007). Typically, a single RRM domain can recognize four to eight specific ribonucleotides in a single-stranded RNA conformation, depending on the actual amino acid sequence (Auweter *et al.* 2006). The discrepancy in RRM-binding preferences towards specific sequence motifs varies not only between different proteins but also within the same protein; this is demonstrated in **Table 1.1**. Strikingly, recent studies have demonstrated that RRMs are not only involved in RNA recognition but also in protein–protein interactions (Maris *et al.* 2005). An example of such interactions is seen in the Y14–Mago complex in *Drosophila*; the heterodimerization of these proteins involves interactions between the beta-sheet surface of the Y14 RRM domain and the alpha-helical surface of the Mago protein (Fribourg *et al.* 2003). As a result, this protein-protein interaction masks key protruding amino acid residues at the interacting interface that are required for RNA interaction, thus preventing Y14 from binding to its target RNAs. This unique feature

of some RRM domains adds an additional layer of regulation, allowing for the fine-tuning of gene expression by coordinating the activity and binding dynamics of RBPs within an RNP complex.

Table 1.1: RNA sequence motifs recognized by various RBPs containing the RNA-binding domain RRM. Abbreviations – RRM: RNA recognition motif. N: stands for any nucleotide. Table adapted from Elliott and Ladomery (2015).

RRM-containing protein(s)	RRM Domain	Target Sequence
Sxl	RRM1	UUUUUUU
	RRM2	UGU
hnRNP A1	RRM1	UAGG
	RRM2	UUAGG
PABP	RRM1	AAAA
	RRM2	
HuD	RRM1	UUAUUU
	RRM2	UU
	RRM3	UAU
PTB	RRM1	UCNU
	RRM2	CNUNN
	RRM3	UCNU
	RRM4	UCNN
Shep	RRM1	Uncharacterised yet
	RRM2	

1.2 Drosophila Fruit Fly as an Animal Model System for Development

1.2.1 Benefits of Using *Drosophila* as a Model Organism

The fruit fly *Drosophila melanogaster* is a versatile model organism that has been used for over a century to study a diverse range of biological processes, from fundamental genetics to development and disease (Morgan 1910). There are myriad merits of using this dipteran insect over vertebrate models, making *D. melanogaster* indispensable for basic research (Jennings 2011). Most importantly, the fruit fly genome is roughly 60% homologous to that of humans and exhibits less genetic redundancy (Wolf and

Rockman 2011; Bellen and Yamamoto 2015). In addition, approximately 75% of the genes responsible for human diseases have functional homologs in *Drosophila* (Reiter *et al.* 2001). Other advantages of using *Drosophila* include the accumulated wealth of knowledge over time, a short life cycle, the ability to produce large numbers of offspring, and the ease with which they can be genetically manipulated through various ways (Jennings 2011). Another benefit of using fruit flies is their small size, approximately 3 mm long, which contributes to the ease with which they can be cultured and maintained inexpensively in laboratory conditions (Kue and Kumar 2023).

1.2.2 Overview of *D. melanogaster* Development

The *Drosophila* life cycle is relatively short, taking around 10 days to progress from fertilized eggs to sexually active adult flies (Fernández-Moreno *et al.* 2007). This not only means that it is easy to raise a large number of offspring for experimentation but also implies an accelerated research timeline. Genetic experiments that may take months or even years in other vertebrate models, such as mice or zebrafish, can be completed in a matter of weeks when using fruit flies (Taormina *et al.* 2019).

Drosophila is a holometabolous insect, and its life cycle can be divided into four morphologically distinct stages: embryo, larva, pupa, and adult, as depicted in **Figure 1.3** (Kue and Kumar 2023). After mating, adult female fruit flies store sperm in a specialized organs called the spermatheca and seminal receptacle, which are both part of the female reproductive tract of insects (Klowden 2009; Schnakenberg *et al.* 2011). This storage enables the fertilization of hundreds of eggs to be laid over the next few days following mating (Jennings 2011). Egg-to-adult development in *Drosophila* is sensitive to ambient temperature; cold temperatures around 18°C slow progression through its life cycle, while warmer temperatures near 29°C have the opposite effect on fly development (Pulver and Berni 2012). At an ambient temperature of approximately 25°C, the life cycle is completed within roughly 10 days from mating, starting with embryonic development that lasts around 24 hours before the larvae hatch (**Figure 1.3**; Wolpert *et al.* 2019). Within two days of hatching, the worm-like 1st instar larvae moult twice, going through different stages of larval development: transitioning first to the 2nd and then to the 3rd instar larva stages (Bodenstein 1944). The 3rd instar larvae then

continue feeding and foraging for additional two days before migrating away from the food source to begin pupation (Figure 1.3). After roughly 5 days post-fertilization, the prepupa undergoes pupation for a duration of 3-4 days. After this, the fully metamorphosed adult flies eclose from their pupal case and reach sexual maturity within approximately 8 hours post eclosion (Strömnæs and Kvelland 1962). The lifespan of fruit flies can vary depending on various factors, including sex, genotype, temperature, and diet. Despite these age-influencing factors, the typical lifespan under standard laboratory conditions ranges from 60 to 90 days (Helfand and Rogina 2003; Sun *et al.* 2013). This makes *Drosophila* a convenient model organism for genetic and developmental studies.

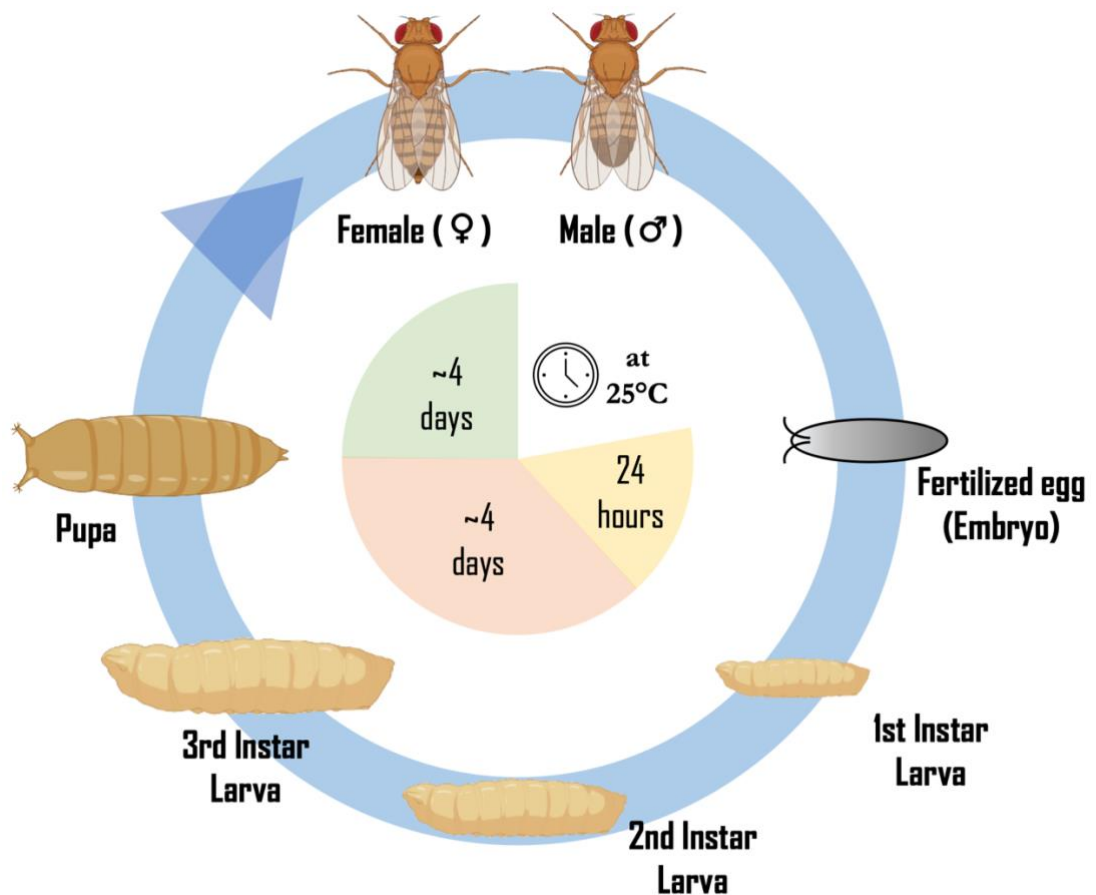


Figure 1.3: *Drosophila melanogaster* life cycle. After adult female and male flies reach sexual maturity and mate, fertilized eggs are laid. At an ambient temperature of 25°C, these eggs then undergo embryogenesis, developing into first instar larvae in approximately 24 hours. After hatching, the 1st instar larvae then grow and progress through two additional stages: the 2nd instar lasts for 1 day, and the 3rd instar lasts roughly 2 days. During these larval stages, *D. melanogaster* larva feeds, accumulating energy that is needed for metamorphosis. At the end of the larval stage, which lasts roughly 4 days, the larvae pupate. During the pupal stage, which lasts for approximately 4 days, adult fly tissues form through metamorphosis. During the pupal stage, metamorphosis

occurs, leading to the formation of adult fly tissues. At the end of the metamorphosis, which lasts approximately 4 days, virgin adult flies emerge from the puparium (also known as the pupal case). After the eclosion of adult flies, their lifespan ranges from 60 to 90 days, depending on the fly's genetic background and rearing conditions (such as temperature, diet, etc.). The diagram was adapted from Wolpert *et al.* (2019) and others (Pulver and Berni 2012; Ong *et al.* 2015; Brischigliaro *et al.* 2023), reconstructed using BioRender.

1.2.3 Ovaries are a Powerful Model Organ to Study the Function of RNA-Binding Proteins

Among the various organs and tissues in *Drosophila*, the female reproductive organ stands out as particularly compelling model for investigating the cellular roles of RNA-binding proteins. Like other bilaterally symmetrical animals, *Drosophila* has an anterior-posterior axis and a dorsal-ventral axis (Lynch and Roth 2011). These axes are determined early in development, specifically during oogenesis. The establishment of fly body axes is influenced by the asymmetric localization of maternally loaded mRNAs and proteins to distinct subcellular regions within the cytoplasm of the developing, unfertilized egg (Roth and Lynch 2009). This in turn provides the egg with positional information along both its future body axes prior to fertilization. The subcellular localization and expression of these determining factors are orchestrated by an intricate network of RBPs, as detailed later in this chapter (See Sections 1.4.5 and 1.4.6).

The *Drosophila* egg develops as part of an intracellularly-connected functional syncytium and is primarily transcriptionally quiescent during its development (Lasko 2012). Consequently, the egg's development is supported by its cellular siblings, a topic that will be discussed in greater detail later in this chapter (refer to Sections 1.4.2 and 1.4.3). RNA-binding proteins therefore play an integral role during oocyte development, regulating the post-transcriptional fate of mRNAs transcribed by the nuclei of these sibling cells, which are then transported and deposited in the egg. In summary, the *Drosophila melanogaster* ovary serves as a valuable model system for elucidating the molecular mechanisms that control the spatiotemporal regulation of gene expression at the RNA level.

1.3 The *Drosophila* Toolkit for Genetic Manipulation

Drosophila, as a model system for fundamental research, offers a genetic toolkit that is more extensive and versatile than those of many other model organisms (Schäfer and

Jäckle 2000; Larkin *et al.* 2020). While the literature offers a comprehensive overview of the vast array of genetic tools and techniques for studying gene function in *Drosophila* (Bratu and McNeil 2015; Hales *et al.* 2015; Şentürk and Bellen 2018; Fischer *et al.* 2023), here we will spotlight two specific systems. These were chosen because they are the primary genetic tools used throughout this study to investigate gene function, expression pattern, and subcellular distribution in *D. melanogaster* oogenesis. The UAS-Gal4 system is designed to induce gene expression in a spatiotemporal manner, while the protein-trap system is used for fluorescently tagging endogenous proteins.

1.3.1 The UAS-Gal4 System

In *Saccharomyces cerevisiae*, also known as baker's yeast, the metabolism of galactose through the glycolytic pathway depends on the galactose-responsive transcription factor Gal4 (Laughon and Gesteland 1982; Traven *et al.* 2006). This positive transcriptional regulatory protein binds to and operates through the upstream activating sequence (UAS) *cis*-acting elements that are found upstream of galactose-inducible gene promoters (Guarente *et al.* 1982; Bram *et al.* 1986). When galactose is present, Gal4-mediated interactions control the expression of the GAL regulon, enabling the organism to metabolize and grow on galactose (Rajeshkannan *et al.* 2022). Given that the yeast transcriptional activator Gal4 has no ortholog in *Drosophila* and that the fruit fly genome lacks endogenous UAS elements, the adoption of the UAS-Gal4 system in *D. melanogaster* became possible as a means for Gal4-inducible and controlled gene expression (Kramer and Staveley 2003; Liu and Lehmann 2008).

In 1993, Brand and Perrimon showed the ability of Gal4 to activate transcription from the UAS element in *Drosophila* (Brand and Perrimon 1993), demonstrating the potential of the Gal4-UAS system for controlling the expression of any transgene of interest in a tissue-specific and temporally controlled manner. Ever since its adoption, the UAS-Gal4 system has become the most commonly used genetic tool in *Drosophila* for driving the expression of genes of interest (Caygill and Brand 2016). Additionally, it has been widely employed for various applications, including the expression of fluorescently tagged proteins, gene knockdown by RNAi, and ectopic overexpression of transgenes.

In *Drosophila*, the Gal4-UAS system has been adopted and successfully operates as a bipartite expression system (**Figure 1.4**). In this two-part system, the expression of the Gal4 transcription factor is driven under the control of a tissue-specific or ubiquitous promoter, while the UAS enhancer elements are located upstream of a particular transgene. Transgenic flies are generated to harbour either the tissue-specific Gal4 driver or the UAS construct. Flies carrying only the Gal4 transgene (i.e., Gal4 driver stocks) will express the Gal4 transcription factor depending on the promoter controlling the transgene. On the other hand, fly stocks with only the UAS construct will remain transcriptionally inactive because the circuit is not complete, thus the UAS-associated transgene is not expressed (**Figure 1.4**). However, when the flies containing the UAS transgene are crossed with a tissue-specific Gal4 driver, the resulting progeny inherit both parts of the bipartite system, thereby completing it. This in turn activates the transcription of the UAS-driven transgene by the Gal4 transcriptional regulator, allowing for targeted gene expression in a tissue-specific manner (**Figure 1.4**). Moreover, it is worth noting that the Gal4-induction of the transgene of interest in this binary expression system is temperature dependent (Schinko *et al.* 2010). Therefore, Gal4 activity can be easily manipulated and is enhanced in flies raised at 29°C. Conversely, the efficiency of the Gal4-UAS system can be downregulated when the flies are maintained at lower temperatures (Zhang *et al.* 2020).

Transgenic flies carrying the Gal4 driver or the pUAS construct are generated by germline microinjection and transformation (Bachmann and Knust 2008). Technically, the transgene carried by the pUAS expression vector is physically introduced into the posterior end of the pre-blastoderm *Drosophila* embryo, where germline cells are formed. This ensures that the recombinant DNA gets incorporated into the fly genome, either via random P-element-mediated integration or site-directed integration (Handler and Harrell-Ii 2001; Bateman *et al.* 2006; Bassett *et al.* 2013; Majumdar and Rio 2015).

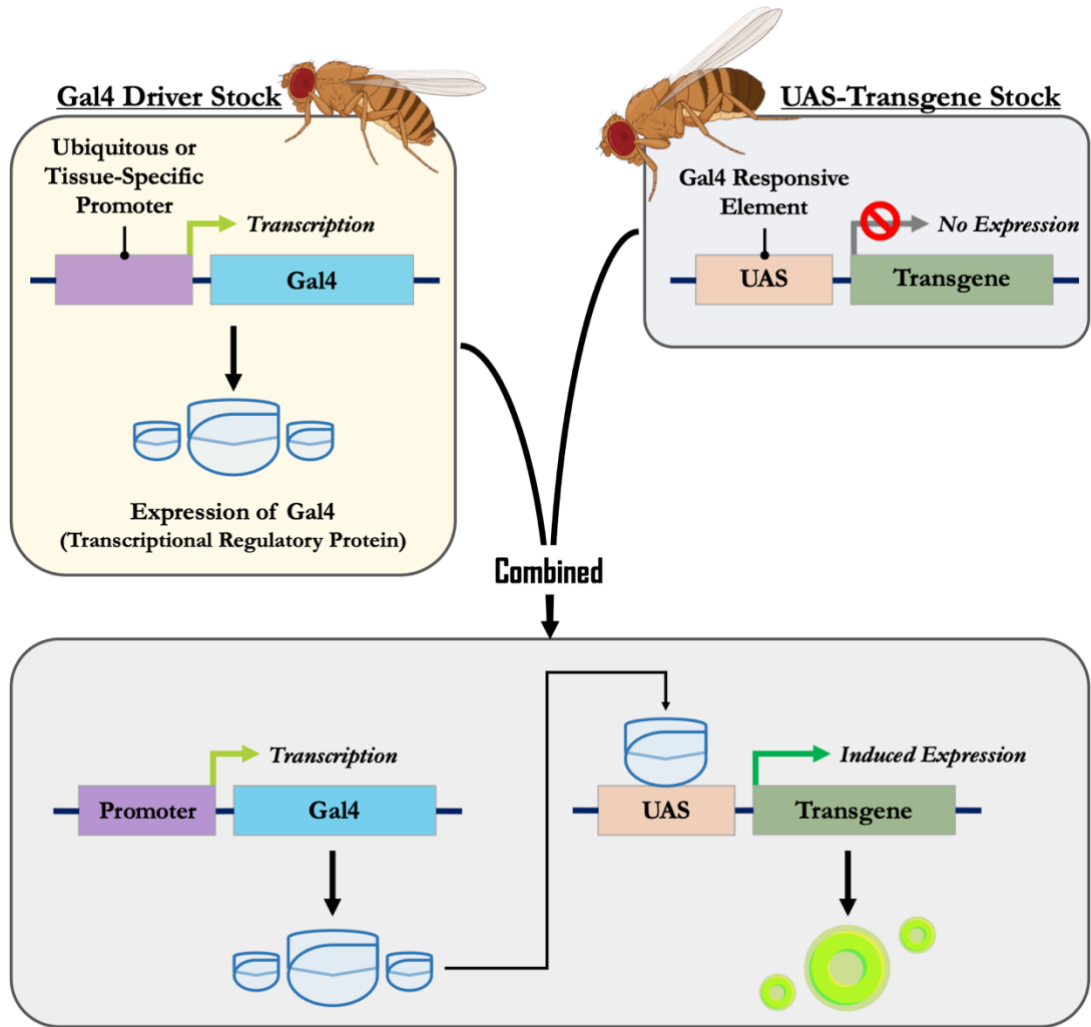


Figure 1.4: The UAS-Gal4 system induces spatiotemporal expression of transgenes in *Drosophila*. Typically, flies carrying either a ubiquitous or tissue-specific Gal4 driver are generated and maintained separately from those harbouring UAS-target transgene constructs. In the absence of the Gal4 transcriptional activator, there is no expression of the transgene of interest. However, when flies carrying the Gal4 driver are crossed with flies containing the UAS-transgene, the UAS/Gal4 bipartite system is complete and becomes operational. In the resulting progeny, transgene expression is restricted to tissues defined by the promoter driving Gal4 expression (Brand and Perrimon 1993).

By using the UAS-Gal4 system, researchers can specifically target the expression of any transgene of interest to precise tissues or even cells. When combined with the temperature-sensitive Gal80 (described below), this system enables the temporal control of gene expression, allowing researchers to activate or suppress genes at designated developmental stages or time points. In this study, we harnessed the versatility of this system for both loss-of-function and gain-of-function analyses. For the loss-of-function analysis, we employed RNAi constructs expressing interfering short

hairpin RNAs (shRNAs) that target specific mRNAs for degradation, effectively reducing the levels of the protein under study. Conversely, in the gain-of-function analysis, we used the UAS-Gal4 system to overexpress particular transgenes, especially to investigate the impact of increasing the levels of the RNA-binding proteins of interest during oogenesis.

Female Germline-Inducible UAS System: Since the introduction of the UAS-Gal4 system (Brand and Perrimon 1993), manipulating gene expression during *Drosophila* oogenesis and early embryogenesis proved to be more challenging than in somatic tissues. The original and widely used pUAS_t plasmid for Gal4-inducible expression of transgenes in flies was not well expressed in the female germline (Brand and Perrimon 1993). In 1998, a modified version of the original vector, known as pUAS_p, was reconstructed to address this limitation, allowing maternal germline expression (Rørth 1998).

To achieve efficient expression of transgenes in the female germline, Rørth introduced several modifications to the number of UAS elements, the promoter, 5'UTR, and the 3'UTR. Instead of the previously used *hsp70*-derived core promoter, the pUAS_p vector uses a germline-compatible promoter, the *P-transposase* promoter (Rørth 1998). The *P-transposase* is a transposon that is naturally active in the female germline (Simmons *et al.* 2002; Deluca and Spradling 2018). Additionally, 14 repeats of the UAS site and two copies of the GAGA motif were used, all of which were placed upstream of the promoter to further enhance the efficiency of germline expression (Rørth 1996; Rørth 1998; Chopra *et al.* 2008). Finally, downstream of the multiple cloning site where the transgene would be inserted, the 3'UTR and terminator regions of the *K10* gene were used to stabilize maternal transcripts (Rørth 1998; Duffy 2002). **Figure 1.5** shows the key differences between the pUAS_t and pUAS_p expression vectors.

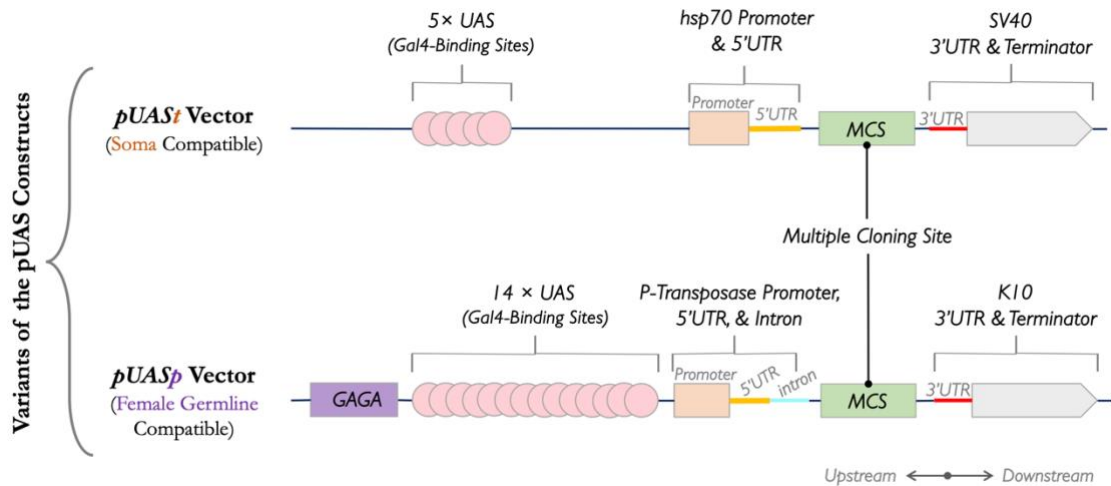


Figure 1.5: Comparison of the pUAS_t and pUAS_p expression vectors. Both vectors are tailored for the UAS/Gal4 system in *Drosophila* (Brand and Perrimon 1993; Rorth 1998). Although pUAS_t facilitates expression across various tissues, pUAS_p is optimized for female germline tissue expression. The vectors differ in the number of UAS enhancer sequences located upstream of the promoter that is used to drive the expression of the target transgene cloned into the multiple cloning site (MCS). Additionally, the pUAS_t employs the *hsp70* core promoter and untranslated regions, whereas the pUAS_p vector uses the *P-transposase* promoter regulatory sequence. Other modifications made to pUAS_p include the addition of the GAGA site upstream of the UAS site, as well as the use of the K10 3'UTR and terminator region downstream of the MCS, instead of the SV40 terminator region found downstream of the MCS in the pUAS_t.

The Gal4–Gal80 Regulatory Axis: The UAS-Gal4 expression system can be spatially regulated depending on the specific promoter used. However, to temporally control the Gal4 transcriptional activity and ultimately the expression of the UAS-coupled transgene, irrespective of the promoter used to drive Gal4, a simple yet powerful component of the yeast GAL regulon is utilised: the Gal80 protein, an inhibitor of the Gal4 transcription factor (Lue *et al.* 1987). When expressed in flies, Gal80 can be used in conjunction with the UAS-Gal4 system to achieve repressible expression of the transgene (Ma and Ptashne 1987). Specifically, Gal80 antagonizes the transcriptional activity of Gal4 by binding to its activation domain, thus preventing interaction between Gal4 and the transcriptional machinery. Different variations of the Gal80 transgene have been engineered to provide the ability to control the timing of its expression, including those designed for systems that are drug-inducible or temperature-sensitive. In the former system, the expression of Gal80 is induced by the presence of either tetracycline or auxin in the diet (Barwell *et al.* 2017; McClure *et al.* 2022). In the temperature-sensitive system, however, Gal80 becomes unstable at high temperatures, and its ability to repress Gal4-transcriptional activity is relieved upon a temperature shift

to 29°C (Matsumoto *et al.* 1978; Barwell *et al.* 2023). The ability to control gene expression with temporal resolution is important when gene manipulation results in embryonic lethality or arrests the development of organs and tissues.

1.3.2 Protein-Trap as a Tool for the Characterization of Gene Expression & Protein Distribution

The GFP-based protein-trap system is another powerful and invaluable tool that allows for the *in vivo* visualization of proteins by inserting a GFP cassette into genes at random locations (Morin *et al.* 2001). This approach was originally introduced by Smith (1996) in cell lines, whereby an epitope was used instead of GFP, and the tagged proteins were detected by immunofluorescence using specific monoclonal antibodies recognizing the epitope. Around the same time, progress was made to adopt a similar approach for use in living organisms, including the haploid unicellular alga *Chlamydomonas reinhardtii* and the metazoan *Drosophila melanogaster* (Jarvik *et al.* 1996). The protein-trap was later adapted in many other organisms using GFP as a reporter tag, thus eliminating the need for antibodies to detect the intracellular distribution pattern of the trapped proteins (Cutler *et al.* 2000; Ding *et al.* 2000; Misawa *et al.* 2000). Unlike other trap strategies, such as enhancer-trap or gene-trap, this method not only serves as a readout of transcription but also allows the detection of the distribution of tagged proteins (Fedorova and Dorogova 2020). Visualizing the spatiotemporal subcellular localization pattern of trapped proteins is essential for understanding and deciphering their gene functions.

The principle behind the GFP protein-trap system is to generate flies in which the full-length endogenous protein is expressed as a GFP-tagged protein (**Figure 1.6**). The expression of this fusion protein is under the control of the endogenous gene promoter, and is regulated by native *cis*-regulatory elements at both the transcriptional and post-transcriptional levels (**Figure 1.6**; Morin *et al.* 2001). Thus, the detection of GFP fluorescence reflects the dynamics of both the gene expression pattern and protein distribution within a cell. Protein-trap flies are generated using a mobile transposable element that carries the coding sequence of GFP lacking both the translation initiation and stop codons (Morin *et al.* 2001). Additionally, the GFP protein-trap cassette is flanked by splice acceptor and donor sequences (**Figure 1.6A**). Upon integration within

an intron of a protein-coding gene, the protein-trap cassette acts as an artificial exon, encoding GFP (Fedorova and Dorogova 2020). Because of the splice acceptor and splice donor sequences on either side of the GFP open reading frame, the GFP artificial exon will be spliced into the mature mRNA and ultimately translated as part of the endogenous protein if in frame (**Figure 1.6B**). The resulting endogenous GFP-fused protein retains the spatiotemporal localization characteristics of the wild-type protein, unless the GFP insertion disrupts the domains or conformation necessary for subcellular localization. Notably, depending on the location of the GFP-cassette insertion within the gene, not all isoforms potentially produced by the gene may be trapped with GFP (Quiñones-Coello *et al.* 2007), revealing isoform-specific expression patterns. Regardless of the trapped isoforms, the fusion of the GFP-reporter serves as an excellent tool to study the cellular and subcellular localization of the tagged proteins. A potential limitation is that GFP may introduce a delay in the detection of rapid turnover changes in protein expression levels caused by the added folding time of the GFP tag.

In addition to being an excellent tool for assessing gene expression in tissues and protein subcellular localization, the GFP protein trap can be used for numerous applications far beyond visualization. For example, they have the potential to be used in various biochemical assays, functional analyses, and screening for particular subcellular distribution patterns. In fact, Besse *et al.* (2009) used *Drosophila* protein trap lines to screen for a novel protein involved in *oskar* mRNA regulation. Analysing the subcellular localization of protein-traps to distinct regions or structures within a cell has been shown to provide insights into the potential biological functions and interactions a protein of interest may participate in (Morin *et al.* 2001; Buszczak *et al.* 2007; Lowe *et al.* 2014). Additionally, many reagents have been developed over the years to utilize the GFP-tag as a bait for immunoprecipitation (ChromoTek GFP-Trap) in order to isolate proteins or RNP complexes and investigate protein-protein or protein-RNA interactions, respectively.

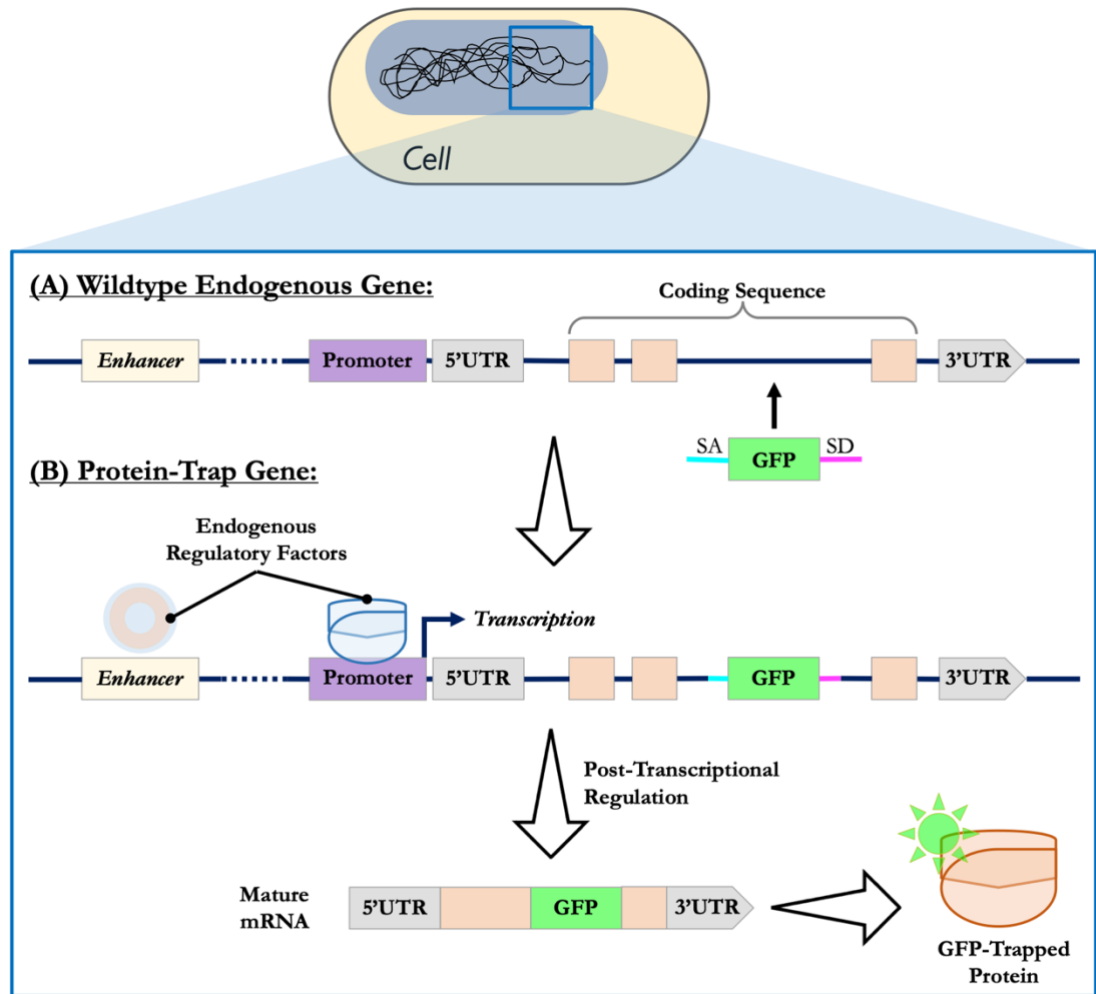


Figure 1.6: The principle of GFP protein-traps for tagging endogenous proteins. (A) Schematic depiction of a hypothetical endogenous protein-coding gene prior to the insertion of the protein-trap cassette. The enhancer is shown in yellow, the promoter in purple, untranslated exons in gray, and exons containing the protein’s coding sequence in orange. The GFP-containing mobile cassette consists of the coding sequence of the green fluorescent protein (depicted in green), a splice acceptor sequence at the 5' end (SA, shown in cyan), and a splice donor sequence at the 3' end (SD, illustrated in magenta). **(B)** Endogenous gene following the insertion of the protein-trap cassette. Once integrated within an intron (represented as lines) of a protein-coding gene, the cassette acts as an artificial exon. Because GFP is inserted into the endogenous locus, it is expressed and subjected to both transcriptional and post-transcriptional regulation, similar to the mechanisms controlling the wild-type gene. When the trapped gene is expressed in a specific cell type and at a particular time by endogenous *cis*- and *trans*-regulatory factors, GFP is transcribed as part of the endogenous gene and remains part of the mature mRNA following RNA processing in the nucleus. In the cytoplasm, the resulting mRNA undergoes post-transcriptional regulation to control its subcellular localization, turnover, and translation. When the GFP coding sequence is in frame, mRNA translation by ribosomes results in a protein fused with a GFP tag. This allows *in vivo* real-time visualization of the endogenous GFP-trapped protein.

1.4 Female Reproductive Organ & Oocyte Development in *Drosophila*

1.4.1 Development of *Drosophila* Ovaries

In many insects, including *Drosophila*, the development of the ovaries begins during embryogenesis (Quan and Lynch 2016). After fertilization, the first cells to form and bud off at the posterior end of the early embryonic syncytium are the pole cells (also called primordial germ cells), a highly specialized cell type that will give rise to germline stem cells within the adult gonads (Kotadia *et al.* 2010). Pole cells (depicted in **Figure 1.11B**) are determined by the accumulation of maternally loaded cytoplasmic determinants that localize to the posterior end of the egg as dense polar granules, composed of RNAs and proteins that distinguish them from somatic cells (Lehmann 2016).

Upon cellularization, pole cells undergo asynchronous cell divisions, resulting in the formation of a cluster of approximately 40 pole cells (Hinnant *et al.* 2020). During gastrulation, the invagination of the posterior midgut primordium passively carries pole cells to the interior of the embryo along the dorsal side (Richardson and Lehmann 2010). Pole cells then polarize to actively and bilaterally migrate towards the mesodermal gonad regions to tightly associate and coalesce with somatic gonadal precursor cells to establish embryonic gonads (Santos and Lehmann 2004). During larval and pupal development, pole cells receive signals from neighbouring somatic cells that specify their fate to give rise to ovarian germline stem cells (GSCs; Hinnant *et al.* 2020). As a result, GSCs undergo a complex reorganization with various subpopulations of somatic cells of the gonad. This process is crucial for the formation of individual ovarioles, the functional egg-producing units of the adult ovary (Sarikaya *et al.* 2012). Importantly, the number and arrangement of the terminal filament stacks within the larval/pupal ovary play a key role in predetermining the number of ovarioles that will eventually form in the adult ovary (Sahut-Barnola *et al.* 1996; Tarikere *et al.* 2022).

1.4.2 Ovarian Morphology of Adult Female *Drosophila*

Insect ovaries are morphologically different in terms of the organization of germ and somatic cells within the mature adult ovary, and they differ in how these cells support oocyte development (Jia *et al.* 2015). For more details on the three different types of ovarian organizational arrangements utilized by the class Insecta, refer to Jia *et al.* (2015)

and Church *et al.* (2021). Like many higher insects, *D. melanogaster* adopted the polytrophic meroistic ovarian morphology (Church *et al.* 2021). In this system, the oocyte develops in a cyst and is interconnected to an adjacent population of supporting cells that supply the developing oocyte with essential mRNAs, proteins, and other cellular material (e.g. nutrients and cytoplasm) through cytoplasmic junctions called ring canals (He *et al.* 2011; Cuevas 2015). Additionally, the cyst is enveloped by a layer of somatic cells forming a distinct independent unit called an egg chamber. The oocyte and its supporting siblings remain in close contact with one another as the egg chambers develop and move through the ovariole (Bastock and St Johnston 2008). Ovarioles in polytrophic meroistic ovaries are comprised of chains of egg chambers, with each egg chamber supporting the development of a single oocyte (Figure 1.7B).

1.4.3 The Process of Oogenesis in *Drosophila*

Oogenesis is a complex process by which germ cells in adult ovaries develop into viable eggs, a topic extensively reviewed in the literature (Himmant *et al.* 2020; Milas and Telley 2022; St Johnston 2023). A female fruit fly possesses a pair of ovaries with each ovary typically composed of approximately 15-18 ovarioles (See Figure 1.7; Giedt and Tootle 2023). Each ovariole is composed of a germarium and a series of egg chamber units of increasing developmental stages, acting as an effective assembly line for continuous egg production (Figure 1.7B; Bastock and St Johnston 2008). Near the anterior tip of the ovariole, in a region called the germarium, reside two populations of stems: germline stem cells and somatic stem cells (SSC).

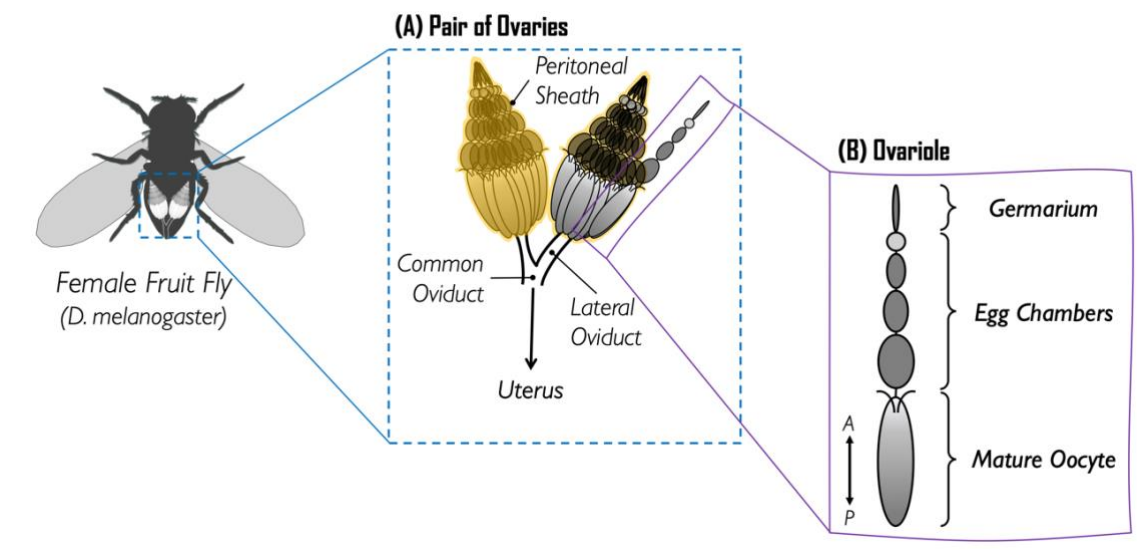


Figure 1.7: Overview of the *Drosophila* female reproductive system. An illustration depicting the relative position of ovaries within a *D. melanogaster* female, which are located at the posterior (Giedt and Tootle 2023). **(A)** The female fruit fly reproductive system is located in the posterior of the abdomen and consists of two ovaries and lateral oviducts. These bilateral oviducts lead the mature egg down the common oviduct towards the uterus and gonopore (Bloch Qazi *et al.* 2003). Each ovary consists of 15–18 ovarioles, all of which are encased by a peritoneal sheath. **(B)** Oogenesis begins at the anterior tip of the ovarioles in a region known as the germarium, where germline cells divide and are packaged into discrete functional units called egg chambers. Ovarioles are made up of linear arrays of egg chambers that are progressively more developed (Hinnant *et al.* 2020).

Oogenesis begins in the germarium with ovarian GSCs undergoing asymmetrical mitotic division to support the continuous formation of oocytes (Cuevas 2015), as depicted in **Figure 1.8B**. One of the two progenies produced remains as a stem cell to replenish the GSC population, while the other daughter cell becomes a cystoblast and goes down the differentiation route (Hinnant *et al.* 2020). The cystoblast undergoes four rounds of incomplete mitotic cell division due to incomplete cytokinesis, forming a syncytium germline cyst of 16 interconnected germline progenitor cells (**Figure 1.8B**). The cluster of 16 cystocytes of a cyst is interconnected via intercellular bridges, known as ring canals (Loyer *et al.* 2015). In region 2a of the germarium, various mechanisms involving the accumulation of maternally loaded determinants, levels of cell adhesion molecules, relative position within the cyst, and among other factors, a single germ cell from the pool of 16 germ cells will acquire the potential to develop into an oocyte, while the remaining 15 germline cells will take the nurse cell fate to support the developing oocyte throughout oogenesis (Milas and Telley 2022; StJohnston 2023). Soon after the oocyte is specified, it moves to the posterior end of the germline cyst in a DE-Cadherin-dependent mechanism (Godt and Tepass 1998a), establishing an anterior-posterior asymmetry within the egg chamber. This initial break in symmetry is crucial in setting the preconditions required to define the future embryonic body axis (González-Reyes and Johnston 1994; Huynh and StJohnston 2004), and eventually shaping the fly adult body.

In the 2b region of the germarium, somatic stem cells give rise to committed cells that form a monolayer of somatic follicle cells, encapsulating the developing germline cyst to assemble an egg chamber (Slaidina and Lehmann 2014). At this stage, the precursor cell specified to differentiate and acquire an oocyte fate, is the only cell that initiates

meiosis. Before the oocyte leaves the germarium, meiosis arrests at prophase I for most of oogenesis (von Stetina and Orr-Weaver 2011). Soon after, the developing egg chamber buds off from the germarium and remains linked to adjacent egg chambers via stalk cells, like beads on a string. Egg chambers undergo complex and orchestrated morphogenetic developmental changes, and these are accompanied by oocyte maturation, where egg chamber development is divided into a series of 14 morphologically distinct stages (Figure 1.8A, Jia *et al.* 2016).

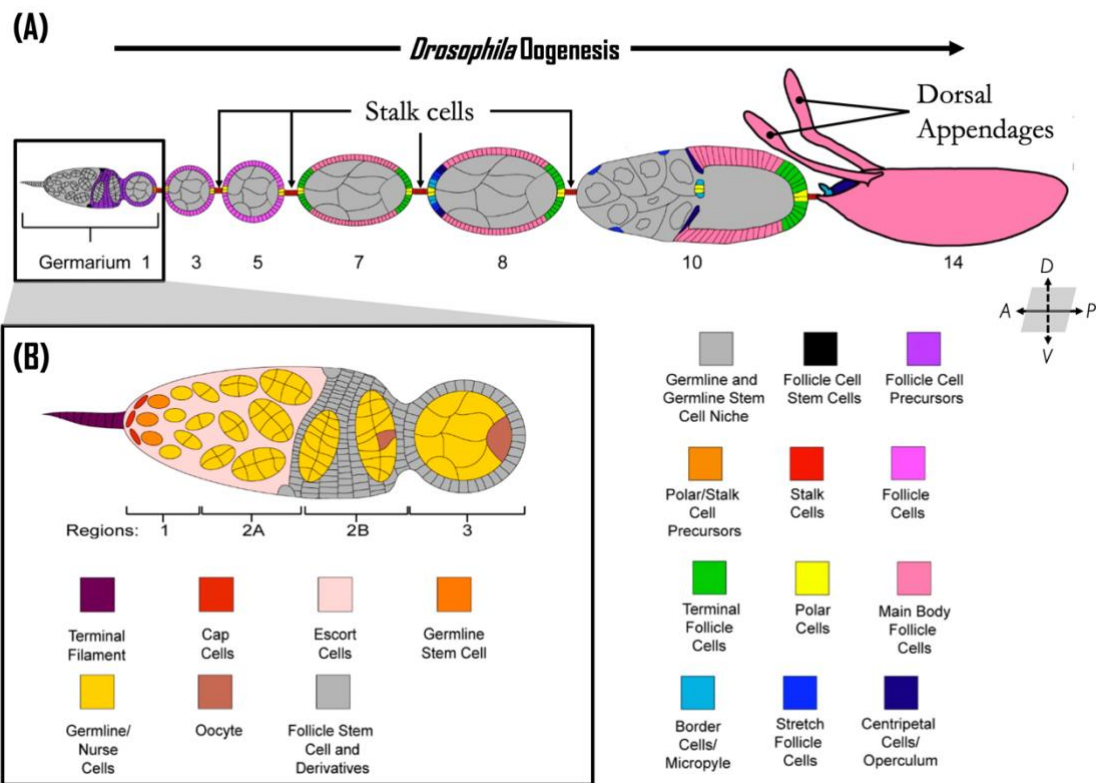


Figure 1.8: *D. melanogaster* oogenesis. (A) Schematic representation of egg chamber development within an ovariole, progressing from the germarium to the mature egg. The different cell types within various stages of egg development are color-coded as indicated by the key provided in the figure. (B) In region 1 of the germarium, germline stem cells located at the anterior tip undergo asymmetric mitotic cell division. One of the daughter cells undergoes four rounds of synchronous division to produce 2, 4, 8, and eventually 16 interconnected germline cysts. Before the germline syncytium buds off from the germarium, it becomes fully encapsulated by a monolayer of follicular epithelium. This happens in region 3 of the germarium, also referred to as stage 1 of oogenesis (in panel A). The egg chamber, which is composed of an oocyte, 15 nurse cells, and somatic follicle cells, undergoes 14 morphologically distinct stages of development. This process results in a fully developed oocyte that is structurally unique and different from the original egg chamber. Diagrams under the open access license were obtained and modified from Lebo and McCall (2021).

Chapter I – Introduction

Upon exiting the germarium, egg chambers continue their development and progress through stages 2–7 of oogenesis, which takes approximately two days (Cuevas 2015). During this time, both the oocyte and nurse cells considerably grow in size. While the nurse cells undergo endoreplication, the chromosomes within the oocyte’s nucleus remain condensed as a karyosome. Simultaneously, the follicle cells divide mitotically to encapsulate the growing germline syncytium. As explained below, the follicular epithelium also undergoes intricate patterning and morphogenetic changes. During mid-oogenesis, vitellogenesis begins, wherein the developing oocyte uptakes and accumulates large amounts of yolk produced by the nurse cells, the overlying follicular epithelium, and others (e.g., fat bodies). Additionally, the *Drosophila* oocyte experiences two distinct stages of cytoplasmic streaming (Quinlan 2016): a relatively slow streaming during mid-oogenesis that onsets at stage 8, followed by a faster streaming during late oogenesis that begins at stage 10. The ooplasmic streaming, which depends on the cortical cytoskeletal network, enables the circulation and distribution of cytoplasmic content within the oocyte (Ganguly *et al.* 2012; Lu *et al.* 2016). Therefore, cellular determinants not properly anchored in place would be displaced during these cytoplasmic streaming events. Most importantly, the major body axes of the embryo are established by this stage of development, as described in a later section.

Later in oogenesis, specifically starting from late stage 11, the oocyte also uptakes cytoplasm, ribosomes, and other small organelles from its germline siblings in a process referred to as nurse cell dumping (Mahajan-Miklos and Cooley 1994; Guild *et al.* 1997; Buszczak and Cooley 2000; Lu *et al.* 2022). This transport of cytoplasmic components occurs through intracellular bridges between the oocyte and nurse cells called ring canals (W. Lu *et al.* 2021), which are formed as a result of incomplete cytokinesis in the germarium. Concurrent with dumping, nurse cells at the anterior tip of the egg chamber undergo programmed cell death (Antel and Inaba 2020). Although the precise mechanism behind nurse cells’ degeneration remains unknown, studies showed that it requires the JNK pathway from neighbouring stretch follicle cells and the involvement of caspase Dcp-1 (McCall and Steller 1998; Timmons *et al.* 2016; Timmons *et al.* 2017). Therefore, by the end of oogenesis, the oocyte is the only remaining germline cell,

which is encased within an eggshell structure produced by the somatic follicle cells during oogenesis.

During the final stages of oocyte maturation and via unknown developmental cues at stage 14, the egg's nuclear envelope breaks down for meiosis to briefly resumes before entering a secondary meiotic arrest in metaphase I and continues until the eggs passes through the oviduct (Xiang *et al.* 2007; Avilés-Pagán and Orr-Weaver 2018). In *Drosophila*, meiosis resumes to completion in response to egg activation, which is triggered through a sperm-independent manner involving rehydration and mechanical pressure during ovulation (i.e., passage of mature stage 14 oocyte through the oviduct; (Kotadia *et al.* 2010; Fellmeth and McKim 2022). The entire process of oogenesis, from an ovarian GSC to a mature egg (depicted in **Figure 1.8**), takes approximately a week (He *et al.* 2011).

Patterning and Morphogenesis of the Follicular Epithelium Throughout Oogenesis:

The follicular epithelium is formed in the posterior-most region of the germarium, where the germline cyst, consisting of the oocyte and germline siblings, is encapsulated by somatic cells (Franz and Riechmann 2010; Fadiga and Nystul 2019). Initially, these immature follicle cells, which have a uniform identity and are characterized by their cuboidal cell shape, constitute the entire population of the follicular epithelium surrounding the egg chamber (Wu *et al.* 2008). Throughout oogenesis, the follicular epithelium differentiates into several distinct subpopulations in response to multiple signalling pathways, including Delta-Notch, JAK/STAT, and Gurken (Assa-Kunik *et al.* 2007; Jaglarz *et al.* 2008). These subpopulations can be identified based on their morphology or position within the egg chamber. The morphogenesis of the *Drosophila* ovarian follicular epithelium has been comprehensively reviewed by Duhart *et al.* (2017).

In the germarium, Delta-Notch signalling from the underlying germline syncytium kicks-off follicular epithelium patterning by coordinating the specification of the terminal follicle cells from the mainbody follicle cells (Grammont and Irvine 2001). Additionally, a small cluster of undifferentiated stalk/polar follicle cell precursors, which separate younger cyst in region 2b from the older cyst in region 3, start to express *fringe*

(Grammont and Irvine 2001). Fringe, a modulator of Notch signalling, is a glycosyltransferase that enhances the selectivity and sensitivity of the Notch receptor to Delta (Panin *et al.* 1997; Blair 2000; Grammont and Irvine 2001).

Just before the egg chamber exits the germarium, the small cluster of fringe-expressing follicle cells that are in direct contact with the germline cyst differentiate into polar cells due to high levels of Notch pathway activation (see red cells in stage 1 of **Figure 1.9**; Grammont and Irvine 2002; Torres *et al.* 2003). Acting as a signalling center, polar cells aid in the patterning of the follicular epithelium along the anteroposterior axis, particularly by influencing neighbouring immature follicle cells through the secretion of the JAK/STAT signalling ligand Unpaired in a morphogen-dependent manner (Baksa *et al.* 2002; McGregor *et al.* 2002; Assa-Kunik *et al.* 2007). In turn, Unpaired secreted by polar cells activates JAK/STAT signalling in the remaining anteriorly located polar/stalk cell precursors that are not in direct contact with the germline cyst (Assa-Kunik *et al.* 2007). These cells, which do not receive Delta-Notch signalling but do receive Unpaired signals, differentiate into interfollicular stalk cells and intercalate to form a bridge connecting adjacent egg chambers to one another (see cells in red; **Figure 1.8A**).

Both polar and stalk cells are among the first follicle cells to differentiate, while the remaining main body follicle cells remain undifferentiated and continue mitotic cell division until mid-oogenesis (Roth and Lynch 2009). By stage 5 of *Drosophila* oogenesis, the follicular epithelium is now subdivided into polar cells at either end of the egg chamber, terminal cells, and mainbody cells (see **Figure 1.9**). Additionally, it is worth noting that, up to this stage, the follicle cells at both terminal regions of the egg chamber are symmetrically prepatterned, meaning that they have similar identities.

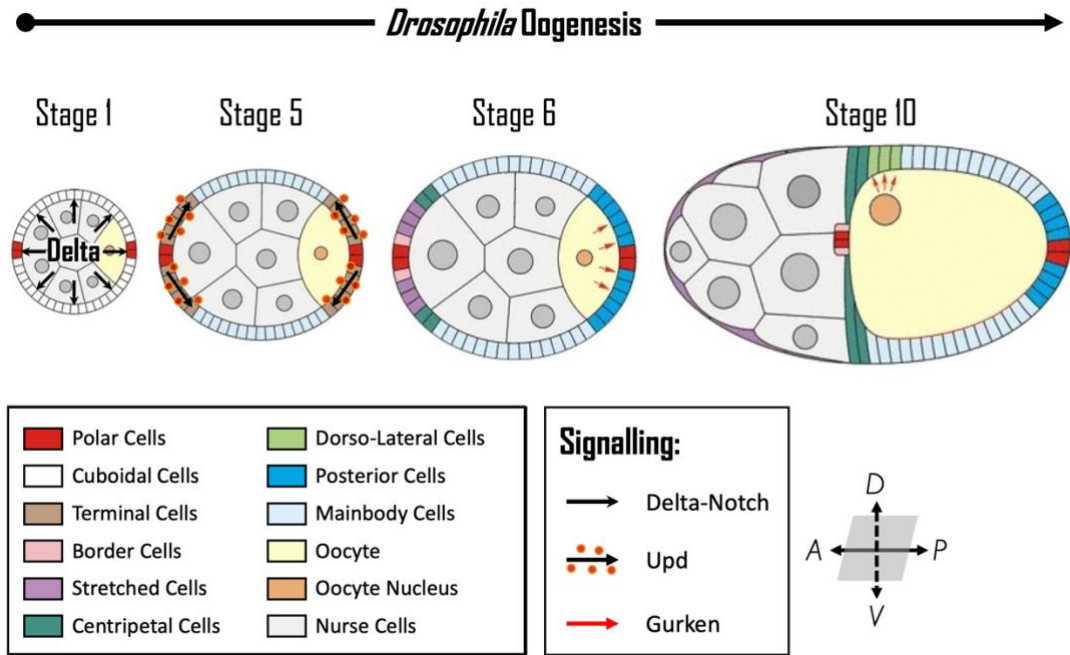


Figure 1.9: Patterning and morphogenesis of the *Drosophila* ovarian follicular epithelium.

Throughout oogenesis, the follicular epithelium, which is the ovarian somatic cells surrounding the 16-cell germline syncytium (comprising 1 oocyte and 15 nurse cells), undergoes extensive and intricate patterning and reorganization (Duhart *et al.* 2017). These cells can be readily distinguished based on their morphology and relative position to the oocyte. Follicle cells are smaller than germline cells. At stage 1 of oogenesis, two main subpopulations of follicle cells are found around the periphery of each egg chamber, a result of the Delta-Notch signalling pathway initiated by the underlying germline cells (shown as black arrows). However, by stage 5, polar cells at the egg chamber termini secrete the ligand ‘unpaired’ (Upd; depicted as black arrows with red circles), which activates the JAK-STAT signalling pathway in neighbouring follicle cells, leading to their differentiation into terminal cells. During mid-oogenesis, the oocyte secretes Gurken (indicated by the red arrows) towards the posterior of the egg chamber, thereby activating EGFR signalling in follicle cells through the Torpedo receptor. Later in oogenesis, Gurken signalling originating from the dorsal-anterior corner of the oocyte leads to the specification of the dorsal fate. By stage 10, different sub-populations of follicle cells are established. The diagram was obtained from Jaglarz *et al.* (2008) with permission and was adapted for this context.

During mid-oogenesis, around stage 6, this symmetry is broken by the secretion of the TGF- α signalling molecule, encoded by the *gurken* gene, from the oocyte’s posterior pole (González-Reyes *et al.* 1995). This localized Gurken signalling at the posterior establishes the anterior-posterior axis and prompts the terminal cells adjacent to the posterior polar cells to adopt the posterior fate, as depicted by the red arrows in **Figure 1.9**. While at the anterior, lack of Gurken signalling enables anterior follicle cells to further differentiate and subdivide (Jaglarz *et al.* 2008), each having distinct and specialized roles later in oogenesis. At the same time, a second wave of Delta-Notch signalling induces the remaining immature follicle cells to exit the endocycle and

differentiate. By stage 6 of oogenesis, as shown in **Figure 1.9**, the anterior terminal follicle cells are further specified into three distinct subpopulations, easily distinguishable based on their proximity to the anterior polar cells (González-reyes and Johnston 1998): the closest being the border cells and stretched cells, while the slightly more distal cells became centripetal cells. Meanwhile, the mainbody follicle cells and terminal posterior cells begin to adopt a columnar cell shape (Jaglarz *et al.* 2008).

After the anterior-posterior axis is established during mid-oogenesis, the nucleus and *gurken* mRNA translocate to the anterior-dorsal region of the oocyte (see Stage 10 in **Figure 1.9**). Here, a second wave of local Gurken signalling specifies the dorsal fate, while its absence on the ventral side determines the ventral fate. Concurrently, during stages late-8 through 10 of oogenesis, a cluster of cells composed of border cells and anterior polar cells migrates towards the oocyte (Rorth 2002; Montell *et al.* 2012). By stage 10 of oogenesis, cells of the follicular epithelium become specified and are properly positioned within the egg chamber, as illustrated in **Figures 1.8A** and **1.9**.

During stages 10B-14, also referred to as the choriogenic stage, follicle cells deposit the vitelline membrane and eggshell over the oocyte (Cuevas 2015). The *Drosophila* eggshell is a multifaceted, proteinaceous extracellular structure that originates from the follicular epithelium (Cavaliere *et al.* 2008). This eggshell is characterized by four morphologically and functionally distinct features: the micropyle, which facilitates sperm entry; the dorsal appendages, which plays a role in gas exchange; the operculum, an opening through which the larva emerges during hatching; and the mainbody shell, a multi-layered hydrophobic protective layering (see stage 14 in **Figure 1.8A**). The dorsal appendages, in particular, as the most prominent structures on the eggshell, serve as a useful morphological readout of proper body axis patterning during oogenesis (Neuman-Silberberg and Schupbach 1994). Their proper formation and positioning are often used as indicators of the correct asymmetric localization and expression of maternally loaded mRNAs. Around stage 10 of oogenesis, a cocktail of signalling molecules (including Gurken, DPP, and ecdysone) at the dorsal-anterior region of the oocyte specifies two separate subpopulations of the dorsal follicle cells. These cells

undergo a series of intricately orchestrated morphogenetic events and eventually give rise to the dorsal appendages, as detailed in Berg (2005).

1.4.4 Border Cell Migration

During early stage 9 of oogenesis, a cluster of approximately 8 somatic cells, comprising of both border cells and anterior polar cells, delaminate from the anterior end of the uniform monolayer follicular epithelium (Rorth 2002). Afterwards, these cells, termed the border cell cluster, coalesce to form a migrating cohesive unit (Montell 2003). Within the border cell cluster, the immobile polar cells are positioned internally and are carried along by the actively migrating outer border cells (Figure 1.10; Han *et al.* 2000). Between stages 9 and 10 of oogenesis, border cells begin their migration from the anterior pole of the egg chamber towards the oocyte, travelling a distance of 150–200 μm over the course of approximately 5 hours (Prasad and Montell 2007). This collective cell migration is guided by a cocktail chemoattractants secreted by the oocyte, including Gurken, Spitz, PVF1, and Keren (Prasad and Montell 2007; Poukkula *et al.* 2011).

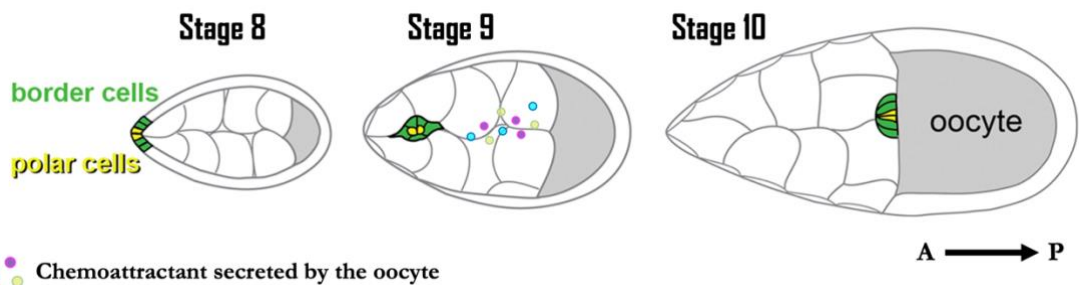


Figure 1.10: Border cell cluster migration within *D. melanogaster* egg chambers. A schematic representation showing border cell cluster migration through egg chambers during stages 8–10 of oogenesis. At the beginning of stage 8, before the onset of migration, anterior polar cells (shown in yellow) and surrounding border cells (shown in green) reside within the anterior end of the follicular epithelium of the developing egg chamber (Montell *et al.* 2012). At this stage, both cell types, along with the entire follicular epithelium, exhibit canonical apical-basal polarity (Veeman and McDonald 2016). The apical side faces the inside of the egg chamber, contacting the nurse cells and oocyte, while the basal side faces the outside of the egg chamber (Montell *et al.* 2012). At the start of stage 9, border and polar cells coalesce to form a cluster that detaches from the epithelium as they begin their migration through the egg chamber. The cluster uses germline nurse cells as a migratory substrate, moving in a guided manner towards the developing oocyte (Montell 2003). This migration is directed by a chemoattractant gradient secreted by the oocyte. By stage 10, the border cell cluster had migrated a distance of about 150 μm over approximately 4–6 hours to reach the anterior side of the oocyte (Prasad and Montell 2007; Stonko *et al.* 2015).

The direction of border cell cluster migration from the anterior side (A) of the egg chamber towards the posterior (P) is indicated by the arrow. The diagram under the open access license was obtained and modified from Stonko *et al.* (2015).

The posterior migration of border cells involves aspects of the epithelial-to-mesenchymal transition (EMT; Lu and Lu 2021). The restricted secretion of the ligand Unpaired by polar cells results in high-level activation of the JAK/STAT signalling pathway in neighbouring somatic cells, which in turn regulates border cell specification and migration (Silver and Montell 2001). Among the genes activated by the JAK/STAT signalling pathway is the *slow border cells (slbo)* gene, which encodes a homologue of the mammalian transcription factor C/EBP (Silver and Montell 2001). Slbo is essential for the expression of genes that specify the fate of border cells and enable them to become motile (Montell *et al.* 1992; Rørth *et al.* 2000). Notably, DE-Cadherin, focal adhesion kinase, β -integrin, and myosin VI are among the target genes of Slbo that are crucial for border cell migration (Montell *et al.* 2012).

For border cells to delaminate from the follicular epithelium, they first must lose their epithelial characteristics (i.e., the canonical apical-basal polarity) by downregulating factors associated with cell-cell adhesion, all while maintaining contact amongst themselves and with anterior polar cells (Veeman and McDonald 2016). As a result of Notch activity and the loss of cell-cell adhesion, the border cell cluster dissociates from the basal lamina and invades its substrate, migrating between the nurse cells towards the oocyte (Prasad and Montell 2007; Wang *et al.* 2007). Additionally, border cells become highly polarized with respect to actin polymerization, membrane movement, contraction, and substrate adhesion molecules, adopting a front-rear polarity that is spatially and temporally orchestrated (Montell 2003; Llense and Martín-Blanco 2008; Lamb *et al.* 2021). This polarity allows the migratory border cell cluster to use both its anchoring rear and front-directed protrusions to adhere to the substrate, sense directional chemotactic signals from the environment, and propel the cell forward (Ridley *et al.* 2003; Montell *et al.* 2012).

The border cell cluster performs two essential functions at the anterior end of the oocyte, and females with defective border cell migration are sterile (Montell *et al.* 1992;

Savant-Bhonsale and Montell 1993). Border cells are required to form a specialized eggshell structure at the anterior end of the oocyte called the micropyle, as indicated in bright blue in stage 14 of **Figure 1.8A** (Montell *et al.* 1992; Montell 2003). This structure contains a pore through which the sperm enters during fertilization. Additionally, border cells contribute to the patterning of the embryo's head regions by expressing *torso-like*, a gene that encodes a secretory protein essential for the localized activation of the receptor tyrosine kinase Torso (Savant-Bhonsale and Montell 1993; LeMosy 2003; Johnson *et al.* 2017).

1.4.5 Establishment of Major Body Axes During Oogenesis

In *Drosophila*, the major body axes will both have been already defined by the time the mature egg is laid (Barresi and Gilbert 2023). This ensures the proper development of the embryo in the next 24 hours and subsequent stages of the animal's life cycle, including the larva, pupa, and adult fly. The anterior-posterior and dorsal-ventral axes are established due to a series of intricate symmetry-breaking events that occur during oogenesis (Roth and Lynch 2009). Each symmetry-breaking event sets the conditions required for the proper onset and completion of subsequent events in highly coordinated morphogenetic processes involving intracellular communication (St Johnston 2023). During the early stages of oogenesis in the germarium, the first visible sign of axis symmetry breaking along the anterior-posterior axis is observed when the oocyte is positioned posteriorly within the germline cyst (see region 3 in **Figure 1.8B**; Godt and Tepass 1998; Milas and Telley 2022). This early asymmetry within the egg chamber lays the foundation for establishing the anterior-posterior axis of the embryo.

During mid-oogenesis, the accumulation of *gurken* mRNA at the oocyte's posterior leads to localized Gurken signalling (see stage 6 of **Figure 1.9**; Roth and Lynch 2009). This activates the EGF receptor Torpedo, which specifies the fate of the posterior terminal follicle cells. Through an unidentified mechanism, these follicle cells signal back to the oocyte, repolarizing its microtubule cytoskeleton (Milas and Telley 2022). Consequently, through a mechanism mediated by centrosomal microtubules, both the oocyte nucleus and *gurken* mRNA relocate to the dorsal-anterior corner of the oocyte (Zhao *et al.* 2012; Tillery *et al.* 2018). Both the oocyte's nucleus and the *gurken* mRNA

are then anchored at their final destination by the motor protein Dynein (Zhao *et al.* 2012). Simultaneously, the plus-end of microtubules becomes restricted to the oocyte's posterior. Once the *gurken* mRNA reaches the dorsal-anterior corner of the oocyte, it is translated into the Gurken protein (González-Reyes *et al.* 1995). For the second time, Gurken then signals the adjacent follicle cells via Torpedo (see stage 10 of **Figure 1.9**). Because of this localized Gurken signalling, both the dorsal fate of the eggshell and the dorsal-ventral axis of the embryo are specified (see stage 14 in **Figure 1.8A**). It is important to emphasize that Gurken is instrumental not just in establishing the *Drosophila* major body axis but also in the formation of the eggshell's dorsal-anterior structures (Queenan Anne Marie *et al.* 1997; Peri and Roth 2000; Duhart *et al.* 2017).

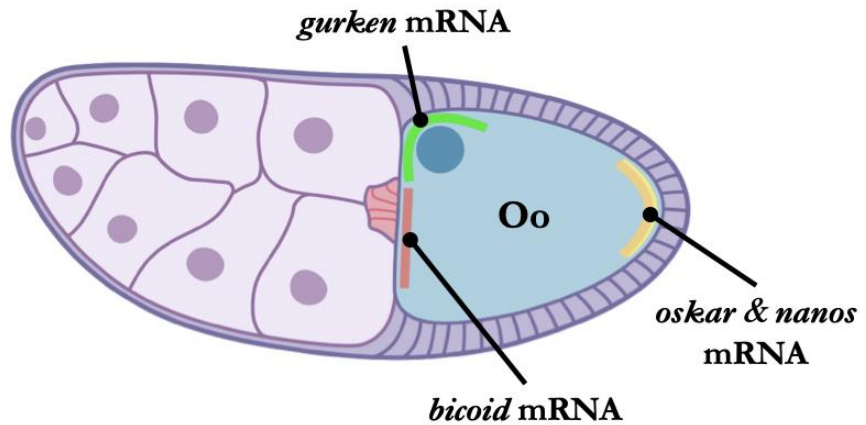
1.4.6 mRNA Localization & Spatial Translational Regulation

The establishment of the anterior-posterior asymmetry within the egg chamber during *Drosophila* oogenesis is essential for the asymmetric subcellular accumulation and translation of maternally expressed mRNAs, which in turn lay out the boundaries for future embryonic body axes patterning and formation (**Figure 1.11**). One of the mechanisms by which this is achieved involves the subcellular distribution of certain mRNAs and their localized translation to distinct regions within the oocyte's cytoplasm (**Figure 1.11A**). Cell polarity within the oocyte is both generated and maintained through microtubule-dependent localization of maternal mRNA molecules (Bastock and St Johnston 2008). This process is facilitated by the reorganization of the intracellular cytoskeleton in response to developmental signalling cues that take place during mid-oogenesis. Additionally, the spatial control of mRNA translation arises from the combinatorial effects of RNA-binding proteins interacting within the mRNP complex. Finally, it is worth highlighting that the early stages of embryogenesis rely on the expression of maternal mRNAs before the zygotic gene expression takes over in a process called the maternal-to-zygotic transition (Lasko 2012).

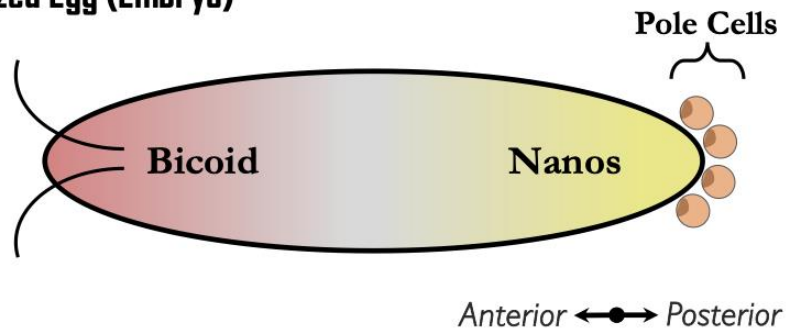
Specification of the embryonic axis is established through asymmetric localization coupled with spatially restricted translation of four key maternal determining mRNAs: *oskar*, *bicoid*, *gurken*, and *nanos* (**Figure 1.11A**). For more details on the localization, anchoring, and translational control of these well-studied mRNAs during oogenesis, as

well as their biological roles during development, refer to the reviews by (Kugler and Lasko 2009; Lasko 2012; Lasko 2020). Briefly, the embryonic anterior-posterior axis is formed via the restricted localization of *bicoid* mRNA to the anterior of the oocyte, while *nanos* and *oskar* mRNAs localize to the posterior (**Figure 1.11A**). On the other hand, the formation of the dorsal-ventral axis is established through the localized translation of *gurken* mRNA at the anterior-dorsal corner of the oocyte, as previously described in Section 1.4.5. After fertilization, the translational repression of *bicoid* and *nanos* mRNAs is relieved, leading to the formation of a Bicoid and Nanos protein gradients across the anterior-posterior axis of the embryo (Lasko 2011), where Bicoid defines the anterior boundary, while Nanos marks the posterior (**Figure 1.11B**). This gradient, in turn, initiates a cascade of zygotic gene expression that directs embryonic patterning and paves the way for proper *Drosophila* development (**Figure 1.11C**; Johnston and Nüsslein-Volhard 1992; Wolpert *et al.* 2019; Barresi and Gilbert 2023). Here, we will primarily focus on the localization and local translational control of *oskar* mRNA. This is achieved throughout oogenesis by the complex post-transcriptional regulation of *oskar*'s mRNP complex.

(A) Stage 10 Egg Chamber



(B) Fertilized Egg (Embryo)



(C) Fly Development

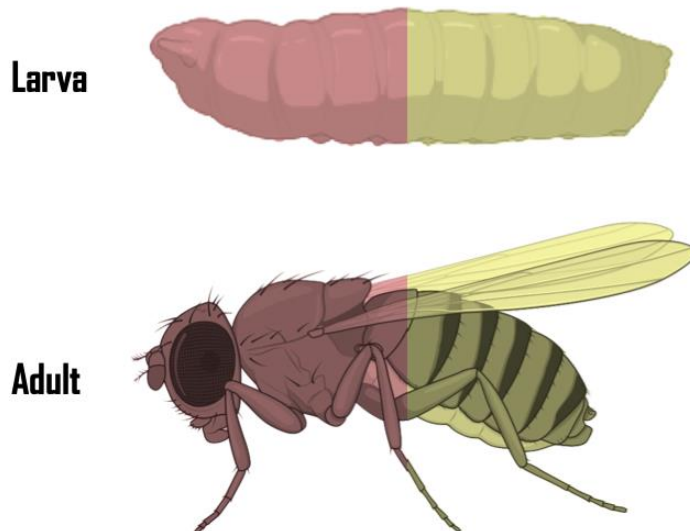


Figure 1.11: Patterning of the *Drosophila* body axes. (A) The major body axes of *Drosophila* are established in the ovary during oogenesis through the asymmetric localization of maternal mRNA within the oocyte (Oo, cell shown in blue). By stage 10 of oogenesis, *bicoid* mRNA (in red) localizes to the anterior cortex of the oocyte, while *gurken* mRNA (in green) marks the dorsal-anterior region, positioned adjacent to the oocyte's nucleus. In contrast, both *oskar* and *nanos* mRNAs (in yellow and orange) localize to the posterior cortex. (B) After fertilization, both *bicoid* and *nanos*

mRNAs are translated, and their proteins then specify the anterior and posterior boundaries of the embryo, respectively. **(C)** The early specification of the anterior (shown in red) and posterior (shown in yellow) boundaries by the Bicoid and Nanos proteins during embryogenesis initiates a cascade of events leading to proper segmentation and patterning of the larva. This then paves the way for proper patterning during metamorphosis, ensuring the formation of the adult body with its complete set of structures, organs, and systems. The diagram was constructed using BioRender based on the literature (Wolpert *et al.* 2019; Barresi and Gilbert 2023).

Localization of *oskar* mRNA During Oogenesis: In the germarium and during the early stages of oogenesis, the microtubule cytoskeletal network is organized such that the microtubule minus-ends are concentrated within the oocyte (Riechmann and Ephrussi 2001). As a result, the transportation of *oskar* mRNA, which is transcribed in the nuclei of polyploid nurse cells, relies on a conserved transport pathway involving Dynein, the minus-end directed motor protein complex (Lasko 2012). During mid-oogenesis, the oocyte's microtubule cytoskeleton undergoes reorganization in response to Gurken signalling at the posterior pole. Consequently, the plus-end microtubules become concentrated towards the posterior pole of the oocyte, while the minus-end microtubules accumulate at the anterior and lateral sides of the oocyte (Steinhauer and Kalderon 2006). From then onward, *oskar* mRNA becomes localized to the posterior cortex of the oocyte (shown in yellow in **Figure 1.11A**). However, this time it shifts to using Kinesin-1, a plus-end directed motor protein complex, for its transport along the microtubule cytoskeletal network (Brendza *et al.* 2000). The entire process of targeting *oskar* mRNA localization to the oocyte heavily depends on RNA-binding proteins, including Egl, Stau, Hrp48, Sqd, Glo, Otu, dDcp1, Mago, Y14, and others (Kugler and Lasko 2009). Interestingly, studies have demonstrated that interactions with RNA-binding proteins in the nucleus can imprint mRNAs destined for asymmetric localization within the cytoplasm. This comes after Hachet and Ephrussi (2004) showed that the splicing of *oskar*'s first intron is essential for its posterior localization. Once *oskar* mRNA is localized to the posterior, it is anchored in place through a microfilament-mediated mechanism involving cortical F-actin and the actin-based motor protein myosin V (Krauss *et al.* 2009; Lu *et al.* 2020). Additionally, the accumulation of *oskar* mRNA at the oocyte's posterior pole has been shown to involve both the long isoform of Oskar (described below) and endocytosis (Vanzo and Ephrussi 2002; Vanzo *et al.* 2007).

Translation control of *oskar* mRNA: Throughout the migration of *oskar* mRNA from the nurse cells to the posterior pole of the oocyte, the mRNA is subjected to complex temporal and spatial translational control to ensure its repression (Vazquez-Pianzola and Suter 2012). Once the *oskar* mRNA localizes to the oocyte's posterior cortex, it becomes derepressed due to dynamic changes in the mRNP composition, leading to the translation and accumulation of the Oskar proteins (Benton and Johnston 2002; Morais-de-Sá *et al.* 2013). The *oskar* mRNA utilizes different translational start sites within the same open reading frame to produce two isoforms of the Oskar protein: Long and Short (Markussen *et al.* 1995). Although both isoforms are involved in directing the establishment of the pole plasm and the formation of pole cells, they play different biological roles. The Long Oskar protein anchors both *oskar* mRNA and the Short isoform to the oocyte's posterior cortex by remodelling the actin cytoskeleton (Vanzo and Ephrussi 2002; Babu *et al.* 2004; Tanaka and Nakamura 2011; Yang *et al.* 2015). On the other hand, the Short Oskar isoform is necessary for the assembly of the germ plasm, from which the primordial germ cells are specified at the posterior pole (**Figure 1.11B**), giving rise to the embryo's germ cells (see Section 1.4.1; Yang *et al.* 2015; Lehmann 2016; Blondel *et al.* 2021).

In addition to coordinating the assembly of germ plasm, Oskar is essential for directing the localization of *nanos* mRNA to the posterior (Ephrussi *et al.* 1991), which in turn plays an indispensable role in patterning the posterior abdominal segments of *Drosophila* embryos (**Figure 1.11A-C**). Improper regulation of *oskar* levels or its ectopic localization results in severe body patterning defects, ultimately leading to lethality. In transgenic flies with multiple copies of the *oskar* gene, Smith *et al.* (1992) showed that overexpression of *oskar* mRNA in the female germline directs ectopic localization and expression of Oskar along the anteroposterior axis, resulting in two distinct phenotypes: firstly, embryos in which the head, thoracic, and anterior abdominal segments are replaced by posterior abdominal segments; and secondly, ectopic pole cell formation. These results demonstrate the importance of restricting Oskar activity to the posterior pole of the oocyte. The *oskar* mRNP relies on numerous RNA-binding proteins with overlapping translational repression activities to ensure localized translation and to restrict Oskar activity to the posterior pole of the oocyte.

Among these proteins are PTB, Bru, Cup, Hrp48, Yps, Me31B, Sqd, Exu, and Bic-C (Mansfield *et al.* 2002; Chekulaeva *et al.* 2006; Besse *et al.* 2009; Kugler and Lasko 2009; Wang *et al.* 2017).

Although several proteins have been implicated in the activation of *oskar* mRNA translation when it reaches the oocyte's posterior cortex, including Orb, Aubergine, and specifically the dsRBD5 of Staufen (Wilson *et al.* 1996; Chang *et al.* 1999; Micklem *et al.* 2000), little is known about this process. Derepression of translation requires functional interactions between *trans*-acting regulatory factors and both the 5' and 3' untranslated regions of *oskar* mRNA (Gunkel *et al.* 1998; Kanke and Macdonald 2015; Kim *et al.* 2015). Recently, Dold *et al.* (2020) implicated the RNA-binding protein Mkrn1 as a key activator of *oskar* translation. In their study, they demonstrated that Mkrn1 antagonizes Bruno and promotes the recruitment of the poly(A) binding protein through its binding to the 3'UTR of *oskar*.

Altogether, an arsenal of proteins ensures not only a robust mechanism for the localization and translational repression of *oskar* mRNA but also confines protein expression to the posterior pole of the oocyte. Given the developmental importance of the Oskar protein, there is a pressing need to decipher the complete RNA-binding protein composition that regulates the translational control of the *oskar* mRNP complex.

1.5 The *Drosophila* Alan Shepard (Shep) Protein

This project will focus on the potential role of *Drosophila* Alan Shepard (Shep) in the post-transcriptional regulation of maternally loaded mRNAs during oogenesis. Our current understanding of the biological role of Shep and the rationale behind its selection for investigation in this project are described below.

1.5.1 The *shep* Gene in *Drosophila*

The *Drosophila* gene *alan shepard* was originally identified in a forward genetic screen aiming to identify genes involved in sensory-motor responsiveness to gravity (Armstrong *et al.* 2006). The gene encodes a highly conserved RNA-binding protein (see **Figure 1.12**) that contains two RNA-recognition motifs located towards the protein's C-

terminus (Matzat *et al.* 2012). Shep orthologs are found in species ranging from *C. elegans* to higher vertebrates, including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Xenopus tropicalis*, *Danio rerio*, and *Daphnia magna* (Armstrong *et al.* 2006; Hu *et al.* 2011; Perez *et al.* 2021).

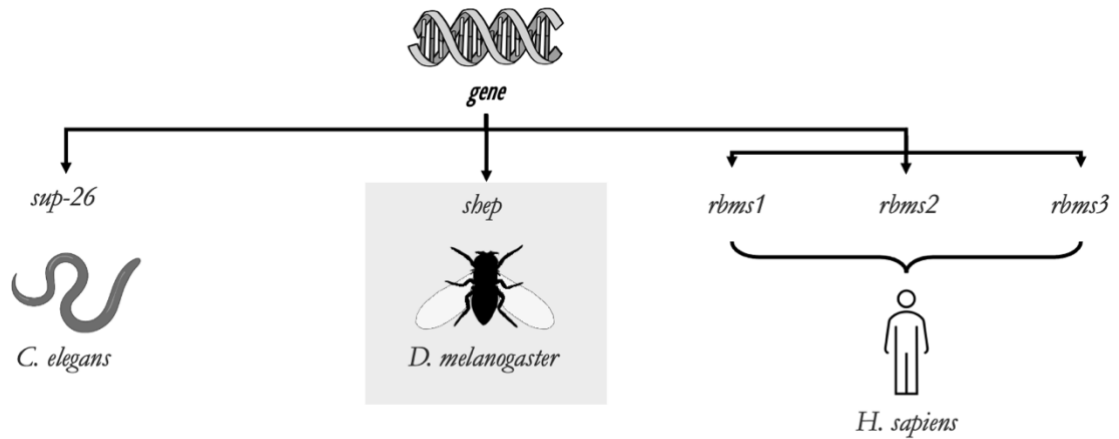


Figure 1.12: Drosophila *shep* gene orthologs. The *Drosophila melanogaster shep* gene encodes an RNA-binding protein that is highly conserved and is found in simple multicellular organisms (like *C. elegans*) all the way to higher vertebrates (such as *Homo sapiens*). The *D. melanogaster* Shep protein and its *C. elegans* ortholog (Sup-26) are encoded by a single gene. Whereas, their human counterpart is encoded by three different genes (*rbms* 1–3).

Chromosomal Location, Transcriptional Diversity, & Alternative Splicing: In *D. melanogaster*, the gene is situated on the negative strand of the left arm of the third chromosome (Gramates *et al.* 2022). Additionally, according to the modENCODE embryonic dataset, the *shep* gene has six different transcription start sites (See **Figure 1.13**; Gramates *et al.* 2022). Furthermore, FlyBase annotations indicate that through a combination of the transcriptional start sites and alternative splicing, the gene produces eight distinct transcripts (namely A, B, D, E, F, G, H, I), from which six unique protein isoforms are translated. This highlights the potential complexity of the *shep* gene, suggesting that it may undergo extensive regulation at both the transcriptional and post-transcriptional levels. Such flexibility in transcriptional start sites and alternative splicing allows for differential control of gene expression, not only in terms of its expression levels but also in a tissue-specific manner. Interestingly, Matzat *et al.* (2012) reported that the *shep* gene also produces transcript C, which was supposedly experimentally validated by EST data available at the time. However, this transcript does not appear to

be annotated in the current iterations of FlyBase as of 2019. Notably, at the protein level, Shep-C exhibits a high degree of similarity to Shep-E (detailed below).

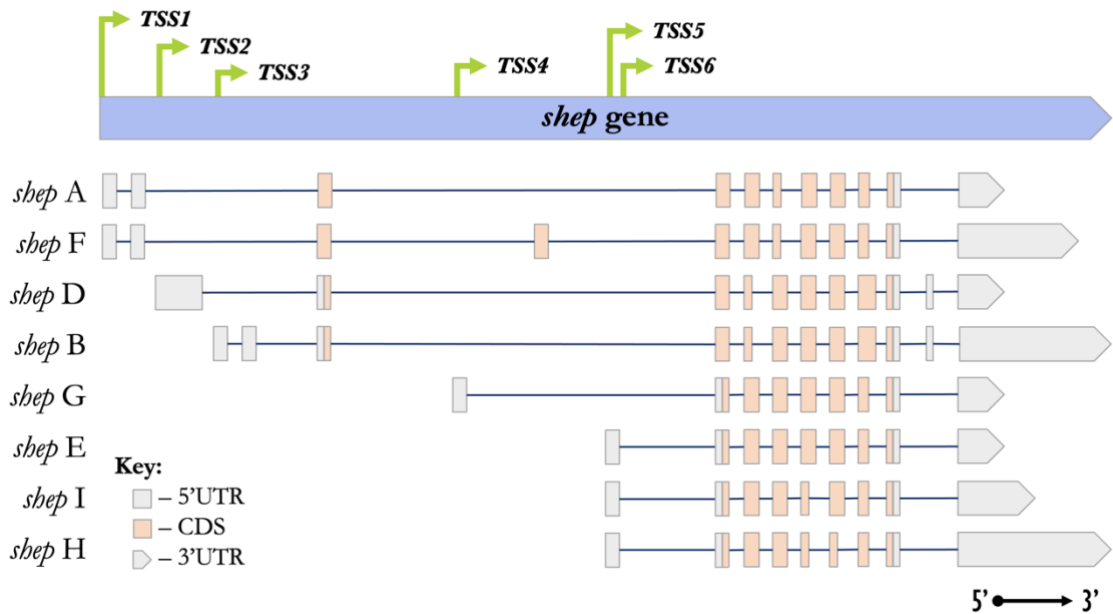


Figure 1.13: Gene region of *shep* in *Drosophila melanogaster*. A schematic representation of the *shep* gene region showing eight of the transcripts described in FlyBase. These *shep* transcripts (A, B, D, E, F, G, H, and I) arise from a combination of alternative transcriptional start sites (green arrows, TSS 1–6) and alternative mRNA splicing. The exons of the transcripts are color-coded to denote different regions of the mRNA: untranslated regions (gray) and coding sequences (orange). The lines connecting them represent introns. The diagram is not drawn to scale. Refer to FlyBase for more details.

Shep Protein Isoforms & their Structural Differences: Although the Shep protein isoforms have distinct N-terminal domains, they all have two highly conserved RNA-binding domains called RNA recognition motifs (RRMs), as shown in **Figure 1.14** (Appendix 7). Notably, isoforms B/D lack two amino acids within RRM1, as detailed in Appendix 7, which could define its binding specificity relative to other Shep isoforms. Additionally, all isoforms contain an additional 11 amino acid linker between the RRM domains, with the exception of isoforms A, F, and C (see green bars in **Figure 1.14**). Interestingly, isoforms B/D contain a unique 7 amino acid stretch at the C-terminus, as indicated by the orange bar in **Figure 1.14** (also shown in the alignment; see Appendix 7). Generally, the different isoforms of the Shep protein can be grouped based on their sizes: those that are around 60 kDa (i.e., isoforms A/F), 50 kDa (i.e., isoforms B/D), or 40 kDa (i.e., isoforms C, E, G, H, and I). Matzat *et al.* (2012) compared *Drosophila* extracts from different larval tissues and demonstrated that the expression of Shep

isoforms is tissue-specific, e.g., all isoforms appear enriched in larval brains, while in the wing discs, isoforms A/F are predominantly enriched. Additionally, they not only demonstrated that Shep-E antagonizes gypsy chromatin insulator activity in the nervous system but also showed that its RNA-binding capacity is essential for its function in this context (Matzat *et al.* 2012; Chen *et al.* 2019). These are among the only studies that delve into investigating the role of specific isoforms, while others investigated the role of Shep in general by using deletions or RNAi, which target all isoforms. Therefore, the distinct biological roles of Shep isoforms remain poorly characterized.

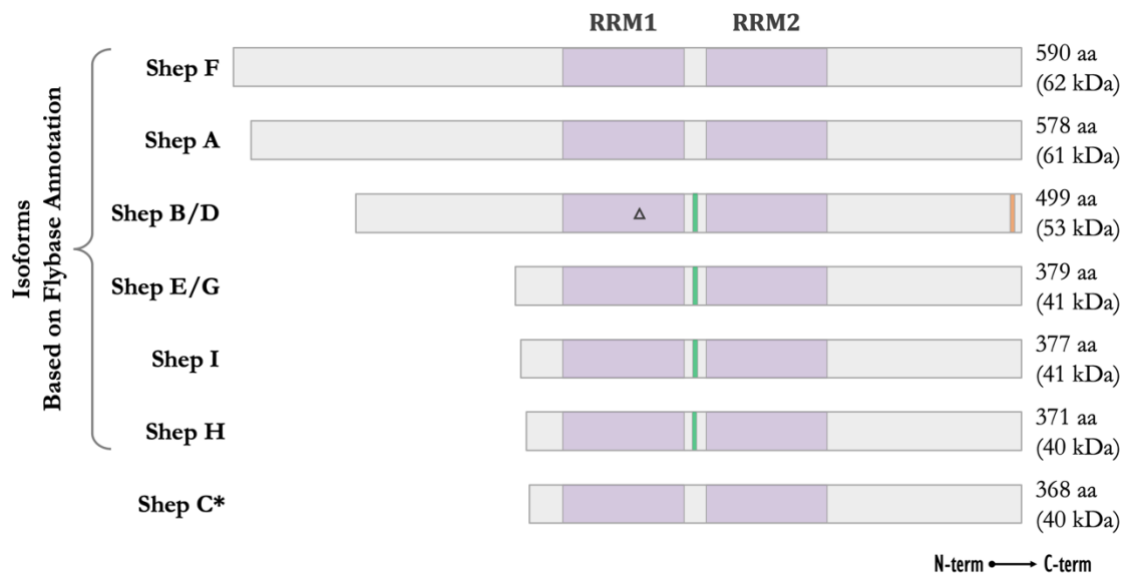


Figure 1.14: Isoforms of the *Drosophila* Shep protein. Schematic representation illustrating the structural differences among the various isoforms of the Shep protein. The Shep protein possesses two RNA-binding domains, known as RNA-recognition motifs (RRMs), which are depicted in purple. The RRM domains are separated by a linker region in all isoforms; however, some isoforms, namely B/D, G/E, I, H, and C, contain additional amino acids within this linker region, which are highlighted in green. The alternative amino acid stretch at the C-terminus of isoforms B/D is shown in orange. Isoforms B/D lack two amino acids in RRM1: K (Lysine) and G (Glycine), as indicated by the Delta symbol (Δ). Refer to Appendix 7 for a detailed alignment of the isoforms presented here. The diagram is not drawn to scale. For reference, isoforms A, B, D, E, F, G, H, and I are currently described in FlyBase (Gramates *et al.* 2022), while isoform C, indicated by the asterisks, is mentioned by Matzat *et al.* (2012).

1.5.2 Biological Roles of *Drosophila* Shep & its Non-*Drosophila* Orthologs

Ever since *shep* was first shown to play a role in the nervous system (Armstrong *et al.* 2006), it has been primarily studied within that context. To date, little is known about its role in other tissues. In the nervous system, the *shep* gene is expressed in a wide

range of differentiated neurons and glia, and the Shep protein is localized to both the nucleus and cytoplasm of these cell types (Chen *et al.* 2018). In the nucleus, Shep has been identified as a negative regulator of the gypsy chromatin insulator complex, which is a DNA-protein complex that regulates chromatin structures and gene expression by partitioning the genome into distinct domains (Matzat *et al.* 2012). Shep antagonizes the gypsy chromatin insulator by physically interacting with its core components, thereby disrupting the insulator activity. Inhibition of gypsy insulator activity by Shep is crucial for promoting neuronal maturation (Matzat *et al.* 2012; Chen *et al.* 2019). A recent study by the same group demonstrated that the RNA-binding capacity of Shep is required to antagonize the gypsy insulator activity (Chen *et al.* 2019). This role of Shep is tissue-specific and operates primarily in the central nervous system (Matzat *et al.* 2012). Moreover, during pupal neuronal remodelling, Shep regulates gene expression by inhibiting enhancer accessibility and chromatin looping (often referred to as enhancer-promoter looping; Chen *et al.* 2021). Overall, Shep in the nucleus provides a mechanism for upregulating the expression of target genes by suppressing the insulator activity.

In the cytoplasm, Shep has been found to function in an array of different neuronal cell types (Tunstall *et al.* 2012; Schachtner *et al.* 2015; Chen *et al.* 2018). These include peptidergic neurons (such as bursicon neurons in both the abdominal ganglion and subesophageal ganglion) and sensory neurons (like olfactory neurons, nociceptors, and proprioceptors). Here, Shep has mainly been shown to function during metamorphosis and in later stages of neurogenesis, specifically in processes involving remodelling, dendritic terminal branching, and maturation. Loss of *shep* leads to several pronounced defects, including late pupal lethality and a reduced adult lifespan (Chen *et al.* 2014; Chen *et al.* 2017). Additionally, depletion of *shep* results in wing expansion defects and uncoordinated adult locomotor activities. Notably, it also influences female courtship behaviour, whereby virgin females reject males. At the anatomical level, there was a reduced neuropil area (Chen *et al.* 2014).

Perhaps a notable finding by Olesnicky *et al.* (2018) is that in *Drosophila*, Shep interacts with *oskar* mRNA during sensory neuron dendrite morphogenesis. It is worth noting

that the interaction between the Shep protein and *oskar* mRNA was identified merely as part of their RNA-seq data, and no further examination was conducted to investigate the biological significance or nature of the interaction. In the same study, using immunoprecipitation-mass spectrometry, Shep was shown to associate with a range of RNA-binding proteins with diverse cellular roles, such as Yps, pAbp, and Sqd. In the ovaries, however, these proteins are well-known for regulating the subcellular distribution and translation of both *oskar* and *gurken* mRNAs (Mansfield *et al.* 2002; Clouse *et al.* 2008; Kugler and Lasko 2009; Vazquez-Pianzola *et al.* 2011). Additionally, they showed that *orb* mRNA is a target of Shep (Olesnicki *et al.* 2018). In fly ovaries, the Orb protein has been implicated in releasing the translational repression of *oskar* mRNA. Furthermore, Orb is important for the subcellular localization and translation of *grk* mRNA (Chang *et al.* 2001; Kugler and Lasko 2009).

Given the various interactions of Shep in the nervous system, which hint at its potential role in post-transcriptional regulation in the ovaries, we became increasingly interested in further investigating its role during *Drosophila* oogenesis. Given its potential as a post-transcriptional regulator of *oskar* and *gurken* mRNAs, which encode crucial factors for embryonic axis patterning. While several studies have demonstrated a role for Shep in the nervous system of *D. melanogaster*, surprisingly, little is known about its molecular function, specifically the mechanisms through which it influences post-transcriptional regulation.

The *C. elegans* ortholog of Shep, suppressor-26 (Sup-26), plays a crucial role in post-transcriptional regulation of the sex determination pathway and is also required for dendrite morphogenesis of sensory neurons (Mapes *et al.* 2010; Schachtner *et al.* 2015). Sup-26 promotes male development by controlling the expression level of Tra-2 protein (Mapes *et al.* 2010). It does so by directly binding, through its RNA-binding domain, to the *cis*-regulatory elements found on the 3'UTR of *tra-2* mRNA, thereby repressing its translation. Additionally, the same study demonstrated that Sup-26 directly associates with poly(A)-binding protein 1. The interaction between these proteins hints towards a potential role for Sup-26 as an inhibitory PABP-interacting protein that antagonizes the translation-inducing activity of poly(A)-binding protein 1 (PABP1) at the 3'UTR. In the

nervous system, the loss of *sup-26* expression in sensory neurons results in a significant reduction in the number of dendritic termini branching (Schachtner *et al.* 2015). These studies highlight the importance of post-transcriptional regulation in mediating development, and both cell function and morphology. A similar mechanism for sex determination has been described for Shep in the small planktonic crustacean *Daphnia magna*, commonly known as water flea. In *D. magna*, Shep associates with the 5'UTR of *doublesex1*, a male-determining gene, and suppresses its translation (Perez *et al.* 2021). Both these non-*Drosophila* orthologs of Shep provide interesting insights into the diverse processes by which *Drosophila* Shep can translationally repress its target mRNAs, either by binding to their 5'UTR or 3'UTR.

The mammalian orthologs of Shep, RBMS1–3, have also been demonstrated to play a role in mRNA stability and translational regulation. They have been extensively studied in the context of tumorigenesis (Sun *et al.* 2018; Block *et al.* 2021). For example, in breast cancer, Sun *et al.* (2018) showed that RBMS2 functions as a tumour suppressor by positively regulating the expression of cyclin-dependent kinase inhibitor 1 through mRNA stabilization, thereby halting cell proliferation. Moreover, RBMS3 plays a pivotal role in mouse pancreatic development by binding to the 3'UTR of *ptfla* mRNA, which is a key pancreatic transcription factor, thereby facilitating its translation (Lu *et al.* 2012). Interestingly, RBMS3 was sufficient to induce epithelial-to-mesenchymal transition (EMT) in two immortalized mammary epithelial cell lines (Block *et al.* 2021). Given this, it would be interesting to determine whether Shep is involved in EMT within border cell migration during later stages of *Drosophila* oogenesis.

1.6 Aims & Objectives

In *Drosophila*, the major body axes are specified early in development and are governed by a combination of developmental, genetic, and molecular mechanisms. Specifically, RNA-binding proteins play a crucial role in controlling gene expression at the post-transcriptional level during oogenesis. They regulate mRNA turnover and oversee the asymmetric localization and spatial-temporal translation of key maternal mRNAs, which encode important determinants for embryonic axis patterning, such as Bicoid, Oskar, Nanos, and Gurken.

Previous research by Besse *et al.* (2009) demonstrated a role for PTB as a translational repressor of *oskar* mRNA. Interestingly, our lab's yeast two-hybrid screen identified Shep as a potential partner of PTB (López de Quinto, unpublished). Therefore, we hypothesized that Shep in the oocyte could contribute to the post-transcriptional regulation of *oskar* and possibly other key maternal mRNAs. Although Shep is known to play a role in regulating the gypsy chromatin insulator complex and neuronal development within the nervous system, its role during oogenesis remains to be characterized. Despite its established function in the nervous system, the molecular mechanisms underlying Shep's post-transcriptional regulation of these processes in the cytoplasm remain largely uncharted. However, when considering its non-*Drosophila* orthologs, it can be postulated that Shep has the potential to regulate mRNA stability and translation.

The primary aim of this project is to characterize the expression pattern of *D. melanogaster* Shep in ovarian tissues at both the mRNA and protein levels, as well as to explore its biological function during oogenesis. To achieve these general aims, the following specific objectives were set:

- 1) Characterize the expression patterns of the *shep* gene in the ovary at both the mRNA and protein levels in order to (i) identify the mRNA and protein isoforms expressed, (ii) to elucidate the subcellular localization pattern of different Shep polypeptides, and (iii) explore the mechanisms driving *shep* expression in the ovary.
- 2) Evaluate the cytoplasmic role of Shep in the post-transcriptional regulation of key maternal mRNAs during oogenesis, which guide the establishment of the embryo's major body axes.
- 3) Identify and characterize Shep interactions with mRNP complexes during oogenesis, and to confirm its association with the RNA-binding protein PTB *in vivo*.

∞ Chapter II ∞

Materials &
Methods

2 Chapter II – Materials & Methods:

2.1 Standard Molecular Biology Techniques

To address the aims and objectives of this research project, several new recombinant DNA constructs were generated using standard techniques of molecular biology, including general cloning procedures. Refer to Appendix 4 for a list of newly generated recombinant constructs. The experimental procedures used to generate these constructs are outlined in this section. Sambrook and Russell (2001) provides a comprehensive description of the standard molecular biology procedures and techniques used in this project, and served as a primary reference throughout this study.

2.1.1 Cloning Vectors

In genetic engineering and molecular biology research, cloning vectors are circular double-stranded DNA molecules which act as carriers for foreign DNA fragment to generate new recombinant constructs. Cloning vectors contain three key features: (1) an origin of amplification that allows for replicating multiple identical copies of the original DNA, (2) a selectable marker that enables identification and selection of host cells that have taken up the vector, and (3) a multiple cloning site (MCS) that contains unique restriction enzyme recognition sites to allow for the insertion of foreign DNA fragments (Craig *et al.* 2021). The main application of the expression vectors used in this study were the generation of transgenic flies, *in vitro* transcription, and bacterial protein expression.

The modular Gal4-UAS system was utilized to achieve conditional expression of transgenes in *Drosophila* with a temporal and spatial resolution (Brand and Perrimon 1993). Two modified variations of the Gal4-responsive expression vector (i.e., UAS-based) were used to drive expression of transgenic constructs: pTIGER- and pWalium-based vectors, both of which contain an ampicillin resistance gene as a selection marker (see Appendix 3). To achieve germline expression, Rørth (1998) constructed the pUASp expression vector by fusing 14 copies of the UAS activator to the P-transposase promoter, a germline compatible promoter derived from a P-element transposon that is naturally active in the female germline. Downstream of the MCS, this vector also

contains the 3'UTR and terminator regions of the maternally expressed gene, K10. The pTIGER vector (Targeted Integration Germline Expression UAS Regulated) was derived from the pUASp vector and modified to contain an attB attachment site: a sequence required for plasmid integration into the *Drosophila* genome at genomic attP landing sites by the PhiC31 integrase method (Ferguson *et al.* 2012).

Moreover, a series of pValium vectors (Vermilion-AttB-Loxp-Intron-UAS-MCS) were constructed for expression of RNAi hairpin constructs to downregulate the expression of genes using an *in vivo* Gal4-induced approach (Ni *et al.* 2008; Ni *et al.* 2009; Ni *et al.* 2011). Derivatives of the pValium vectors, generated by Perkins *et al.* (2015), were used for the cloning of TRiP hairpin transgenic constructs. Specifically, pWalium20 was used for effective expression in the soma, while pWalium22 was utilized for the germline. The pWalium-based vectors contain the mini-white gene as a selectable marker, an attB site for targeted integration into the *Drosophila* genome, 10 copies of UAS organized into two pentamers, a core promoter, a bipartite multiple cloning site flanking a *white* intron which has shown to reduce toxicity in bacteria, and an intron of the *ftz* gene followed by an SV40 poly(A) tail to facilitate the processing of the hairpin RNA. To induce effective expression in their respective tissues, pWalium20 and pWalium22 rely on different core promoters, namely Hsp70 and P-transposase promoters, respectively.

For *in vitro* synthesis of RNA probes, the pBluescript-II phagemid vector with the MCS in the KS orientation was used (pBSII-KS; Stratagene, 212207). In addition to the ampicillin resistance gene, which serves as a bacterial selection marker, the vector contains T7 and T3 promoters flanking the MCS to allow for the production of either sense or antisense RNA transcripts, corresponding to the respective strands the RNA is transcribed from (see Appendix 3 for the vector map).

2.1.2 DNA Templates

Various sources of DNA were used as template for different aspects of this project, including the generation of new recombinant DNA constructs, gene expression analysis, and *in vitro* synthesis of RNA probes. Examples of these templates include already

available constructs, commercial expressed sequence tag clones (EST), and cDNA libraries derived from a range of *D. melanogaster* tissues.

2.1.2.1 Existing Expression Constructs

Constructs previously generated by SLQ lab (i.e., our lab) or provided by colleagues and collaborators served as DNA templates for cloning. These constructs were either subjected to restriction enzyme digestion to isolate specific DNA fragments or used directly as templates for DNA amplification (see **Table 2.1**).

Table 2.1: List of recombinant DNA construct used in this project.

Clone	Associated Gene	Application	Ref.
pCRII-TOPO-PTB plasmid	PTB	Digestion of PTB CDS	(Besse <i>et al.</i> 2009)
pT7CFE-CHisU-PTB plasmid	PTB	Digestion of PTB CDS	(Besse <i>et al.</i> 2009)
pBSII-mCherry-nostop plasmid	mCherry (mCh)	Digestion of mCh CDS with no stop codon	(López de Quinto, unpublished)
pUAS _t -attB-Su(var)2-10 plasmid	Su(var)2-10	Digestion of Su(var)2-10 CDS	(Mitchell and Taylor, unpublished)

2.1.2.2 Publicly Available cDNA Clones

For convenience, sequenced full-length EST clones from the Berkeley *Drosophila* Genome Project (BDGP) were purchased through the *Drosophila* Genomics Resource Center (DGRC; <https://dgrc.bio.indiana.edu/Home>; see **Table 2.2**). These EST clones were affixed to Whatman® FTA® discs, and the DNA was subsequently extracted from the discs following the protocol described below.

Resuspension of DNA from Whatman® FTA® Discs: To resuspend the DNA from Whatman® FTA® Discs, the discs were quickly rinsed in 50µL of nuclease-free 1× Tris-EDTA (TE; 10mM Tris-HCl, 1mM EDTA, pH 8.0; Fisher Bioreagents, BP2473) by repetitive pipetting for a few times to wash off any chemicals on the Whatman® FTA® disc that could potentially inhibit bacterial transformation. The DNA was then resuspended in 50 µL of TE and allowed to incubate at room temperature for a few hours or preferably overnight with occasional mixing by vortexing. Subsequently, 1µL

of DNA was transformed into DH5 α competent cells (see Section 2.1.10 for transformation protocol). The discs and resuspended DNA were stored separately and kept at -80°C for long-term storage.

Table 2.2: *Drosophila melanogaster* DNA clones purchased from DGRC.

Associated Gene	EST Clone	Associated ORF	DGRC Stock #	EST FlyBase ID
Shep (FBgn0052423)	LD40028	A & C	12698	FBcl0165851
	RH63980	E	10969	FBcl0246748
Su(var)2-10 (FBgn0003612)	RE73180	D	18711	FBcl0230143

2.1.2.3 Generation of Ovarian cDNA Library

Total RNA Extraction from Adult Female *Drosophila* Ovaries: Total ovarian RNA was purified from approximately 10-20 pairs of freshly dissected ovaries or those preserved at -20°C in RNA stabilization solution, RNeasy[®] (Sigma-Aldrich, R0901). The RNA was then purified according to the manufacturer's instructions using either the GeneJET RNA Purification kit (Thermo Scientific, K0731) with slight modifications to the 'Insect Total RNA Purification' protocol, or the EZ-10 Spin Column Animal Total RNA Miniprep kit (Bio Basic, BS82312) following a slightly adjusted standard protocol. The modifications to the procedure included modifying the Lysis buffer composition and method of homogenization. When using the GeneJET kit, the Lysis buffer was supplemented with β -mercaptoethanol instead of DTT. Ovaries were homogenised by 30 strokes of an Eppendorf pestle in RNase-free 1.5mL microcentrifuge tubes. The pestle was cleaned prior to homogenization with either RNaseZap[®] Solution (Ambion, AM9780) or with an alternative, RNase AWAY[™] Surface Decontaminant (Thermo Scientific[™], 7002).

To elute the purified RNAs, 25 μL of RNase-free water was added directly to the centre of the column membrane and incubated for 2 minutes at room temperature prior to centrifugation. To optimise the yield of RNA purification, the elution step was repeated by passing the eluate (i.e., flow-through from the RNA elution step) back into the same column for the second time. The eluate (i.e., purified RNA) was recovered in an RNase-free microcentrifuge tube and stored at -20°C for long-term storage.

Removal of Genomic DNA (gDNA) from RNA Preparations: Following from the previous step, RNase-free DNase I (1U/μL; Thermo Scientific, EN0521) was used to remove gDNA from RNA preps, according to the manufacturer’s instructions. In an RNase-free microcentrifuge tube, a 10μL reaction containing 1μg of purified RNA was prepared, gently mixed, and spun down prior to 30 minutes incubation in a heat block at 37°C. The genomic DNA removal reaction was terminated by adding 1μl of 50mM EDTA and incubating at 65°C for another 10 minutes. The reaction volume was scaled up or down depending on the amount of RNA needed for downstream applications.

Synthesis of First-Strand cDNA Template: The UltraScript 2.0 cDNA Synthesis kit (PCR Biosystems, PB30.31) was used to synthesize cDNA libraries from total RNA, following the manufacturer’s instructions. A 20μL reaction mixture containing oligo(dT), random hexamers, and 1μg of RNA template was prepared in PCR tubes on ice, properly mixed by pipetting, and briefly spun down before incubation in a thermal cycler (Bio-Rad MJ Mini Personal, PTC1148C) at 50°C for 30 minutes. The cDNA synthesis reaction was halted by incubating at 95°C for an additional 10 minutes. The newly synthesized ovarian cDNA library was then stored at -80°C for long-term storage.

2.1.2.4 Non-Ovarian cDNA Template

Non-ovarian total RNA or their cDNA libraries, such as adult testis and 3rd instar Larval Wing imaginal disc and brain, were obtained from colleagues in the department to serve as positive controls for *shep* gene expression analysis (see **Table 2.3**). In cases where total RNA was obtained, synthesis of first-strand cDNA template was generated following the same protocol described for ovarian cDNA library (see above).

Table 2.3: List of *D. melanogaster* non-ovarian cDNA templates used in this project. †: RNA purification was carried out under RNase-free conditions to avoid RNA degradation. Asterisks indicate colleagues at Cardiff University in Helen White-Cooper’s (*) or Fisun Hamaratoglu’s lab (**).

Source of Library	Sample obtained	Brief Description of Procedure [†]	Ref.
Adult Male Testis	cDNA sample from <i>w¹¹¹⁸</i> control flies	Total RNA was purified from 10 testis using RNAqueous™ Micro Total RNA Isolation kit (Invitrogen™, AM1931). First-strand cDNA was synthesised using	Saurabh Chaudhary*

		GoScript™ Reverse Transcription System (Promega, A2790)	
3rd instar Larval Wing Imaginal Disc (WID)	Total RNA sample from <i>yw</i> control flies	Total RNA was purified from either 40 WIDs or 10 brain tissues using RNAqueous™ Micro Total RNA Isolation kit (Invitrogen™, AM1931).	Jack Bruton & Hoi Ping Weeks* *
3rd instar Larval Brain			

2.1.2.5 Genomic DNA extraction from Adult Ovaries

For gDNA extraction, 5–10 pairs of dissected ovaries were washed twice: once in PBS, followed by a wash in Lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM KCl, 2.5 mM EDTA, 0.45% NP-40, 0.45% Tween-20). These ovaries were then homogenized in 100 µL of Lysis buffer, to which 1 µL of Proteinase K (20 mg/mL; ThermoScientific, EO0491) was added. Aspirate multiple times using a P200 pipette for thorough homogenization. To aid digestion, samples were periodically vortexed and then incubated overnight at 55°C. On the following day, the reaction was terminated by heating at 65°C for an additional hour before centrifugation at 13,000 × g for 5 minutes. The supernatant, containing the gDNA, was recovered in a clean nuclease-free tube and stored at -20°C for long-term storage.

2.1.3 Polymerase Chain Reaction

This section outlines the different polymerase chain reaction (PCR) protocols used for various applications, including amplification of DNA, screening for the presence of transgenic inserts, or gene expression analysis. PCR conditions were adjusted for optimal amplification based on the DNA template, DNA polymerase, primer pair, and length of amplicon. Furthermore, reaction volumes were accordingly scaled based on the amount of DNA required for downstream applications. All reactions were prepared on ice in nuclease-free 0.2 mL PCR tubes (Greiner Bio-One International GmbH, 683201). It is worth noting that the annealing temperature and duration of the elongation phase for each PCR reaction were adjusted depending on the primer pair used and the length of the amplicon, respectively.

2.1.3.1 Endpoint PCR (also referred to as Standard PCR)

For general cloning, DNA fragments were PCR amplified using VeriFi™ DNA polymerase (PCR Biosystems, PB10.43), according to the manufacturer's instructions. A 25µL reaction mixture containing 10-20ng of template was prepared, properly mixed by pipetting, and briefly spun down. PCR reactions were carried out in a thermal cycler (Bio-Rad MJ Mini Personal, PTC1148C) following a sequence of thermocycler cycling conditions recommended by the manufacturer. Subsequently, amplicons were purified using GeneJET PCR Purification kit (Thermo Scientific, K0701) following the manufacturer's instructions, before being used for downstream applications.

2.1.3.2 Colony PCR

As a screening strategy, diagnostic colony PCR reactions were performed using Taq DNA Polymerase (PCR Biosystems, PB10.13), according to the manufacturer's instructions. A 10µL reaction mixture was prepared, properly mixed by pipetting, and briefly spun down before a small amount of bacterial cells was added directly from the plate using a sterile pipette tip and vigorously mixed. PCR reactions were carried out in a thermal cycler (Bio-Rad MJ Mini Personal, PTC1148C) following a sequence of thermocycler cycling conditions recommended by the manufacturer with a slight modification. Instead of the usual one-minute initial degradation step at 95°C, the incubation period of this step was extended to 5 minutes to facilitate the disruption of the bacterial cell walls and plasma membranes. Subsequently, amplicons were run on an agarose gel to check for the absence or presence of a desired insert (see Section 2.1.5 for agarose gel electrophoresis procedure).

2.1.3.3 Reverse Transcription PCR

For semi-quantitative gene expression analysis, a two-step reverse transcription PCR (RT-PCR) was performed using Ultra DNA polymerase (PCR Biosystems, PB10.32) or redTaq DNA Polymerase (PCR Biosystems, PB10.13), according to the manufacturer's instructions. A 25µL reaction mixture containing <100ng of cDNA template was prepared, properly mixed by pipetting, and briefly spun down. PCR reactions were carried out in a thermal cycler (Bio-Rad MJ Mini Personal, PTC1148C)

following a sequence of thermocycler cycling conditions recommended by the manufacturer. Subsequently, amplicons were run on an agarose gel to check for the expression of a desired transcript(s).

For detailed information regarding the RNA immunoprecipitation assay followed by RT-PCR, please refer to Section 1.4.3.

For expression analysis of *shep* transcripts (-A, -B, -D, -E, -F, -G, and -H) in wildtype control ovarian extract, first-strand cDNA was obtained from w^{1118} ovaries, as described in Section 2.1.2.3. Following this, RT-PCR reactions were prepared using the Ultra DNA polymerase mix containing 80 ng/ μ L of template DNA and specific primer pair combinations for a total of 30 PCR cycles. Refer to Appendix 1 for a list of the primers used in the identification of expressed transcripts and to Appendix 1.1 for their approximate annealing sites. Each primer pair combination was optimized via gradient PCR, a process aimed to determine the optimum annealing temperature and identify the presence of non-specific amplification. If a particular transcript was not expressed in the ovaries, as indicated by the lack of amplification, then cDNA template obtained from a non-ovarian total RNA was used as a positive control (see Section 2.1.2.4).

Unlike the other *shep* transcripts, nested RT-PCR had to be employed to determine the expression of transcript 'I'. In the initial part of nested RT-PCR, transcripts 'H' and 'I' were selectively amplified from a 25 μ L of PCRBIO Ultra reaction that contained 10ng of first-strand cDNA template and a transcript-specific primer pair (as indicated in Appendix 1). This was followed by a series of PCR reactions, using the previously amplified DNA as the template. Subsequent DNA gel extractions were performed to specially enrich both transcripts 'H' and 'I' and dilute the cDNA template, as illustrated in **Figure 2.1**. In the second phase of the nested RT-PCR, transcript 'I' was amplified using the enriched and gel-extracted DNA template, accompanied by several control reactions. These controls included a no DNA template control and a test for single-stranded DNA amplification, wherein a reaction composed of all PCR components except for the forward primer was run. The resulting RT-PCR amplicons were analysed through electrophoresis on a 1% agarose gel, as detailed in Section 2.1.5.

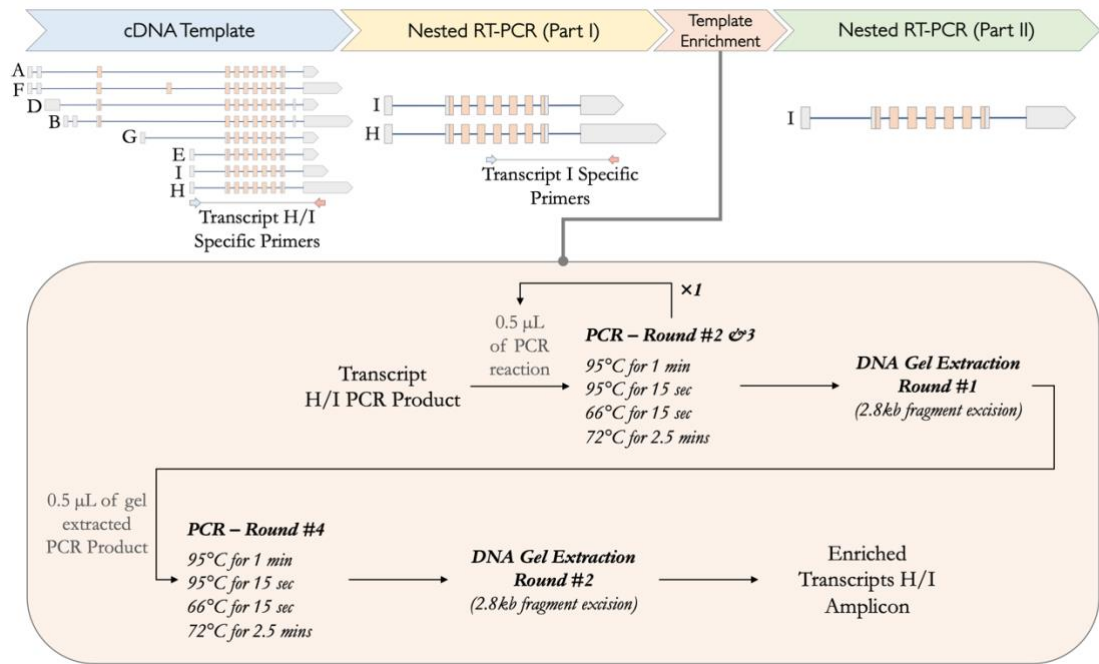


Figure 2.1: Workflow for analyzing transcript I expression in the ovary. When analyzing the expression of transcript I, nested RT-PCR had to be employed. In the initial step of the nested RT-PCR, transcripts H and I were selectively amplified from an ovarian cDNA library using specific primer pairs within a 25 μ L PCR BIO Ultra reaction (see Appendix 1 for primer details). To enrich the specific transcripts of interest and dilute the ovarian cDNA template library, a series of endpoint PCR reactions and DNA gel purifications were conducted, mirroring the conditions of the initial nested RT-PCR step. After enriching transcripts H and I, the second step of nested RT-PCR was carried out using a different set of primers that are now specific to transcript I (refer to Appendix 1 for primer details). The resulting PCR amplicons were then visualized on a 1% agarose gel.

2.1.3.4 Quantitative Reverse Transcription PCR

To quantify relative expression levels of mRNAs in different samples, quantitative reverse transcription PCR (RT-qPCR) was carried out using qPCR BIO SyGreen Blue Mix Lo-ROX (PCR Biosystems, PB20.15). A 10 μ L reaction mixture was prepared on ice according to the manufacturer's instructions containing 3 μ L of diluted cDNA template (in HPLC grade water) at a final concentration of 21ng. Optimized primer pairs designed to assay target and reference (i.e., housekeeping) genes/transcripts are listed in Appendix 1. Unlike other reference genes available at the lab, *rp49* was the most appropriate housekeeping gene given that its expression levels were stable across different experimental samples and repeats. For each analysed transcript, standard controls were performed with no template control (NTC) or just RNA template not treated with reverse transcriptase to control for extraneous DNA and gDNA contamination, respectively. Technical triplicates were prepared for each biological

replica (including standard controls). Samples were loaded to a 96-well PCR microplate (StarLab, I1402-9909) and sealed using a polyolefin film (StarLab, E2796-9795) before qPCR reactions were run on LightCycler®96 (Roche). Thermalcycler cycling parameters were set up according to the manufacturer's guidelines which included the recommended melt curve analysis after the completion of amplification cycles to assess amplicon homogeneity. After Ct values were obtained (also known as quantification cycle, Cq), relative levels of target transcripts were calculated according to the 'Livak' method (also known as $2^{-\Delta\Delta Cq}$). The analysis was applied to normalize the data against a reference gene and to determine the relative expression level of target transcripts in test samples relative to the control sample. Normalized expression ratios (i.e., $2^{-\Delta\Delta Cq}$ value) were plotted as a bar chart in GraphPad and tested for statistical significance (see Section 2.8). Error bars represent the mean \pm SEM values derived from three independent biological replicates.

2.1.3.5 Purification of PCR Amplicons

For specific applications, PCR fragments were purified from PCR reactions using the GeneJET PCR Purification kit (Thermo Scientific, K0701). Although the manufacturer's instructions were followed, minor modifications to the elution procedure were made to increase DNA yield in the eluate, as detailed below.

Elution: To maximize the yield of DNA extracted from the column, purified DNA was eluted by adding 25 μ L of 65°C preheated Elution buffer directly to the centre of the column membrane and incubated for 2 minutes at room temperature before centrifugation. Then, the eluate was passed back into the same column and incubated for 2 additional minutes before centrifugation. Finally, the eluate was recovered and kept at -20°C for long-term storage.

2.1.4 Oligonucleotide & Primer Design

Primers and oligonucleotides were manually designed from desired template DNA sequences. Primers were used to amplify DNA fragments, screen for the presence of inserts, or sequence final constructs. Oligonucleotides, on the other hand, were annealed to generate inserts for cloning. The OligoAnalyzer tool (version 3.1) provided by Integrated DNA Technologies was used to analyse the physical and thermodynamic

properties of both the oligos and primers. The tool was used to assess the length, GC content, melting temperature (T_m), and identify potential formation of secondary structures and dimerization.

Sigma-Aldrich Biotechnology synthesized the designed primers and oligos (see Appendix 1). Lyophilized DNA was reconstituted in nuclease-free water to achieve a final concentration of 100 μ M and were then stored at -20°C for long-term storage. Unless otherwise specified, a working stock concentration of 10 μ M was prepared for use in subsequent downstream applications.

Primer pairs designed for PCR amplification were first tested using a gradient PCR to determine the optimal annealing temperature, maximize the yield of the desired product, and avoid mispriming and nonspecific amplification. Annealing temperature (T_a) of a primer pair was calculated by averaging the predicted melting temperatures of the forwards and reverse primer, then subtracting 5°C from the average. To determine the optimum annealing temperature for a primer pair, a temperature gradient PCR was performed across a range of annealing temperatures set around the calculated T_a (e.g., $\pm 5^{\circ}\text{C}$), as long as it did not exceed the extension temperature. Amplicons were analysed by agarose gel electrophoresis.

For qPCR analysis, primers were designed to span a splice junction (if possible) to avoid or detect amplification from contaminating genomic DNA, and amplify products with a size range of 80-250 nucleotides. All qPCR primer pairs (listed in Appendix 1) were previously validated to ensure their specificity and efficiency. Primers for measuring the relative expression levels of target transcripts were designed by the SLQ lab, except where otherwise stated.

Primer pairs, used for RT-PCR analysis, were designed in accordance with the qPCR primer guidelines, outlined in (Quellhorst and Rulli 2012). Primers for *shotgun* (*shg*; *D. melanogaster* DE-Cad; FBgn0003391) and *nanos* (*nos*; FBgn0002962) on Giuliani *et al.* (2014) and Peng and Gavis (2022), respectively. When possible, these published primers were redesigned with slight alteration to the primer sequence to include a GC clamp at the 3' end.

2.1.5 Agarose Gel Electrophoresis

Agarose gel electrophoresis was employed to analyze nucleic acid products derived from PCR, *in vitro* transcription, and restriction digestion. This technique separates DNA or RNA fragments based on their size by running them through an agarose-based gel using a spatially uniform electric field. In principle, smaller fragments travel farther than larger ones in gel electrophoresis, resulting in distinct bands on the gel for easy identification and analysis.

Preparation of Agarose Gels: Agarose gels at concentrations of 1-2% w/v were prepared by dissolving agarose powder (Fisher BioReagents™, BP1356) in 1×TBE buffer, which contains 89mM Tris-base, 89mM Boric acid, and 2mM EDTA^{pH8} in dH₂O. For gels visualizing DNA fragments, SYBR™ Safe DNA gel stain (Invitrogen, S33102) was added to the molten agarose, according to the manufacturer's instructions. Molten agarose was poured to a gel mould placed in the electrophoresis tank and allowed to set at room temperature for approximately 30-45 minutes. The solidified gel was then submerged in 1×TBE buffer before loading the ladder and samples.

Sample Preparation: To facilitate loading and visualization of DNA or RNA samples onto the gel, 12µL of a DNA sample or 10µL of an RNA sample was prepared, containing DNA loading dye (Thermo Scientific, R0611) or RNA loading dye (Thermo Scientific, R0641), respectively. Additionally, the loading dye aids in visualizing the samples during loading and enables tracking the migration of nucleic acid fragments throughout the electrophoresis. It is worth noting that the RNA loading dye contains ethidium bromide, allowing for direct visualization of RNA bands after electrophoresis when the gel is exposed to ultraviolet or blue light. Alongside nucleic acid samples, a molecular weight standard was loaded with each run, either the GeneRuler™ 1 Kb Plus DNA Ladder (Thermo Scientific, SM1331) or the RiboRuler High Range RNA Ladder (Thermo Scientific, SM1823). To achieve optimal resolution of RNA bands during electrophoresis, both RNA samples and ladder were denatured by heating at 70°C in a heat block for 10 minutes before loading onto the gel.

Electrophoresis Conditions & Gel Visualization: Nucleic acids were electrophorized at a constant voltage of 75 volts for approximately 45 minutes or until bromophenol blue

had migrated to about two-thirds of the gel's length. The SYBRTM Safe DNA stain added directly to the gel facilitate the visualization of DNA bands, while ethidium bromide present in the RNA loading dye assists with the visualization of RNA bands. Following electrophoresis, the gels were photographed in a GelDoc-It® 310 documentation system (from UVP) using UV light. For cloning applications, a Dark Reader® Transilluminator (Clare Chemical Research, DR46B) was used to visualize and excise DNA bands.

2.1.5.1 DNA Gel Extraction

Agarose gel extraction is a commonly used method to purify DNA fragments from agarose gels for downstream applications, such as cloning. The GeneJet Gel Extraction kit (Thermo Scientific, K0691) was used to extract and purify DNA fragments separated by agarose gel electrophoresis. It is important to note that fresh electrophoresis buffer was always used when running DNA fragments for gel purification. While the manufacturer's instructions were followed, the elution procedure of the protocol was slightly modified to increase DNA yield within the eluate, as described below.

DNA Band Excision: After electrophoresis, the gel was visualized under a dark reader transilluminator to locate the DNA fragment of interest. The band of interest was then carefully excised from the agarose gel using a clean scalpel. The excised gel slice containing the DNA fragment of interest was transferred into a nuclease-free Eppendorf tube and weighed using a laboratory precision balances (KERN & Sohn GmbH, EW 220-3NM). To facilitate DNA binding to the purification column in subsequent steps, an equal volume of Binding buffer (provided by the kit) was added to the excised gel slice (e.g., 100µL of buffer to every 100mg gel). The tube was then incubated in a heat block at 60°C for 10 minutes, or until the gel slice was completely dissolved. During incubation, the sample was regularly mixed by vortexing to ensure complete dissolution. The DNA sample was then transferred onto a spin column and processed as per the instructions provided by the kit for subsequent purification.

Elution: To maximize the yield of purified DNA eluted from the column, 25µL of 65°C preheated Elution buffer was added directly to the centre of the column membrane and incubated for 2 minutes at room temperature before centrifugation. The eluate was then

passed back through the same column and incubated for 2 additional minutes before centrifugation. Finally, the eluate was recovered and stored at -20°C for long-term storage.

2.1.6 Restriction Endonuclease Digestion of DNA

Restriction enzyme digestion takes advantage of naturally occurring enzymes that cleave DNA at a specific sequence, making them valuable for cloning or screening applications. In this study, restriction digestion of DNA was carried out using FastDigest® enzymes from Thermo Fisher Scientific engineered for rapid DNA digestion (see Appendix 2 for list of enzymes used and their recognition sites). In a nuclease-free microcentrifuge tube, reactions were prepared using the appropriate substrate DNA and restriction enzyme(s), following the manufacturer's instructions. The reactions were gently mixed, spun down, and then incubated in a heat block at 37°C for 30 minutes to allow for enzymatic cleavage of the DNA at specific recognition sites. Digested DNAs were processed immediately or stored at -20°C for long-term storage. Reaction volumes were accordingly scaled up or down depending on the desired DNA concentration for downstream applications. A brief description of the procedure in terms of reaction volumes prepared, amount of DNA digested, and post-digestion treatments applied are described below based on the downstream application of the digested DNA, such as cloning, *in vitro* transcription, or screening for the presence of insert.

Digestion of Plasmid DNA for Cloning: For cloning applications, 2µg of plasmid DNA was digested in a 40µL reaction. The digested vectors were typically dephosphorylated using the protocol described in Section 2.1.7, then purified using the GeneJET PCR Purification kit from Thermo Scientific, and finally eluted in 25µL of Elution buffer provided by the kit.

Digestion of PCR Fragments for Cloning: For cloning applications, 1µg of amplicon was digested in a 150µL reaction. Amplicons, with exposed phosphate groups at their termini, were purified using the GeneJET PCR Purification kit from Thermo Scientific and then eluted in 25-50µL of the provided Elution buffer.

Digestion of Plasmid DNA for *in vitro* Transcription: For *in vitro* synthesis of RNA probes, 7µg of the plasmid containing the recombinant DNA of interest was linearised in a 150µL reaction using restriction enzymes that generate blunt end or 5'-overhang downstream of the insert. Restriction enzymes that generate 3'-overhang have been reported to generate spurious long transcripts (Schenborn and Mierendorf 1985; Triana-Alonso *et al.* 1995), and were therefore avoided. If necessary, 3'-overhangs can be blunted by T4 DNA polymerase prior to transcription. After linearization, DNA was purified using the GeneJET PCR Purification kit (Thermo Scientific) and then eluted in 25µL of the Elution buffer supplied with the kit.

Diagnostic Digestion of Plasmid DNA: To screen for the presence of the desired insert, up to 1µg of recombinant constructs, purified from overnight bacterial cultures, were digested in a 20µL reaction containing the FastDigest GREEN buffer before analysing digestion products by agarose gel electrophoresis to confirm the presence and size of the expected DNA fragment.

2.1.7 Dephosphorylation of Plasmid DNA

A standard procedure in the cloning workflow is the removal of the 5'-phosphate groups from DNA termini to minimise self-ligation and reduce unwanted background of non-recombinant constructs. Dephosphorylation of digested DNA was therefore performed on digested vectors using FastAP™ Thermosensitive Alkaline Phosphatase (1U/µL; Thermo Scientific, EF0651) according to the manufacturer's instruction. This was achieved by adding 2µL of Alkaline Phosphatase, which is compatible with the FastDigest buffer used for restriction digestion, directly to the reaction containing the digested DNA. The reaction was then allowed to incubate at 37°C for an additional hour. Subsequently, the enzymes were either inactivated by incubating the reaction at 75°C for 5 minutes or removed by purification using the GeneJET PCR Purification kit (Thermo Scientific). Dephosphorylated DNAs were processed immediately for ligation with the desired transgene or stored at -20°C for long-term storage.

2.1.8 DNA Ligation

To generate recombinant constructs, DNA inserts with cohesive ends were ligated into digested vectors using T4 DNA ligase provided by the Rapid DNA Ligation kit (5U/μL; Thermo Scientific, K1422), according to the manufacturer's instructions. A 5μL ligation reaction was prepared on ice containing 50ng of digested vector and insert DNA at a 3:1 molar ratio excess over the vector (see equation below). The mixture was then gently mixed, spun down, and incubated at room temperature for 30–60 minutes to ensure ligation of vector and insert DNA. After the ligation was completed, the transgenic constructs were transformed into chemically competent *E. coli* cells using the protocol outlined in Section 2.1.10. The remaining ligation mixture was stored at -20°C for long-term storage.

Equation to calculate the mass of insert required for ligation:

$$\text{Mass of insert [ng]} = \frac{\text{mass of vector [ng]} \times \text{Length of insert [bp]}}{\text{length of vector [bp]}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}}$$

2.1.9 Bacterial Growth Media

Luria-Bertani (LB) media is a nutrient-rich bacterial growth medium widely used for the cultivation of various bacterial strains, including *E. coli*. It is routinely used for the preparation of plasmid DNA, recombinant proteins, and bacterial cultures for a variety of research purposes. Ever since Bertani published the formula in 1951 as part of his first paper on lysogeny (Bertani 1951), several formulations of LB have been developed. In general, all formulations of LB contain tryptone, yeast extract, and sodium chloride to provide essential nutrients for bacterial growth, but the amounts and ratios of these components differ between the formulations. The two commonly used formulations are the Miller and Lennox media (see **Table 2.4**). The Lennox formula has a lower salt concentration compared to the Miller growth medium. In this study, the Lennox media was used when cloning very large constructs (>17Kb). On the other hand, the Miller formula was used for routine bacterial growth or for protein expression. Agar was added to the LB formula to make LB-agar plates, which were used to facilitate the isolation and identification of bacterial colonies.

Table 2.4: Formulation of Luria-Bertani media for microbial cultivation. N/A: not applicable.

Media Component(s)	Composition per 1L		
	Agar Plate(s)	Liquid Broth	
		Miller (Typical Recipe)	Lennox (Low-Salt Recipe)
Tryptone	10g	10g	10g
Yeast Extract	5g	5g	5g
Sodium Chloride (NaCl)	10g	10g	5g
Agar	12g	N/A	N/A
Distilled water	Up to 1L	Up to 1L	Up to 1L

Preparation of LB Broth & LB Plates: To prepare LB Miller liquid broth and agar plates, 25g or 37g of pre-mixed granules (Fisher BioReagents™, BP1426 and BP1425) were dissolved in 1L of distilled water, respectively. Similarly, LB Lennox liquid broth was prepared by dissolving 20g of pre-mixed granules (Sigma-Aldrich, L3022) in 1L of distilled water. The medium was then briefly mixed by swirling and sterilized by autoclaving at 121°C for 20 minutes. For selective cultivation and selection of *E. coli*, the autoclaved liquid media were allowed to cool down at room temperature before the antibiotics were added (Table 2.5) under sterile conditions (e.g., in a laminar flow chamber or near a lit Bunsen burner)

Table 2.5: Antibiotic stock & working concentrations for plasmid selection. *: Ampicillin and kanamycin were dissolved in water, while Chloramphenicol was dissolved in absolute ethanol. †: Antibiotic working concentration for both LB plates and liquid broth. ⚡: Antibiotic working concentration for liquid broth to culture sensitive clones.

Antibiotic	Stock Concentration* (mg/mL)	Working Concentration (µg/mL)
Ampicillin (Fisher BioReagents™, BP1760)	100	100 [†] or 50 [⚡]
Chloramphenicol (Fisher BioReagents™, BP904)	25	34 (for LB plates) 63 (for LB broth)
Kanamycin (Fisher BioReagents™, BP906)	50	50 [†]

Whereas for the agar plates, the autoclaved molten agar mixture was only allowed to briefly cool for 15-30 minutes or until it was just comfortable to handle (at approximately 60°C) before antibiotics were added (Table 2.5), to avoid the agar from fully solidifying. The media were then gently swirled to avoid the formation of bubbles,

especially in the agar-containing media. Under a laminar flow cabinet, the molten agar media was immediately poured into sterile 90mm diameter Petri dishes, ensuring even distribution at the bottom of the plate. The plates were left to solidify at room temperature before being inverted and stored at 4°C, along with the liquid broth.

2.1.10 Bacterial Transformation

Non-pathogenic strains of *E. coli*: In this research project, various strains of chemically competent *E. coli* cells were used based on the specific application. For routine cloning, the MAX Efficiency™ DH5 α (Invitrogen, 18258012) was the strain of choice. Meanwhile, the High Efficiency DH10 β strain (NEB, C3019H) was selected for the cloning of complex and large constructs (>17Kb). Lastly, the BL21(DE3) *E. coli* strain (NEB, C2527I) was used for protein expression.

Preparation of Chemically Competent Cells: To prepare competent cells from *E. coli*, bacteria from glycerol stocks were first plated on a non-selective LB-agar plates (see Section 2.1.17), and a single colony was inoculated under aseptic conditions in non-selective LB liquid media (i.e., broth lacking antibiotics) and kept growing for 6-8 hours at 37°C. In a sterile 100mL conical flask, 50mL of fresh culture was prepared by inoculating a non-selective LB broth containing 20mM glucose with 1-5 μ L of the previously prepared liquid culture. The culture was then grown at 18°C with constant shaking at 200 rpm until they reached the exponential growth phase (OD_{600} =0.4 - 0.6). Preparation of chemically competent *E. coli* cells was carried out using the Mix&Go! *E. coli* Transformation kit (Zymo Research, T3001), according to the manufacturer's instructions. In sterile 1.5mL Eppendorf tubes, 50 μ L aliquots of chemically competent cells were flash-frozen in liquid nitrogen and immediately stored at -80°C until needed.

DNA Transformation: For efficient bacterial transformation of plasmid DNA, a 50 μ L aliquot of competent cells was gently handled and thawed on ice for 5-10 minutes. Near a lit Bunsen burner, 1-3 μ L of plasmid DNA or ligation product was added to the thawed cells, which was then gently mixed by flicking the bottom of the tube several times. Before plating, the transformation mixtures undergo different recovery conditions depending on the antibiotic resistance gene carried by the construct. In the case of ampicillin selection, the transformation mixture was simply incubated on ice for

10 minutes before plating. For chloramphenicol and kanamycin selection, however, the transformation mixtures were first recovered on ice for 10 min. Then, the cells were outgrown for 1 hour at 37°C in 2 volumes of non-selective LB Broth, with or without 20mM glucose, while being under constant shaking at 200 rpm. This recovery step helps with cell growth and the build-up of antibiotic resistance before plating the transformants on selective plates.

Plating of Transformants: LB agar plates designated for either selective or non-selective selection were taken from storage at 4°C and allowed to warm in a 37°C incubator for about an hour. Near a lit Bunsen burner, the transformation mixtures were pipetted onto the appropriate agar plates and spread evenly. The plates were then inverted and incubated at 37°C for roughly 16-18 hours (overnight), unless specified otherwise. For large recombinant constructs, especially those over 17 Kb in size, the plates were incubated at 30°C for a minimum of 24 hours.

2.1.11 Inoculation of Bacterial Overnight Culture

Overnight cultures were grown for several applications, such as colony screening for the presence of an insert, plasmid DNA preparation, or recombinant protein expression. Depending on the downstream application, different culture volumes were prepared near a lit Bunsen burner using pre-warmed LB liquid media, containing appropriate selective antibiotic, if necessary. To screen colonies for DNA inserts, 2-5mL of LB medium was added to sterile 13mL bacterial culture tube (Sarstedt, 62.515.006). For plasmid DNA purification, 20-100 mL of LB medium was added to an autoclaved 100-250mL Erlenmeyer flask. Unless otherwise stated, a single random bacterial colony was directly transferred to the liquid LB media using a sterile inoculating loop. The loop was then shaken vigorously to ensure complete bacterial transfer. In a shaking incubator, cultures were grown overnight at 37°C with continuous shaking at 250 rpm. On the following day, overnight cultures were harvested by centrifugation at 5,000 × g for 10-20 minutes. The bacterial pellet was collected without the supernatant and then immediately used for downstream applications or stored at -20°C for long-term storage.

2.1.12 Purification of Plasmid DNA

Various protocols and kits were followed to purify plasmid DNA from bacterial overnight cultures for uses in downstream applications. Regardless of the protocol used, kits utilize the alkaline lysis method to isolate plasmid DNA from *E. coli* chromosomal DNA (Birnboim and Doly 1979). For applications such as colony screening and plasmid DNA amplification, the ‘Standard Plasmid DNA Miniprep’ protocol was followed using either EZ-10 Spin Column Plasmid DNA Miniprep kit (Bio Basic, BS414) or Hybrid-Q™ Plasmid Rapidprep kit (GeneAll®, 100-102). However, when purifying plasmid DNA for microinjections for fly transgenesis, the ‘Endotoxin-Free Plasmid DNA Miniprep’ protocol was preferred, as it helps embryo survival during the germline transformation procedure by minimizing potential endotoxin contamination. The E.Z.N.A.® Endo-Free Plasmid DNA Mini kit (Omega Bio-Tek, D6950) was used for routine constructs, while the PureLink™ HiPure Midi Plasmid DNA Purification kit (Invitrogen, K2100-04) was used for complex and large constructs (>17Kb). The protocols were followed according to the manufacturer’s instructions, with minor modifications made to the elution procedure when using the silica-based columns provided by the miniprep kits.

Elution: To maximize the yield of purified plasmid DNA eluted from the silica-based columns, 25µL of 65°C preheated Elution buffer was added directly to the centre of the column membrane and incubated for 2 minutes at room temperature before centrifugation. Then, the eluate was passed back into the same column and incubated for 2 additional minutes before centrifugation. Finally, the eluate was recovered and stored at -20°C for long-term storage.

2.1.13 DNA Sanger Sequencing of Constructs

All recombinant constructs generated or used in this project were sequenced by Eurofins Genomics (Ebersberg, Germany) before their use in any of the downstream applications. This sequencing step not only confirms the clone identity but also ensures that fusion proteins are in-frame, with no occurrence of nucleotide substitutions, insertions, or deletions. Primer used for sequencing are listed in Appendix 1. Samples

of plasmid DNA were prepared for Sanger sequencing using the TubeSeq Service, following the recommended DNA and primer concentrations.

Sample Preparation of Plasmid DNA for Sequencing: In a 1.5mL Eppendorf tubes, a 15µL sample was prepared containing 65-100 ng/µL of purified plasmid DNA in double distilled water. In case the primers were not provided as part of Eurofins Genomics ‘Standard Primers’, 2µL of 10µM primer was added to each tube, resulting in a total volume of 17µL premixed sample.

2.1.14 *Drosophila* Transgenesis – Germline Transformation of Transgenic Constructs

The PhiC31 (ϕ C31) integrase system was used to generate transgenic fly stocks, expressing a transgene in a tissue-specific manner. The system utilizes ϕ C31 integrase to enable site-specific integration of transgenic constructs into the *Drosophila* genome by mediating unidirectional site-specific recombination between recognition sites, namely bacterial attachment site (attB) and a phage attachment site, attP (Groth *et al.* 2004; Bischof *et al.* 2007; Markstein *et al.* 2008; Knapp *et al.* 2015). All newly generated constructs bear an attB attachment site, allowing for their integration into genomic attP landing sites previously inserted into the fly genome (see **Table 2.6**). Transgenic constructs were sent for microinjection to the University of Cambridge Fly Facility (England, UK) through their *Drosophila* embryo Microinjection Service: phiC31 integrase-mediated transgenesis plan 5. To improve embryo viability, plasmid DNA purified from endo-free miniprep kits was used to prepare samples for microinjection into *Drosophila* embryos. A 20µL sample of plasmid DNA, with a final concentration between 400-500 ng/µL, was prepared in a 0.5mL microcentrifuge tube and diluted with HPLC grade water if necessary.

Table 2.6: attP landing sites used for site-specific integration of transgenic constructs. * The letter ‘L’ stands for ‘left’ and indicates the chromosomal arm where the attP landing site is located. The chromosomal location of attP landing sites were obtained and summarised from Markstein *et al.* (2008) and van der Graaf *et al.* (2022).

Landing Site	Chromosomal location *	Cytology location	Location relative to nearest genes
attP2	3L	68A4	Intragenic: <i>MocsI</i>
attP40	2L	25C7	Intragenic: <i>Msp-300</i>

To confirm the integration of transgenic constructs into the *Drosophila* genome, genomic DNA was extracted from flies of an established transgenic stock (refer to Section 2.1.2.5). A 10 μ L Taq DNA polymerase PCR reaction (PCR Biosystems, PB10.13) was then performed using transgene-specific primers and the extracted gDNA as a template for amplification. The resulting PCR product was then analysed using agarose gel electrophoresis to verify the insertion and presence of the transgene.

2.1.15 Spectrophotometric Quantification of Nucleic Acid

Concentration of DNA or RNA molecules was determined by measuring the optical density (OD) using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, v3.7), following the protocol described in Desjardins and Conklin (2010). Briefly, after calibrating the instrument with 1 μ L of an appropriate blank solution (e.g., Elution buffer, TE buffer, etc.), 1 μ L of the sample was pipetted onto the spectrophotometer. The software accompanying the spectrophotometer then calculated the nucleic acid concentration in the sample based on the OD values and known extinction coefficients for DNA or RNA.

For precise and accurate readings, a Qubit assay was used to quantify nucleic acids. Specifically, RNA samples and standards were prepared using the Qubit™ RNA High Sensitivity kit (Invitrogen, Q32855) following the manufacturer's instructions. Concentration measurements were then obtained with the Qubit™ 3.0 Fluorometer (Invitrogen, Q33216) as per the manufacturer's user guide manual.

2.1.16 Quality Assessment of RNA Samples

RNA ScreenTape analysis was used to evaluate the integrity and size range of RNA samples obtained from RNA immunoprecipitation assays (described in Section 2.4.3). This was important to determine the compatibility of the immunoprecipitated RNA for use in downstream RNA-seq experiments. After RNA immunoprecipitation and RNA purification, the concentration of RNA samples was measured using a NanoDrop™ 1000 Spectrophotometer and Qubit assay to determine which RNA ScreenTape analysis to perform: high sensitivity (HS; 0.5-10 ng/ μ L) or broad range (BR; 25-500 ng/ μ L). The HS ScreenTape RNA Ladder (Agilent Technologies, 5067-5581) and RNA samples were prepared in an optical tube (Agilent Technologies, 401428) with

HS RNA ScreenTape buffer (Agilent Technologies, 5067-5580) following the instructions provided by the manufacturer. In a Mastercycler X50a (Eppendorf, 6313000042), RNA ladder and samples were heated to 72°C for 3 minutes and then immediately placed on ice for 2 minutes. RNA ladder and samples were then placed into an automated platform, the 4200 TapeStation system (Agilent Technologies, G2991A), for analysis with the HS RNA ScreenTape (Agilent Technologies, 5067-5579). The instrument was operated using the Agilent TapeStation Controller Software (version 4.1.1) and the raw data obtained from electrophoresis was analysed using the Agilent TapeStation Analysis Software (version 4.1.1).

2.1.17 Bacterial Glycerol Stocks

For long-term storage of *E. coli* strains and transformants, bacterial glycerol stocks were generated from overnight cultures in either selective or non-selective LB medium. The procedure was carried out under sterile conditions and in close proximity to a lit Bunsen burner to ensure sterility and prevent microbial contamination. In a sterile 1.5mL microcentrifuge tube, 800µL of bacterial culture was mixed with 200µL of 87% autoclaved glycerol by gentle inversions. Bacterial glycerol stocks were then immediately stored at -80°C for long-term preservation.

Recovery of Bacteria from Glycerol Stocks: A sterile inoculating loop or a P20 pipette tip was used to scrape off bacteria from the frozen glycerol stock. The bacteria were then streaked onto an appropriate LB agar plate (either selective or non-selective) using the ‘Quadrant Streaking’ method of plating. The plates were subsequently incubated at 37°C overnight.

2.1.18 Chemical-Based Disinfection & Decontamination of Bacterial Cultures

To ensure proper disposal of liquid bacterial cultures, the supernatant of spun cultures and surfaces contaminated by culture spillage were treated with Chemgene HLD₄H disinfectant (StarLab, XTM308) or an alternative suitable biocide, such as Virkon. An adequate volume of disinfectant was added to the supernatant from centrifuged bacterial cultures. This mixture was then allowed to incubate overnight to ensure proper disinfection of bacteria prior to disposal down the laboratory sink. Alternatively, the

Prestige Medical Classic Media autoclave was employed to achieve a complete sterilization by maintaining the bacterial culture at 121°C under pressure for 28 minutes.

For the disposal of solid wastes generated throughout the molecular cloning procedure, waste materials were collected in Sterilin™ biohazard plastic waste bags (Thermo Scientific, 509HT) and were then sterilized by autoclaving. These solid wastes include Petri dishes, inoculation loops, culture tubes, and other contaminated laboratory consumables.

2.2 Molecular Cloning – Generation of New Recombinant Constructs

As part of this project, several new constructs were generated containing different transgene fragments for a range of downstream applications (see Appendix 4). These constructs can be categorized into three main groups: constructs intended for generating transgenic flies, plasmids required as templates for *in vitro* transcriptions, and plasmids specifically designed for protein expression. While a detailed description of the cloning procedure is provided in the following section, it is important to highlight that DNA templates such as *D. melanogaster* ovarian cDNA libraries, pre-existing constructs, or EST clones were typically used for PCR amplification. Additionally, the sequence of interest was amplified using primers designed to introduce restriction sites compatible with those present in the multiple cloning site of the cloning vectors. In other instances, the fragment of interest was directly excised from pre-existing constructs or EST clones and then cloned into the MCS of the vectors. The specifics pertaining to the DNA templates are elaborated in Section 2.1.2. A complete list of the primer pairs and vector maps used in this project can be found in Appendix 1 and 3, respectively.

2.2.1 Constructs to Generate Transgenic Fly Stocks

2.2.1.1 Generation of shep TRiP Hairpin Constructs for in vivo RNAi Knockdown Analysis

The *shep* RNAi hairpin was designed based on a protocol published by Jian-Quan Ni and Norbert Perrimon for the Transgenic RNAi Project (TRiP; Perkins *et al.* 2015) at Harvard Medical School. The 71 nucleotides long hairpin oligos were designed based on the miR1 scaffold, where the top oligo consists of two parts, namely the passenger strand and guide strand, forming a miRNA duplex. These two strands are separated by

the sequence “tatgcttgaatataacta”, as shown in **Figure 2.2**. The passenger strand refers to the sense sequence common to all splice RNA forms, while the guide strand is complementary to the passenger strand, which incorporates into the RNA-induced silencing complex (RISC) and silences gene expression. The bottom oligo, however, is designed to be the reverse complement to the top oligo.

For cloning purposes, the oligos were designed to contain overhangs of *NheI* and *EcoRI* restriction sites at the 5' end of the passenger strand or 3' end of the guide strand, respectively, as depicted in **Figure 2.2**. The top oligo contains the sequence “ctagcagt” at the 5' end of the passenger strand and the sequence “gcg” at the 3' end of the guide strand, whereas the bottom strand contains the sequence “aattcgc” at the 5' end of the passenger strand and the sequence “actg” at the 3' end of the guide strand. The sequences used for the top and bottom oligos of the *shep* RNAi hairpin were designed based on the TRiP stock ‘GL00659’ (refer to Appendix 1 for oligo sequences), primarily because they had been previously designed and validated to target all *shep* transcripts while avoiding unspecific binding to off-targets. The sequences of the top and bottom oligos were obtained from the RSVP database, accessible at https://www.flyrnai.org/cgi-bin/RSVP_search.pl.

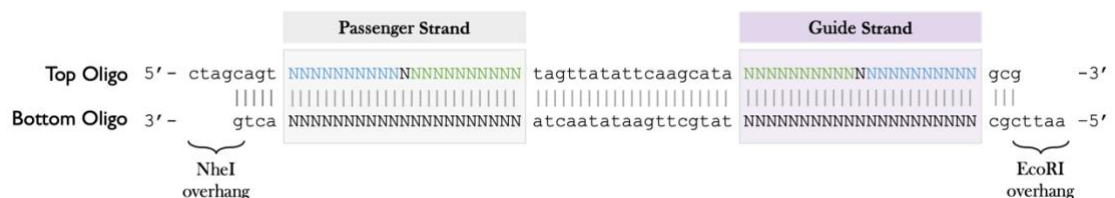


Figure 2.2: TRiP hairpin oligonucleotides design. The TRiP RNAi system utilizes short hairpin microRNA technology to efficiently silence gene expression. The design of the system revolves around the miR1 scaffold, which is composed of two 71-nucleotide-long oligonucleotide sequences, namely top and bottom. The top oligo is composed of the passenger strand and guide strand, which are separated by the sequence “tatgcttgaatataacta” to form a miRNA duplex when transcribed. The passenger strand (highlighted in gray) is the sense sequence that is common to all splice variants, while the guide strand (highlighted in purple) is the antisense sequence of the passenger strand which is indicated by the green-blue coloured nucleotides (N). The bottom oligo is designed as the reverse complement of the top oligo. Upon annealing of top and bottom oligos, the *NheI* and *EcoRI* restriction site overhangs are produced, which allows for directional cloning into knockdown vectors. After *in vivo* transcription and RNAi processing, the guide strand is incorporated into the RISC and is responsible for silencing gene expression (Perkins *et al.* 2015).

Top and bottom oligos were annealed in 50 μ L of Annealing buffer (10mM Tris-HCl pH 7.5, 0.1M NaCl, 1mM EDTA), containing a final concentration of 2 μ M of each oligo. The annealing reaction was mixed, incubated at 95°C for 5 min, and then slowly cooled down to room temperature. The resulting annealed DNA oligos, carrying *NheI* and *EcoRI* overhangs, serve as a DNA insert. To generate the pWalium-*shep* hairpin constructs, pWalium20 and pWalium22 were double digested with *NheI* and *EcoRI* (refer to Section 2.1.1 for detailed information on cloning vectors). As opposed to our standard procedure where digested vectors are treated with alkaline phosphatase to remove phosphate groups from the 5'-ends and minimize self-ligation, the digested pWalium vectors were not dephosphorylated. Instead, the insert DNA was directly cloned downstream of the UAS sites into the MCS of the pWalium-20 or -22 vectors, generating constructs designed for effective hairpin expression either in the soma (pWalium20-*shep*) or in the germline (pWalium22-*shep*). The constructs were transformed into DH5 α cells, and the resulting colonies were then screened for the presence of insert using diagnostic colony PCR products (refer to Appendix 1 for primer sequences) followed by agarose gel electrophoresis.

After colony screening, constructs with the expected insert size were sequenced before endo-free purification of plasmid DNA was carried out using the E.Z.N.A.® Endo-Free Plasmid DNA Mini kit from a 20mL overnight LB culture with 100 μ g/mL ampicillin (refer to Sections 2.1.11-12). Transgenic flies were generated by the Fly Facility at Cambridge University using the PhiC31 integration method, as described in Section 2.1.14. All newly generated pWalium constructs bear an attB attachment site allowing for their integration into attP landing sites. The pWalium20-*shep* construct was integrated into the attP40 landing site located on the left arm of the 2nd chromosome, whereas pWalium22-*shep* was integrated into the attP2 landing site located on the left arm of the 3rd chromosome. Here, constructs were generated for these specific chromosomal locations because transgenic flies with *shep* RNAi constructs on the respective chromosomes were commercially available through the Bloomington Drosophila Stock Centre (refer to Appendix 4.3 for a list of fly stocks).

2.2.1.2 Generation of pTIGER-mCh:Shep Constructs

To produce transgenic flies for visualizing Shep localization pattern and for gain-of-function analysis in the germline (i.e., overexpression experiments), three isoforms of Shep, namely -A, -C and -E, were N-terminally tagged with mCherry and cloned into the pTIGER plasmid. The isoforms are depicted in panel A of **Figure 2.3**.

Cloning of mCherry into pTIGER for N-terminal Tagging: The mCherry coding sequence, which lacks a stop codon, was obtained from the pBSII-mCherry-nostop construct available in our lab (López de Quinto, unpublished). The coding sequence was excised using *KpnI* and *SpeI* which cuts upstream of the mCherry start codon and just downstream of the last codon, respectively. After DNA gel purification, the digested fragment was ligated into *KpnI-SpeI* double digested and dephosphorylated pTIGER-attB vector downstream of the UAS sites into the MCS to generate pTIGER-attB-mCherry construct. The construct was subsequently processed according to the standard cloning procedure. To briefly describe the procedure, DH5 α -transformed colonies were screened for the presence of inserts using diagnostic endonuclease digestions and thoroughly sequenced before the standard plasmid DNA purification was carried out using EZ-10 Spin Column Plasmid DNA Miniprep kit from a 20mL overnight LB culture with 100 μ g/mL ampicillin. For the remainder of this thesis this generated construct will be referred to as pTIGER-mCherry or pTIGER-mCh for short. These constructs served as a backbone for subsequent subcloning of N-terminal mCherry-tagged proteins.

Cloning of Shep Isoforms into pTIGER-mCherry: To individually tag the three isoforms of Shep (-A, -C, and -E) with mCherry at the N-terminal, the Shep full-length coding sequences including the stop codon were PCR-amplified from the EST clones as DNA template, using VeriFi™ DNA polymerase. Refer to Sections 2.1.2.2 for details on the EST clones and Section 2.1.3.1 for information on PCR. The coding sequences of Shep -A and -C were obtained from the LD40028 clone, whereas Shep-E was acquired from the RH63980 clone. The primers referred to in Appendix 1 introduced a *BglII* restriction site at the 5' end of the amplicon and *XbaI* restriction site at the 3' end. The PCR products were purified using the GeneJET PCR Purification kit. After

endonuclease digestion, the PCR fragments were ligated into *Bg*II-*Xba*I digested and dephosphorylated pTIGER-mCh vector, downstream of mCherry coding sequence that lacks a stop codon. The resulting ligation products (i.e., the newly generated constructs) were named pTIGER-mCh:Shep -A, -C, and -E. These constructs were designed so that the coding sequence of Shep is in-frame with that of the mCherry to generate a recombinant fusion protein.

The generated constructs were then transformed into DH5 α cells and screened using *Bg*II-*Xba*I diagnostic restriction digestion. The constructs were then thoroughly sequenced before endo-free purification of the plasmid DNA was carried out using the E.Z.N.A.® Endo-free Plasmid DNA Mini kit from a 20mL overnight LB culture with 100 μ g/mL ampicillin. Generation of the transgenic flies was performed by the Fly Facility at Cambridge University using the PhiC31 integrase system for site-specific integration of constructs into the *D. melanogaster* genome, as previously outlined in Section 2.1.14. All newly generated pTIGER constructs contain an attB attachment site, which allows for their integration into attP landing sites previously introduced into the fly genome. All three constructs were integrated into the attP40 landing site, located on the left arm of the 2nd chromosome. Additionally, pTIGER-mCh:Shep-E construct was also integrated into the attP2 landing site located on the left arm of the 3rd chromosome.

2.2.1.3 Generation of Shep mutant Constructs

To generate Shep mutants lacking their RNA-binding capacity for functional analysis in the female germline, the same three isoforms of Shep previously tagged with mCherry were mutated and re-cloned into the pTIGER vector. To achieve a Shep mutant with impaired RNA-binding, key residues in the RNA binding domains were mutated, more specifically in the beta-3/RNP1 region of each RNA recognition motif (RRM; **Figure 2.3**). The mutations introduced were based on a previous study conducted by Chen *et al.* (2019), where Shep-E mutant was generated and biochemically tested (i.e., EMSA) to confirm its incapability to bind RNA targets. The selected residues were changed to alanine to reduce the likelihood of altering the protein folding.

The different Shep mutants, namely -A, -C, and -E, were generated sequentially from one another. In other words, after Shep-A was subjected to site-directed mutagenesis, it

was used as a template for PCR to first obtain the Shep-C mutant, from which the Shep-E mutant was derived. For consistency and as a sign of courtesy to the previous study, these newly generated Shep mutants with mutations in both RRM1 and RRM2 will be referred to as Shep^{RRM} throughout the thesis.

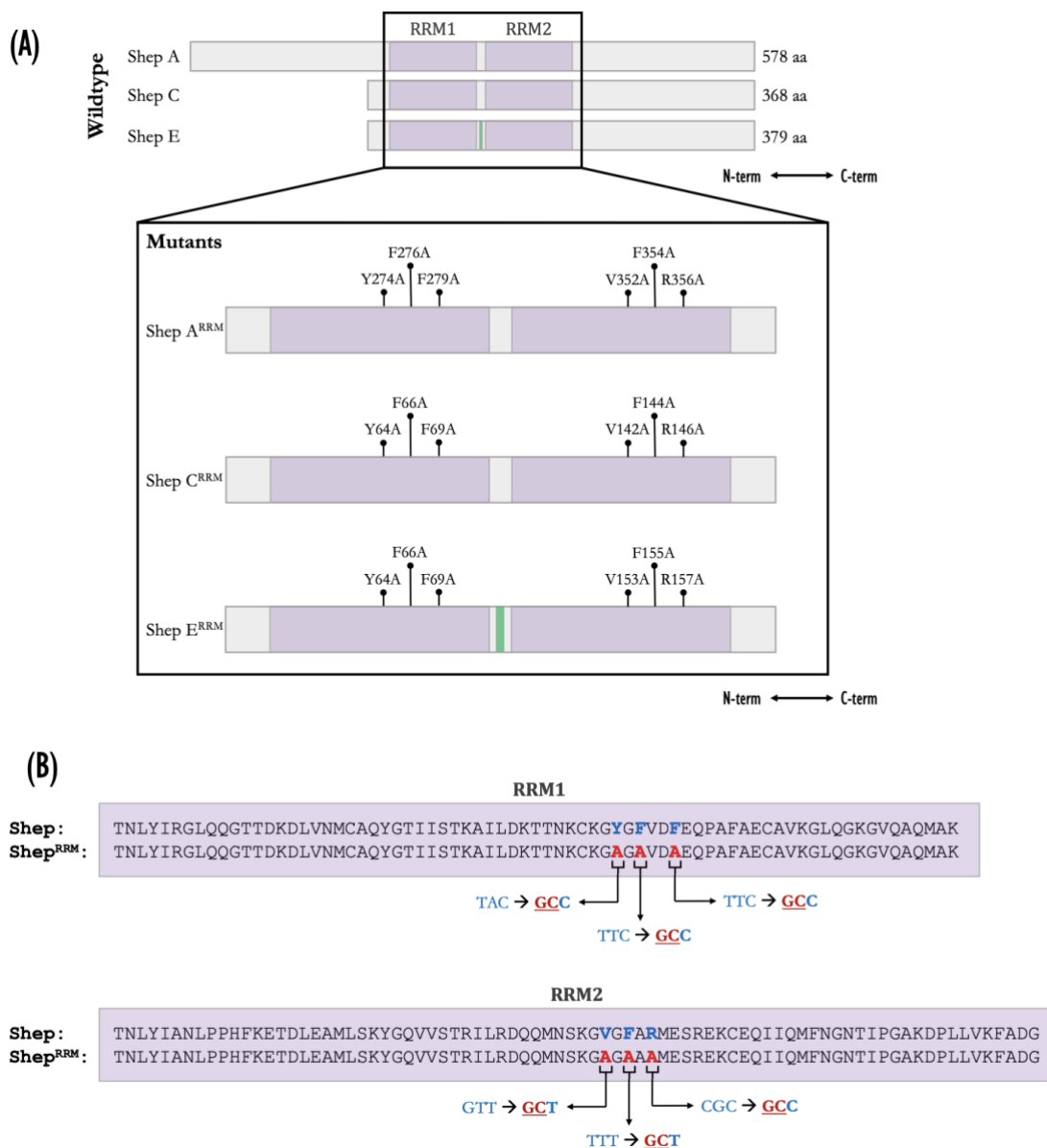


Figure 2.3: Generation of Shep RRM mutants unable to bind RNA. (A) A schematic diagram of the three isoforms cloned in this study, namely Shep-A, -C, and -E. All isoforms of Shep contain two RNA recognition motifs (RRMs; shown in purple), conferring its ability to bind RNA molecules. Unlike isoforms -A and -C, Shep-E has an additional 11 amino acid linker between the RRM domains (shown in green). The zoom-in highlights the specific residues in Shep RRM mutants that were targeted for mutagenesis. These RNA-binding mutations include changing amino acid residues such as Tyrosine (Tyr, T), Phenylalanine (Phe, F), Valine (Val, V), and Arginine (Arg, R). Diagram is not depicted to scale. **(B)** An alignment between the wildtype and mutant sequence, highlighting the mutations introduced. The mutants were generated by site-directed

mutagenesis to mutate three key amino acid residues in the RNP1 domains of each RRM1 and RRM2 to alanine (Ala, A). The targeted nucleotides and residues are highlighted in blue (wildtype) and red (mutant) in the sequences. The newly generated mutants are expected to impair the RNA binding capacity of Shep and therefore may have significant effects on RNA processing and regulation.

Mutagenesis of Shep Isoform A: The Shep-A RRM mutant was generated by two rounds of PCR-based site directed mutagenesis using the QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies, 200521), following the manufacturer's instruction. Oligonucleotide primers were designed according to the guidelines provided by the manufacturer. A set of primer pairs were designed to introduce three-point mutations in each of the RNP1 folds in both RRM1 and RRM2 domains (refer to Appendix 1 for primer sequences), as indicated in **Figure 2.3**. In the first round of mutagenesis, the PCR reaction was carried out using the pTIGER-mCh:Shep-A construct as a template and the appropriate primer pair to introduce the mutations in RRM1 (Y274A, F276A, F279A in the RNP1 domain). The PCR reaction was treated with DpnI (provided by the kit) to digest parental methylated and hemimethylated DNA template before transformation. The resulting PCR product, Shep-A^{RRM1} mutant construct (pTIGER-mCh:Shep-A^{RRM1}), was transformed into XL10-Gold ultracompetent cells provided by the kit. Transformants were plated on ampicillin plates, colonies were screened for insert, and constructs were verified by sequencing to contain the mutations in RRM1.

For the second round of mutagenesis, the same procedure was repeated for RRM2 but instead using the pTIGER-mCh:Shep-A^{RRM1} construct as template and a different primer pair to introduce the mutations in RRM2 (V352A, F254A, R356A in the RNP1 domain). Once the construct was verified to carry all RNA-binding mutations, Y274A, F276A, F279A in the RNP1 domain of RRM1 plus V352A, F254A, R356A in the RNP1 domain of RRM2, the mutated Shep-A^{RRM} CDS was double digested using *Bg*II and *Xba*I for DNA gel purification. Then, the Shep-A^{RRM} CDS fragment was ligated into a fresh *Bg*II-*Xba*I digested and dephosphorylated pTIGER-mCh vector, downstream of mCherry coding sequence that lacks a stop codon. This ensures that any mutations unaccounted for that may have been induced in the vector's backbone are avoided. An endo-free purification of the plasmid DNA was carried out using the E.Z.N.A.® Endo-free Plasmid DNA Mini kit from a 20mL overnight LB culture with 100 µg/mL

ampicillin. The newly generated mutant was named pTIGER-mCh:Shep-A^{RRM} and will be referred to as Shep-A^{RRM} throughout the thesis.

Obtaining Shep-C RNA-binding mutant: Since Shep-C is a shorter version of Shep-A, the coding sequence of Shep-C RNA-binding mutant was thereby directly amplified from pTIGER-mCh:Shep-A^{RRM} construct via PCR using VeriFiTM DNA polymerase and primers #16 and #17 referred to in Appendix 1. The primers introduce a *Bgl*II restriction site at the 5' end of the amplicon and *Xba*I restriction site at the 3' end. The resulting PCR product was purified using the GeneJET PCR Purification kit and then digested with *Bgl*II and *Xba*I restriction enzymes. Once the overhangs were exposed, the PCR fragments were ligated into a *Bgl*II-*Xba*I digested and dephosphorylated pTIGER-mCh vector, downstream of mCherry coding sequence lacking a stop codon. The resulting ligation product was then transformed into DH5 α cells. Transformants were plated on ampicillin plates, colonies were screened for the insert, and constructs were verified by sequencing to contain the mutations in the RNP1 domain of both RRM1 (Y64A, F66A, and F69A) and RRM2 (V142A, F144A, and R146A). After confirming the mutations through sequencing, an endo-free purification of the plasmid DNA was carried out using the E.Z.N.A.[®] Endo-Free Plasmid DNA Mini kit from a 20mL overnight LB culture with 100 μ g/mL ampicillin. The newly generated mutant was named pTIGER-mCh:Shep-C^{RRM} and will be referred to as Shep-C^{RRM} throughout the thesis.

Generation of Shep-E RNA-binding Mutant: Isoform E is very similar to Shep-C but with an additional 11 amino acids in the linker region between the RRM domains. Therefore overlap extension PCR was utilized to obtain the coding sequence of Shep-E RRM mutant from the pTIGER-mCh:Shep-C^{RRM} construct. This involved two rounds of PCR amplifications using the VeriFiTM DNA polymerase, and a set of short outermost and overlap primers. The outermost primers were designed to introduce a *Bgl*II restriction site at the 5' end of the amplicon and *Xba*I restriction site at the 3' end for downstream cloning application. Whereas, the overlap primers were designed following the guidelines used by Hilgarth and Lanigan (2020) to incorporate overlapping sequences into the PCR products in initial round of amplification. This is

crucial for the overlap extension amplification in the second round, which is necessary to incorporate the 11-amino acid linker sequence.

In the initial round of PCR amplification, two amplicons were generated in separate conventional PCR reactions using primers #16 and #78 for the first reaction and #79 and #17 for the second reaction, as depicted in **Figure 2.4**. The resulting amplicons, namely Amplicon #1 (300 bp) and Amplicon #2 (860 bp), share a common overlapping region which consists of the 33bp coding sequence encoding the 11-amino acid linker. The amplicons were then purified using the GeneJET PCR Purification kit to remove the template DNA (i.e., pTIGER-mCh:Shep-C^{RRM} construct) that may interfere with downstream amplification. For the second round of PCR amplification, an overlap extension PCR reaction was assembled using the amplicons as DNA templates at an equimolar ratio (86 ng of Amplicon #1 and 246 ng of Amplicon #2) along with the shorter outermost primers used in initial round of PCR (i.e., #16 and #17). The incorporation of overlapping sequences creates complementary ends on the two amplicons, which allows them to anneal and serve as templates for extension in the second round of amplification, resulting in a single continuous DNA product. The resulting amplicon, namely Amplicon #3, is the full-length coding sequence of Shep-E^{RRM} mutant. Amplicon #3 was purified using the GeneJET PCR Purification kit and then cloned into pTIGER-mCh vector as previously described for Shep-C^{RRM} construct. Once mutations in the RNP1 domain of both RRM1 (Y64A, F66A, and F69A) and RRM2 (V153A, F155A, R157A) were verified via sequencing, an endo-free purification of the plasmid DNA was carried out using the E.Z.N.A.® Endo-Free Plasmid DNA Mini kit from a 20mL overnight LB culture with 100 µg/mL ampicillin. The newly generated mutant was named pTIGER-mCh:Shep-E^{RRM} and will be referred to as Shep-E^{RRM} throughout the thesis.

Germline Transformation: Transgenic flies were generated by the Fly Facility at Cambridge University using the PhiC31 integrase system for site-specific integration of constructs into the *Drosophila* genome, this is outlined in Section 2.1.14. All three constructs were integrated into the attP40 landing site, located on the left arm of the 2nd chromosome.

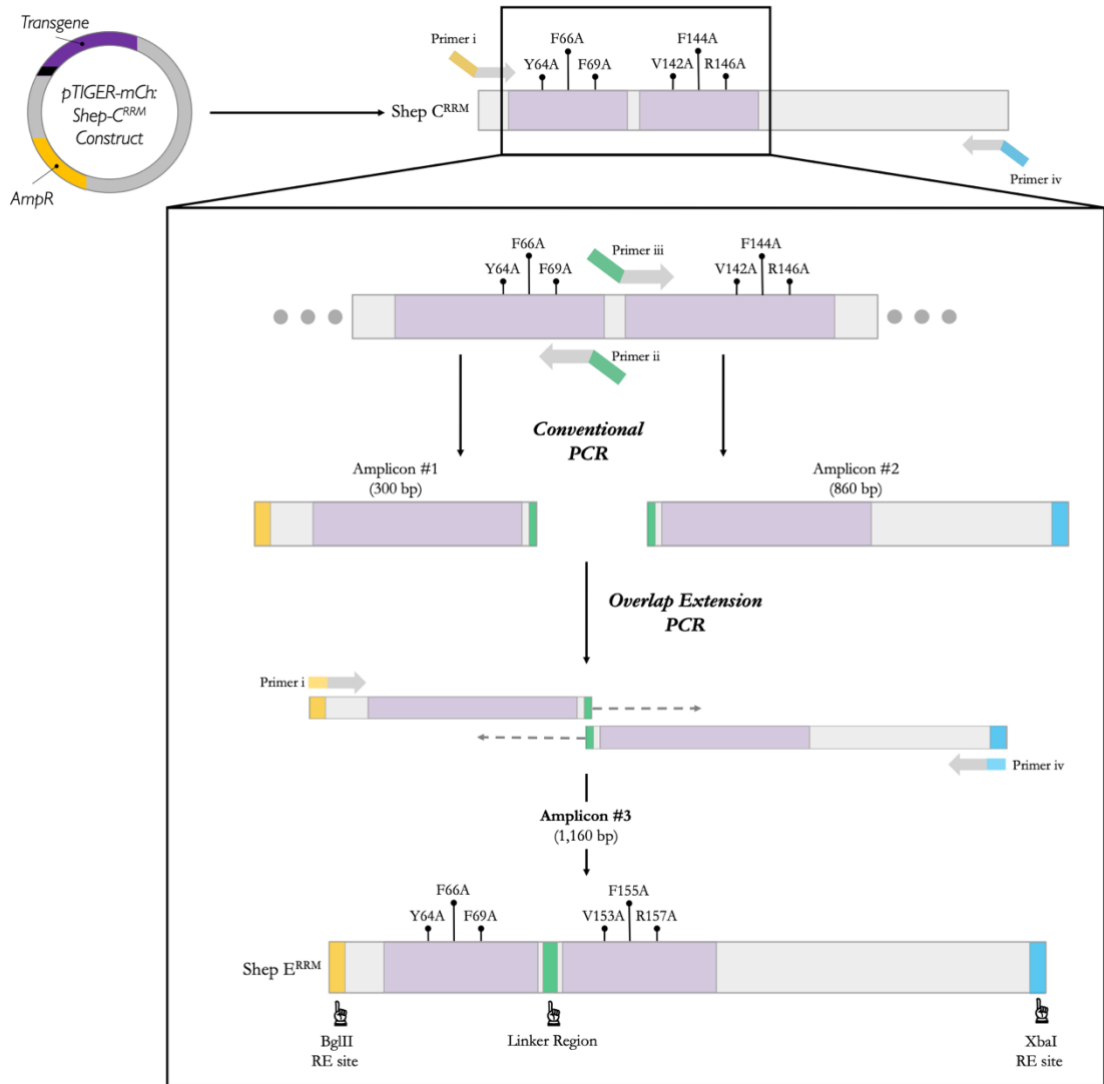


Figure 2.4: Workflow for obtaining Shep-E RRM mutant via Overlap Extension PCR. A brief graphical overview of the workflow utilised to obtain a Shep-E RRM mutant from the Shep-C^{RRM} construct. Both isoforms of Shep (C and E) are composed of two RNA recognition motifs (RRMs; shown in purple) and a seven amino acid stretch at the C-terminus (shown in orange). However, Shep-E has an additional 11 amino acid linker between the RRM domains (shown in green) which is absent in isoform-C. To generate a construct of Shep-E with the mutations in Shep-C^{RRM} plus the eleven-amino-acid linker region, the workflow involved two stages of polymerase chain reaction: conventional PCR and overlap extension PCR. The first stage uses conventional PCR, short outermost primers (i and iv), and long overlap primers (ii and iii) to generate two separate amplicons from Shep-C which contains the mutations in RRM domains needed to impair RNA-binding. Additionally, it also incorporates a 33bp overlapping sequence encoding the linker region through the long overlap primers (ii and iii; depicted as green parts of the primer) needed for obtaining Shep-E. The outermost primers (i and iv) also incorporate *Bgl*II (dark yellow) and *Xba*I (blue) restriction sites, respectively, needed for cloning applications. Amplicon #1 is generated using primers i and ii (300 bp) while Amplicon #2 is generated using primers iii and iv (860 bp). In the second stage of the workflow, an overlap extension PCR is utilised to obtain a full-length coding sequence of Shep-E with the RRM mutations (Amplicon #3, 1,160 bp). The reaction uses Amplicon #1 and Amplicon #2 as DNA templates at an equimolar ratio along with the outermost primers (i and vi). Diagram is not depicted to scale.

2.2.2 pBluescript-II Constructs for *in vitro* Synthesis of RNA Probes

To investigate RNA-protein interactions, several pBSII clones were generated to synthesize RNA probes *in vitro* for RNA-affinity pulldown assay. This section provides a detailed description of the newly generated pBSII-KS⁺ and other pre-existing constructs used in this project for *in vitro* synthesis of biotin- and digoxigenin-labelled RNA probes. Refer to Appendix 4.1 for a list of the constructs that were generated and used for this application.

2.2.2.1 Generation of pBSII-KS *shotgun* Constructs

The 5'UTR, CDS, and 3'UTR of *shotgun* (*shg*; FBgn0003391, *Drosophila* DE-Cadherin) were individually cloned into the pBSII-KS⁺ vector (Stratagene) to generate biotinylated RNA probes for their use in RNA-affinity pulldown assay. The VeriFi™ DNA polymerase was used to PCR amplify the untranslated regions and full-length coding sequence of *shg* from a *D. melanogaster* ovarian cDNA library as template (see Sections 2.1.2.3 and 2.1.2.3). The primer pairs used to amplify the untranslated regions introduced an *Xba*I restriction site at the 5' end of the amplicon and an *Xho*I restriction site at the 3' end (refer to Appendix 1). Whereas *Xba*I and *Eco*RI restriction sites were incorporated into the 5' and 3' ends of the CDS sequence, respectively. After PCR amplification, the resulting amplicons were digested with the appropriate endonuclease enzymes and then ligated downstream of the T7 promoter into the MCS of pre-digested and dephosphorylated pBSII-KS⁺ vector. The constructs were transformed into DH5 α cells and then screened using diagnostic restriction digestion, specifically *Xba*I-*Xho*I double digestion for *shg* UTRs and *Xba*I-*Eco*RI for the CDS construct. After colony screening, the constructs with an expected insert size were sequenced. Subsequently, standard plasmid DNA purification was carried out using EZ-10 Spin Column Plasmid DNA Miniprep kit from a 20mL overnight LB culture with 100 μ g/mL ampicillin. The newly generated constructs were named pBSII-KS⁺ *shg*-5'UTR, -CDS, and -3'UTR.

To use as DNA template for *in vitro* transcription, these constructs were linearized with *Xho*I for the UTRs and *Eco*RI for the CDS. The linearized constructs were then

purified using the GeneJET PCR Purification kit before proceeding to the *in vitro* synthesis of RNA probes described in Section 2.3.1.

2.2.2.2 Generation of pBSII-KS *gurken* Constructs

The 5'UTR, CDS, and 3'UTR of *gurken* (*grk*; FBgn0001137) were individually cloned into the pBSII-KS⁺ vector (Stratagene) to generate sense biotinylated RNA probes for their use in RNA-affinity pulldown assay. The regions of interest were PCR amplified from a *D. melanogaster* ovarian cDNA library as template, using the VeriFi™ DNA polymerase. The primer pairs used to amplify the untranslated regions and full-length coding sequence of *grk* introduced *Xba*I and *Eco*RI restriction sites at the 5' and 3' ends of each amplicon, respectively. Cloning of PCR fragments into the pBSII-KS⁺ vector, screening for the presence of the insert, and sequencing of the constructs were conducted as outlined for the pBSII-KS *shg* constructs (see Section 2.2.2.1). The newly generated constructs were named pBSII-KS⁺ *grk* -5'UTR, -CDS, and -3'UTR. For *in vitro* synthesis of biotinylated sense RNA probes, all three constructs were linearized with *Eco*RI.

2.2.2.3 Other pBSII Clones used for *in vitro* Synthesis of RNA probes:

Biotinylated sense RNA probes of different *oskar* regions (M1M2, Short-Osk, and 3'UTR) and *tsunagi* (aka *Y14*, CDS) were generated as described in Besse et al. (2009). Also, other biotinylated probes were synthesized from constructs already available in the lab, which includes the UTRs and CDS of *Delta* and *Notch*, or the 3'UTR only of *bicoid*, *nanos*, and *K10*. The coding sequence of *notch* was split in two halves, roughly 4Kb each, to generate two pBSII constructs for *Notch* CDS (1 and 2), to avoid complications with probes synthesis. Furthermore, the construct used to synthesize digoxigenin (DIG) labelled *oskar* antisense probe for fluorescent *in situ* hybridization was a gift from Anne Ephrussi (EMBL, Germany) and generated as described in Ephrussi and Lehmann (1992) and Vanzo and Ephrussi (2002).

2.2.3 Constructs for Protein Expression in Bacteria

To induce expression in bacteria, constructs were generated by inserting the coding sequence of the proteins of interest into bacterial expression vectors compatible with

the T7 RNA polymerase induction system (Chamberlin *et al.* 1970; Studier and Moffatt 1986; Iost *et al.* 1992). These vectors allow for efficient transcription of the inserted coding sequence under the control of the T7 promoter, enabling high-level protein expression. In this study, we generated tagged fusion proteins using two types of epitopes. The first is the widely used 6 consecutive histidine residue tag, while the second is a novel 12-amino acid peptide tag from ChromoTek called the Spot tag.

2.2.3.1 Generation of *pSpot-PTB* and *pSpot-Su(var)2-10* Constructs

For PTB and Su(var)2-10, the presence of unique restriction sites on either side of the CDS enabled direct subcloning of the CDS into a bacterial expression vector using the traditional ‘Cut-&Paste’ cloning method. The full-length coding sequence of PTB was obtained from the pT7CFE-CHisU-PTB construct through double-digestion using *Bam*HI and *Xho*I enzymes, which cut upstream of the start codon and just downstream of the stop codon, respectively. Similarly, the full-length coding sequence of Su(var)2-10 was obtained from the pUAS_t-attB-Su(var)2-10 construct, a generous gift of Robert Mitchell, through double digestion using *Eco*RI and *Xho*I. After DNA gel purification, digested fragments were ligated into an appropriately double-digested and dephosphorylated pSpot1 expression vector, downstream of the Spot-tag coding sequence (ChromoTek, ev-1). The resulting constructs were named pSpot1-PTB and -Su(var)2-10. The newly generated constructs were transformed into DH5 α cells and then screened using diagnostic colony PCR, as described in Sections 2.1.11-13 and 2.1.3.2. The constructs were thoroughly sequenced before a plasmid DNA purification was carried out from 20mL overnight culture with 50 μ g/mL kanamycin, using EZ-10 Spin Column Plasmid DNA Miniprep kit, see Sections 2.1.11-13.

2.2.3.2 Generation of *pSpot-Shep* Constructs

To generate N-terminal Spot-tagged Shep fusion proteins for protein expression and subsequent purification, the full-length CDS of the different Shep isoforms (-A, -C, and -E) including the stop codon were PCR amplified from the DGRC Gold EST clones, using VeriFi™ DNA polymerase as previously described in Section 2.1.3.1. The primers referred to in Appendix 1 introduced a *Eco*RI restriction site at the 5' end of the amplicon and *Hind*III restriction site at the 3' end. Shep amplicons were double

digested using *EcoRI* and *HindIII* to expose restriction overhangs for cloning. After restriction digestion, the amplicons were purified using the GeneJET PCR Purification kit and ligated into digested and dephosphorylated pSpot1 expression vector, downstream of the Spot-tag coding sequence (ChromoTek, ev-1). The resulting constructs were named pSpot1-Shep -A, -C, and -E. The newly generated constructs were transformed into DH5 α cells and then screened using diagnostic colony PCR. The constructs were thoroughly sequenced before a plasmid DNA purification was carried out from 20mL overnight culture with 50 μ g/mL kanamycin, using EZ-10 Spin Column Plasmid DNA Miniprep kit.

2.2.3.3 Generation of pNEBExpress-Shep Constructs

Generation of pNEBExpress-His Expression Vector: The *in vitro* protein synthesis kit from NEB (E5360) provides a construct containing the CDS of the metabolic enzyme dihydrofolate reductase (DHFR), as a positive control for protein expression experiments. The CDS of DHFR was removed using *BamHI*, and then the expression vector backbone was purified from the agarose gel using the GeneJET Gel Purification kit. The expression vector backbone was allowed to self-ligate, transformed into DH5 α competent cells, and grown overnight at 37°C. Colonies were then screened for the absence of the DHFR CDS by diagnostic digestion. After the right clone was identified, a large preparation of plasmid DNA purification was carried out from 20mL overnight culture with 100 μ g/mL ampicillin, using EZ-10 Spin Column Plasmid DNA Miniprep kit. The resulting vector was sequenced before being used as a cloning vector for *in vitro* protein expression applications. This cloning vector will subsequently be referred to as pNEBExpress expression vector.

Generation of pNEBExpress-His-ShepE Construct: To generate N-terminal His-tagged Shep for protein expression and subsequent purification, the full-length coding sequence of Shep-E including the stop codon was PCR amplified from the RH63980 DGRC Gold EST clone, using VeriFi™ DNA polymerase. The primers referred to in Appendix 1 introduced a *EcoRI* restriction site at the 5' end of the amplicon and *XhoI* restriction site at the 3' end. Shep-E amplicons were double digested using *EcoRI* and *XhoI* to expose restriction overhangs for cloning. After restriction digestion, the

amplicons were purified using the GeneJET PCR Purification kit and ligated into digested and dephosphorylated pNEBExpress cloning vector, downstream of the His-tag coding sequence to generate a His-tagged Shep fusion protein. The resulting construct was named pNEBExpress-His:ShepE. The newly generated construct was transformed into DH5 α cells and colonies were subsequently screened using diagnostic colony PCR. The final construct was thoroughly sequenced before plasmid DNA purification was carried out from 20mL overnight culture with 100 μ g/mL ampicillin, using EZ-10 Spin Column Plasmid DNA Miniprep kit.

Generation of pNEBExpress-His-Spot-ShepE Construct: To generate Shep fusion protein for *in vitro* expression with both His- and Spot-tag at the N-terminal, the top and bottom N-Spot oligos were annealed in 50 μ L of Annealing buffer (10mM Tris-HCl pH 7.5, 0.1M NaCl, 1mM EDTA) containing a final concentration of 2 μ M of each oligo. Sequences of the top and bottom oligos can be found to in Appendix 1. The annealing reaction was mixed, incubated at 95°C for 5 min, and then slowly cooled down to room temperature. The resulting annealed oligo carries *Bam*HI overhangs at both ends, allowing for its insertion into pNEBExpress cloning vector downstream the His-tag. Contrary to the standard procedure, the pNEBExpress cloning vector was digested using *Bam*HI but was not dephosphorylated to ensure a successful ligation. After ligating the dsDNA oligo into the cloning vector, the construct was transformed into DH5 α cells and colonies were subsequently screened using diagnostic colony PCR to ensure the oligo's presence in the correct orientation. After the construct was sequenced, the plasmid DNA was purified from 20mL overnight culture with 100 μ g/mL ampicillin using EZ-10 Spin Column Plasmid DNA Miniprep kit.

Next, the pNEBExpress-His:Spot vector was double digested using *Eco*RI and *Xho*I, dephosphorylated using alkaline phosphatase, and purified using the GeneJET PCR Purification kit. After purifying the linearized vector, the full-length coding sequence of Shep-E, including the stop codon—which was previously generated by PCR as described for the pNEBExpress-His-ShepE construct (see above)—was ligated downstream of the Spot-tag. The pNEBExpress-His:Spot:ShepE construct was designed so that the coding sequence of Shep is in-frame with that of both the histidine and Spot tags to generate a

recombinant dual-tagged fusion protein. After thoroughly sequencing the final clone, plasmid DNA was purified from a 20mL overnight culture containing 100 µg/mL ampicillin using the EZ-10 Spin Column Plasmid DNA Miniprep kit.

Generation of pNEBExpress-His-ShepE-Spot Construct: To generate a Shep fusion protein with both N-terminal His-tagged and C-terminal Spot-tagged for protein expression and purification, the full-length CDS of Shep-E, this time lacking the stop codon, was PCR amplified from the RH63980 DGRC Gold EST clone, using VeriFi™ DNA polymerase. The primers introduce *EcoRI* and *XhoI* restriction sites at the 5' end and 3' end of the amplicon, respectively. The Shep-E amplicon was cloned into *EcoRI-XhoI* double-digested and dephosphorylated pNEBExpress cloning vector downstream the His-tag. The resulting pNEBExpress-His:ShepE^{noStop} construct was then transformed into DH5α cells, plated on ampicillin LB agar plates, screened via restriction digestion, and thoroughly sequenced. Afterwards, the pNEBExpress-His:ShepE^{noStop} construct was digested using *XhoI* to allow for the ligation of the C-Spot dsDNA oligo, which introduced a short linker region followed by the Spot tag with an added stop codon. The top and bottom C-Spot oligos were annealed as previously described for the N-terminal Spot tagging (see above). This construct was then transformed into DH5α cells, grown on ampicillin LB agar plates, screened via colony PCR, and thoroughly sequenced. Next, the final clone was purified from a 20mL overnight culture containing 100 µg/mL ampicillin using the EZ-10 Spin Column Plasmid DNA Miniprep kit.

2.3 General Biochemical Techniques

2.3.1 *In vitro* Transcription

To identify and characterize RNA-binding protein interactions, biotinylated sense RNA transcripts were generated from recombinant pBSII constructs using *in vitro* transcription. To achieve this, recombinant pBSII constructs were first linearized and then used as templates for the T7 *in vitro* transcription reaction (**Figure 2.5**). This method allows for efficient labelling and isolation of RNA transcripts of interest for subsequent RNA-affinity pulldown assays to study protein-RNA interactions.

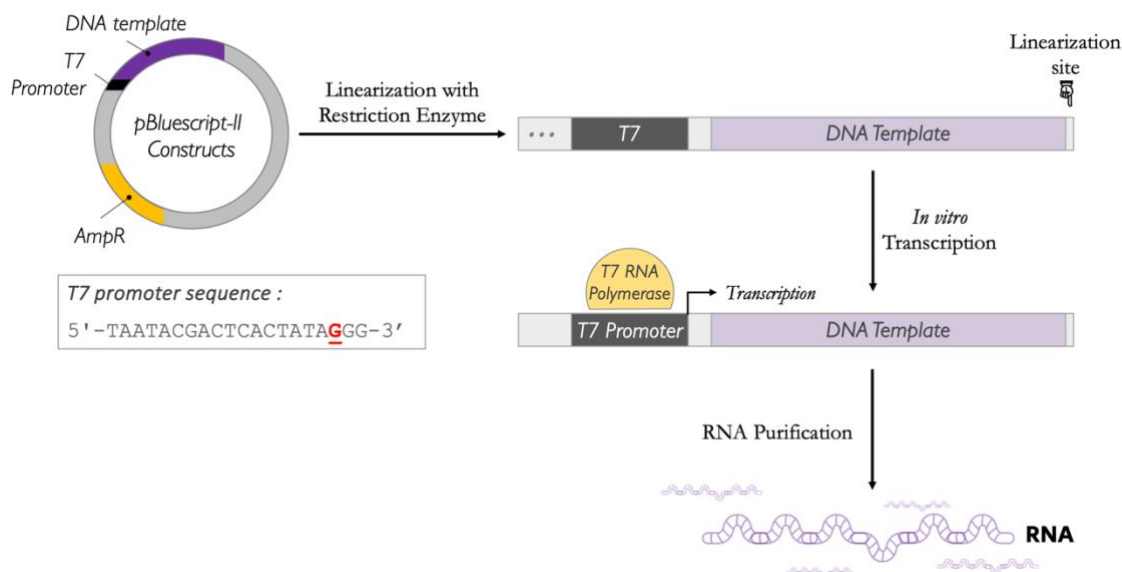


Figure 2.5: A brief overview of the T7 *in vitro* transcription workflow. *In vitro* transcription is a widely used method in the field of molecular biology to synthesize large quantities of high-quality RNA probes suitable for various downstream applications such as RNA-protein interactions. The workflow involves generating a pBluescript-II construct containing the sequence of the RNA to be synthesized, preparing the template by linearizing the construct downstream the DNA template, and setting up the transcription reaction with T7 RNA polymerase. Once transcription is complete and the resulting newly transcribed RNA is purified, quality control checks of transcript RNA are performed to assess size, integrity, and purity by gel electrophoresis and spectrophotometry. The nucleotide base ‘Guanine’, which is underlined and color-coded letter for reference, in the T7 sequence represents the first ribonucleotide that gets transcribed in the newly *in vitro* synthesised RNA transcript. RNA-protein interactions are an essential aspect of gene regulation, and *in vitro* synthesized RNA probes enable the study of these interactions with high specificity and sensitivity.

Linearization of Plasmid DNA Template: In a 150µL restriction digestion reaction, 7µg of the pBSII construct was linearized with an appropriate FastDigest restriction enzyme that cuts downstream of the transgene to be transcribed. Restriction enzymes were carefully chosen to produce either blunt ends or 5'-overhangs, as 3'-overhangs have been reported to generate spurious transcripts (Schenborn and Mierendorf 1985). Appendix 4.1 provides details on the specific restriction enzymes used to linearize the different pBSII constructs. The resulting linearized plasmid was subsequently purified using the GeneJET PCR Purification kit (Thermo Scientific) and eluted in 25µL nuclease-free Elution buffer or 1:1 ratio of RNase-free water to RNase-free TE buffer.

***In vitro* Synthesis of RNA Probes:** To generate sense RNA transcripts, the linearized pBSII construct was used as template to perform *in vitro* transcription using the TranscriptAid T7 High Yield Transcription kit (Thermo Scientific, K0441), according

to the manufacturer's instructions. In an RNase-free microcentrifuge tube on ice, 20 μ L of a T7 *in vitro* transcription reaction was prepared, containing 1.5 μ g of linearized plasmid template and Biotin-16-UTP (Roche, 11388908910). To preserve the RNA folding and minimize the disruption to its three-dimensional structure, a final concentration of 10mM UTP was used with a 1:10 molar ratio of modified to unmodified UTP. The reaction mixture was thoroughly mixed, briefly spun down, then the tube was sealed with parafilm and incubated overnight at 37°C.

Removal of Template DNA: After completion of the *in vitro* transcription reaction, the template DNA was removed by adding 1 μ L of RNase-free DNase I (1U/ μ L; provided by the kit), thoroughly mixed, and incubated at 37°C for 30 minutes.

Precipitation & Purification of RNA Transcripts: Lithium chloride (LiCl)-based precipitation method was employed to purify RNA transcripts and efficiently remove proteins, excess free-ribonucleotides, and digested DNA fragments (Gautam 2022). To precipitate the RNA, samples were incubated overnight at -20°C in 55 μ L of 4M LiCl (Invitrogen, AM9480). Following the overnight RNA precipitation in LiCl, the RNA pellet was collected by centrifugation at 13,000 rpm for 15 minutes at 4°C. The RNA pellet was washed with 500 μ L of RNase-free 75% ethanol without disrupting the pellet, and then subjected to a second round of centrifugation. The ethanol wash was discarded, and the pellet was allowed to air-dry at room temperature for 1-3 hour(s) before the RNA was resuspended in 50 μ L of RNase-free TE. Next, the newly synthesized RNA transcripts underwent quality control checks to assess their size, integrity, and purity by agarose gel electrophoresis and spectrophotometry. Once the quality of RNA transcripts was ensured, they were stored at -20°C or -80°C for long-term storage.

2.3.2 Preparation of *Drosophila* Ovarian Lysates

Ovarian cell lysates were prepared using various methods depending on the downstream application. Briefly, crude and cytoplasmic S10 extracts were prepared by homogenization of ovarian tissue in either SDS-PAGE Loading buffer or Hypotonic buffer, respectively. Following homogenization, the cellular lysate was then centrifuged to remove cellular debris, and the supernatant fractions containing the proteins were

collected. The resulting extracts were used in various downstream biochemical assays, including Western blot, RNA-protein interaction, and protein-protein interaction. The crude extract was typically run directly onto SDS-PAGE gels to visualize proteins of interest, while cytoplasmic S10 extracts were used as a source of soluble proteins in various biochemical assays to investigate RNA-protein or protein-protein interactions.

Crude Protein Extract: In a 1.5 microcentrifuge tube, 10 pairs of ovaries were rinsed twice with ice-cold PBS before being re-suspended in 100 μ L of 2 \times SDS-PAGE Loading buffer to achieve a 1:10 ratio of sample to Loading buffer (National Diagnostics, EC886). The ovaries were homogenized using 30 strokes of an EppendorfTM Micropestle (Fisher Scientific, 10683001). The resulting ovarian lysate was heated at 90 $^{\circ}$ C for 10 minutes before cellular debris were cleared by centrifugation at 16,000 \times g for 2 minutes. Before loading the extracts into the wells, an equal volume of analytical water was added to bring the final concentration of the SDS-PAGE Loading buffer to 1 \times .

Cytoplasmic S10 Protein Extract: In a 1.5 microcentrifuge tube, 20-25 pairs of ovaries were rinsed four times with ice-cold buffers: twice with PBS, once with Hypotonic buffer (10mM HEPES-KOH^{pH7.4}, 10mM potassium acetate, 1.5mM magnesium acetate, 2.5mM DTT; see **Table 2.7**), and once in Hypotonic buffer containing cOmpleteTM Mini EDTA-free protease inhibitor cocktail (PI; Roche, 04693159001) at a working concentration of 1 \times . The ovaries were then resuspended in Hypotonic buffer containing protease inhibitors, using a volume equivalent to that of the ovaries (e.g., 10 μ L of buffer for every 10 pairs of ovaries). The ovaries were homogenised on ice using an EppendorfTM Micropestle with 30 strokes of the pestle. After homogenization, the lysate was centrifuged at 10,000 rpm for 10 minutes at 4 $^{\circ}$ C to remove cellular debris. The supernatant was collected, and glycerol was added to achieve a final concentration of 5% for protein stabilization. The resulting S10 extract was kept on ice and immediately used in downstream biochemical assays or stored at -20 $^{\circ}$ C until required.

Table 2.7: Recipe to prepare 15mL of Hypotonic Lysis buffer from stock solutions. This buffer is also commonly known as the Osmotic Lysis buffer in the literature. DTT: 1,4-Dithiothreitol.

Buffer Components	Volumes per 15mL	Final [Conc.]
HEPES-KOH^{pH7.4} Solution (500 mM; Millipore, 5310)	300 μ L	10mM
Potassium Acetate Solution (5 M; Sigma-Aldrich, 95843)	30 μ L	10mM
Magnesium Acetate Solution (1 M; Sigma-Aldrich, 63052)	22.5 μ L	1.5mM
DTT Solution (500 mM; Roche, 11583786001)	75 μ L	2.5mM
Analytical water (Fisher Scientific, 10449380)	14.5 mL	N/A

2.3.3 Immunoblotting – Fluorescent Western Blot Detection Method

Preparation of Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) Gel: The polyacrylamide concentration of the resolving gel for SDS-PAGE was determined based on the molecular weight range of the proteins of interest. In this study, proteins were separated under denaturing and reducing conditions using a handcast SDS-PAGE gel composed of a 5% stacking gel and 8-15% resolving gel (Table 2.8). To cast 1mm thick polyacrylamide gels, a glass cassette was assembled using the Mini-PROTEAN® Tetra Cell System (Bio-Rad, 1658003) according to the manufacturer's instructions. A 10-well comb was used to polymerize polyacrylamide gels to analyse up to 9 samples alongside a marker on one gel.

Table 2.8: Recipe for preparing resolving and stacking gels. * During gel casting the components were added sequentially in the order shown below.

SDS-PAGE Gel Layer	Gel Concentration
Stacking Gel	5%
Resolving Gel	8%, 9%, 10%, or 15%
Reagent(s)*	Reference
Analytical water	Fisher Scientific (10449380)
30% Acrylamide/Bis Solution (37.5:1)	National Diagnostic (EC890)
4× Resolving buffer (1.5M Tris-HCl, 0.4% SDS, pH 8.8)	National Diagnostic (EC892)
4× Stacking buffer (1.5M Tris-HCl, 0.4% SDS, pH 6.8)	National Diagnostic (EC893)

10% Ammonium Persulfate (APS)	Bio-Rad (1610700) & Sigma-Aldrich (A9164)
TEMED	Bio-Rad (1610800)

Gel Electrophoresis: The samples were prepared to achieve a final concentration of 1× SDS-PAGE Loading buffer before being loaded into the wells. Each well was loaded with 5-10µg of protein extract, which is equivalent to 1-2 pairs of ovaries, respectively. With each run, 10µL of Precision Plus Protein™ Dual Color Standards protein ladder (Bio-Rad, 1610374) was run alongside samples. The samples and ladder were heated at 95°C for 5 minutes before loading into the wells. The gel was electrophoresed at a constant current in Tris-Glycine SDS-PAGE Running buffer (National Diagnostics, EC870). The stacking gel was run at 15mA for 30 minutes, while the resolving gel was run at 25mA for an additional hour or until the 25 kDa band marked with pink has migrated to the very bottom of the gel. The given electrophoresis running conditions were designed for running a single mini gel. Therefore, amperages were adjusted accordingly if multiple gels were run simultaneously in the same electrophoresis tank.

Protein Transfer: The proteins separated on polyacrylamide gels were transferred to a Mini 0.2µm nitrocellulose membrane (Bio-Rad, 1704158) using the Trans-Blot® Turbo™ Transfer System (Bio-Rad, 1704150) following the manufacturer's instructions. The polyacrylamide gel was sandwiched according to the manufacturer's instructions, as shown in **Figure 2.6**, and electrophorized using the '7-minutes mixed MW' transfer protocol. After semi-dry transfer of the proteins, the nitrocellulose membrane was dried between Whatman™ 3MM Chr Chromatography Paper (GE Healthcare, 3030-917) for 5-10 minutes at room temperature before proceeding to the next step of the immunodetection protocol.

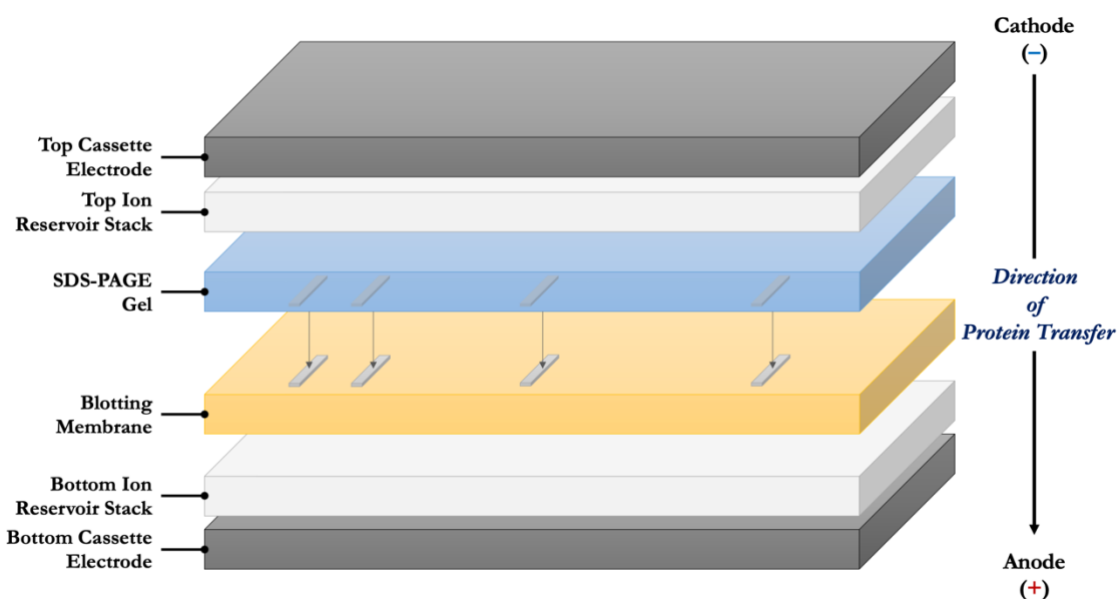


Figure 2.6: Assembly of semi-dry transfer sandwich using the trans-blot turbo system. For a successful transfer of proteins from the gel to the membrane, components are carefully assembled and orientated in respect to the charged plate electrodes. The nitrocellulose membrane is situated nearest to the positive electrode (+, anode) at the bottom of the cassette. Meanwhile, the SDS-PAGE gel is placed over the membrane, closer to the negatively charged electrode (-, cathode). This arrangement ensures that SDS-bound proteins, which acquired an overall negative charge during sample preparation and electrophoresis, migrate towards the positively charged electrode and onto the membrane.

Fluorescent Immunodetection of proteins: The nitrocellulose membrane was rehydrated in PBS before being blocked with 8mL of Intercept® PBS Blocking buffer (LI-COR Biosciences, 927-70001) at room temperature for 1 hour with continuous agitation on a rocking platform. The membrane was then incubated overnight with primary antibodies diluted in Intercept® T20 PBS Antibody Diluent (LI-COR Biosciences, 927-75001) at 4°C with shaking (refer to **Table 2.9** for antibody dilutions). After incubation with primary antibodies, the membrane was rinsed three times with PBST (0.1% Tween-20 in PBS) and then washed three more times in PBST for 10 minutes each. The washed membranes were then incubated with fluorescently labelled secondary antibodies diluted in PBST solution containing 5% w/v of skimmed milk (Marvel, 5000354909832) for 1 hour at room temperature with shaking (refer to **Table 2.9** for antibody dilutions). The membrane was then rinsed and washed as described for the primary antibody before being washed with PBS for 5 minutes. After washing, proteins were detected by fluorescence using the Odyssey® CLx Infrared Imaging

System (LI-COR Biosciences, Model# 9140). Image analysis was performed using Li-Cor software ‘Image Studio’ (version 5.2) and ImageJ (version 1.8.0_181 64 bits).

A modified protocol of the immunodetection procedure described above was used when detecting Oskar protein. The modified protocol differs from the standard protocol in that PBS-based buffers were substituted by TBS-based buffers. Therefore, following protein transfer, the dried membrane was rehydrated in TBS (150mM NaCl, 50mM Tris-HCl buffer, pH 7.6; Calbiochem, 524750), and was blocked in TBSM (TBS containing 5% w/v skimmed milk). In addition, primary and secondary antibodies were diluted in TBSTM (0.1% Tween® 20 in TBS with 5% skimmed milk). During the washing steps, membranes were washed with TBST (0.1% Tween® 20 in TBS).

Table 2.9: List of primary & secondary antibodies used for fluorescent immunoblotting.

Target Protein	Antibody	Host Organism	Dilution	Reference
Primary Antibody:				
PTB	Polyclonal	Rat	1:3,000	Besse <i>et al.</i> (2009)
eGFP	Monoclonal	Rat	1:1,500	Chromotek (3H9)
mCherry	Monoclonal	Mouse	1:1,000	Chromotek (6G6)
Shep	Polyclonal	Rabbit	1:1,000	Matzat <i>et al.</i> (2012)
Oskar	Polyclonal	Rabbit	1:2,000	Vanzo and Ephrussi (2002)
Kinesin Heavy Chain	Monoclonal	Rabbit	1:10,000	Cytoskeleton (AKIN01)
Dynein	Monoclonal	Mouse	1:2,000	Millipore (MAB1618)
Fluorophore-Conjugated Secondary Antibody:				
IRDye® 680RD Rabbit IgG	Polyclonal	Goat	1:10,000 or 1:15,000	LI-COR Biosciences (926-68071)
IRDye® 800CW Rabbit IgG	Polyclonal	Goat		LI-COR Biosciences (926-32211)
IRDye® 800CW Rat IgG	Polyclonal	Goat		LI-COR Biosciences (926-32219)
IRDye® 800CW Mouse IgG	Polyclonal	Goat		LI-COR Biosciences (926-32210)
IRDye® 680RD Mouse IgG	Polyclonal	Goat		LI-COR Biosciences (926-68070)

Membrane Stripping for Reprobing Using the High pH Method: To remove signals from nitrocellulose membrane to probe with different antibodies, the membrane was washed in PBS for 5 minutes on a rocking platform and then submerged in stripping solution (0.2M NaOH; Fisher Chemicals, S492053) for another 10 minutes.

Subsequently, the membrane was rinsed and washed twice with PBST (0.1% Tween-20 in PBS) for 10 minutes. The stripped membrane can be reprobbed for the next round of immunodetection, proceeding from the blocking step of the protocol.

Total Protein Staining: For total detection of proteins in the 800nm channel, the membrane was stained with 1:3 dilution of Revert™ 700 Total Protein Stain (LI-COR Biosciences, 926-11011; supplemented with methanol) in analytical water for 5-10 minutes at room temperature with gentle agitation on a rocking platform. The membrane was briefly rinsed twice with distilled water and then submerged in distilled water before proceeding to acquire an image of the membrane in the 700nm channel using the Odyssey® CLx Imager. Refer to the manufacturer's manual for more information on 'Total Protein Quantification' and 'Normalization Calculations and Analysis'.

Molecular Weight Estimation: A detailed description of protein's apparent size on electrophoresed SDS-PAGE gels or probed nitrocellulose membranes is provided in the 'Computational Analysis' section of this Chapter (see Section 2.7.4).

2.3.4 Bacterial Protein Expression

Proteins of interest were expressed by either transforming BL21(DE3) *E. coli* cells (NEB, C2527I) or using a cell-free system to purify recombinant proteins. These proteins can then be used for various applications, such as antibody generation, *in vitro* protein-protein interaction assays, *in vitro* protein-RNA interaction assays, etc.

Protein Expression in Bacteria: For *in vivo* bacterial protein expression, 50ng of the expression vectors suitable for the T7 RNA polymerase-IPTG induction system were transformed into BL21(DE3) chemically competent *E. coli* cells and plated on appropriate selective media to grow overnight at 37°C. A liquid culture was prepared by inoculating a single colony in a selective LB broth media for 6-8 hours or overnight at 37°C. A fresh culture was prepared by adding 100µL of the overnight culture to 11mL of selective LB broth media and incubated at 37°C at 250 rpm until the OD_{600nm} reached 0.4-0.6 (approximately 2.5 – 3 hours). Once the optimal optical density was reached, protein expression was induced by adding 1mM IPTG (Sigma-Aldrich, I6758) to the

remaining culture, achieving a final concentration of either 1 μ M or 0.5 μ M. The IPTG-induced bacterial cell culture was then incubated at 25°C at 230rpm. A 1mL aliquot of the culture was collected hourly for 4 hours, after which the remaining cell culture was moved to 16°C for overnight incubation. In the meantime, the collected samples were stored at 4°C until needed for further processing.

To extract the proteins expressed by the BL21 cells, the IPTG-induced bacterial cells were first harvested by centrifugation at 16,000 \times g for 10 minutes. Then, the harvested cells were resuspended in NEBExpress *E. coli* Lysis Reagent (NEB, P8116) by brief pipetting or vortexing until the suspension is homogenous. The exact volume of the Lysis buffer used depends on the optical density reading at 600nm (OD600) of the IPTG-induced bacterial culture upon harvesting, as described in the manufacturer's manual. The homogenate was incubated at room temperature with gentle shaking for 20 minutes or until the suspension is clear or visible. Lysates were then centrifuged at 16,000 \times g for 10 minutes at 4°C to pellet insoluble materials and cell debris. The cleared extracts were then transferred into a sterile microcentrifuge tube to be used as the soluble fraction. For the insoluble fractions, the pellet formed after centrifugation of the bacterial cell lysate was treated differently. Proteins in the insoluble pellets were resuspended in 50mM Tris-HCl (pH 7.5) using an equal volume to that of the NEBExpress *E. coli* Lysis Reagent employed during the lysis step. These fractions could be stored long-term at -20°C or used immediately for further analysis or purification.

Cell-Free Protein Synthesis: For *in vitro* bacteria-free protein expression, a 50 μ L reaction of NEBExpress® Cell-Free *E. coli* Protein Synthesis System (NEB, E5360) was setup containing 125ng of expression vectors suitable for the T7 RNA polymerase induction system. Reactions were prepared according to the manufacturer's instructions in Eppendorf tubes, gently mixed by vortexing, sealed to prevent evaporation (Parafilm, PM992), and incubated overnight at 30°C with vigorous shaking at 250rpm. With every protein expression experiment, a no-template control (NTC) and a dihydrofolate reductase (included in the kit) reactions were set as negative and positive controls, respectively. After completion of the *in vitro* protein synthesis, reactions were stored at -20°C for long-term storage or immediately processed for analysis or purification.

SDS-PAGE Analysis: A 10 μ L sample of either the soluble or insoluble fraction was prepared containing SDS-PAGE Loading buffer, heated at 95°C for 5 minutes, and then loaded into 15% SDS-PAGE gel for protein analysis: to examine protein quality and estimate protein quantity. Proteins were separated by electrophoresis following the conditions described for fluorescent Western blot experiment, refer to Section 2.3.3.

Coomassie Protein Staining: After electrophoresis, proteins separated on SDS-PAGE gels were washed once with distilled water and then incubated directly in 10 mL of InstantBlue® Coomassie Protein Stain (Abcam, ab119211) for 15 minutes. Stained gels were rinsed twice in distilled water before being imaged using the Odyssey® CLx Infrared Imaging System (LI-COR Biosciences, Model# 9140). Image analysis was performed using Li-Cor software ‘Image Studio’ (version 5.2) and ImageJ (version 1.8.0_181 64 bits).

Protein Purification: His-tagged proteins were purified using the NEBExpress Nickel Spin Columns (NEB, S1427), according to the to the manufacturer’s instructions.

2.4 Biochemical RNA-Protein & Protein-Protein Interaction Assays

2.4.1 RNA-Affinity Pulldown Assay

RNA–protein interactions play a crucial role in various cellular processes and significantly contribute to the regulation of post-transcriptional gene expression. Therefore, to identify and characterize these interactions, an *in vitro* RNA-centric method called the RNA-affinity pulldown assay was employed. The RNA-affinity pulldown assay was carried out as previously described by Besse *et al.* (2009). To briefly describe the procedure, this *in vitro* system involves generating pBSII constructs for T7 *in vitro* synthesis of RNA in the presence of biotin-UTP to generate biotinylated sense RNA molecules. The resulting labelled RNA molecules are then immobilized on streptavidin beads and incubated with a protein extract derived from ovarian tissues, allowing soluble RBPs to recognize RNA bait molecules containing their specific RNA-binding motif and selectively precipitating RBPs from the sample. After the binding of RBPs and biotinylated-RNA is complete, the assay takes advantage of the high affinity of biotin for streptavidin to pulldown *in vitro* assembled protein–RNA complexes. The

RBPs that are pulled down because of their interaction with RNA bait are typically identified and detected by western blot (Figure 2.7).

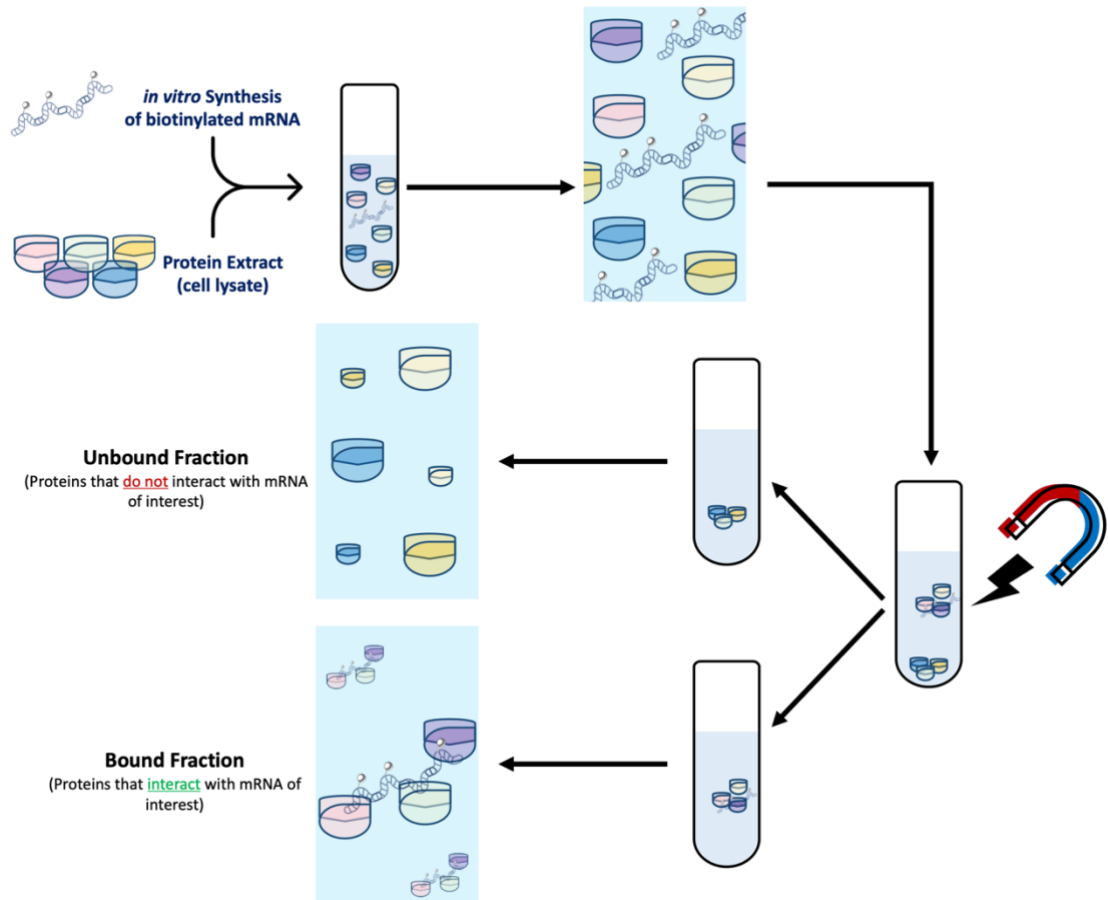


Figure 2.7: Biotinylated RNA affinity pulldown assay for the detection of Protein-RNA interactions. The assay allows for the identification of RNA-binding proteins (RBPs) that bind directly to the RNA of interest in a cell-free system. Moreover, it can also be used to map the interactions to specific regions of the RNA sequence with which the RBP interacts, and to characterize the RNA-binding motif through site-directed mutagenesis. The assay involves *in vitro* synthesis of RNAs of interest in the presence of biotin-UTP to generate biotinylated RNA molecules. These tagged RNAs are then incubated with a cell lysate to allow RBPs to recognize the regions of the RNA to which they bind (bound fraction), while RNA lacking their binding sites do not interact with RBPs (unbound fraction). The RNA-protein complexes are then isolated using streptavidin-coated magnetic beads. RBPs pulled down with biotinylated RNA are typically detected by Western blot analysis.

***In vitro* Synthesis of Biotinylated RNA Probes:** Biotinylated sense RNA probes were synthesized for *tsunagi* (referred to *y14*), *oskar*, *delta*, *notch*, *shotgun* (*Drosophila* DE-Cadherin), *gurken*, *bicoid*, *nanos*, and *K10*. The labelled probes were synthesized as described in Section 2.3.1. The 3'UTR of *oskar* and the coding sequence of *y14* were used as a positive and negative control, respectively.

Preparation of Ovarian-derived Protein Extracts: Ovarian-derived protein extracts were prepared using adult females of the *w¹¹¹⁸* genotype 1–4 days old post-eclosion that were collected in fresh vials. The females were then kept for 2–3 consecutive days at room temperature in the presence of dry yeast and males (at a 3:1 sex ratio). Soluble S10 ovarian extracts were prepared from 25 pairs of ovaries per biotinylated-RNA point, as described in Section 2.3.2. Usually, ovaries were pooled together to make a master ovarian cell extract that is equally divided between RNA-affinity pulldown reactions. The extract was always kept on ice and immediately used, see below the ‘Affinity Pulldown of RNA-RBP Complexes’ section.

Equilibration & Preparation of Streptavidin Magnetic Beads: An aliquot of 40µl magnetic beads conjugated to streptavidin particles (NEB, S1420S) was used per biotinylated-RNA point. The beads were washed three times in 1mL of ice-cold NEB Wash buffer (20mM Tris-HCl at pH 7.5, 1mM EDTA at pH 8.0, 0.5M NaCl) for 1 minute each, on a nutating mixer (hereafter referred to as ‘rotating platform’ in subsequent sections). The wash solution was removed by applying a magnet (DynaMag™-2; Invitrogen, 12321D) for approximately 30 seconds. The magnetic beads were then resuspended in 250µL of NEB Wash buffer. The equilibrated beads were then kept on ice, ready to be used in the next step (see below).

Binding of Biotinylated-RNA Probe to Beads: Equimolecular amounts of biotinylated RNA – typically using 4µg of the control *oskar* 3’UTR as reference – were added to each point of equilibrated streptavidin magnetic beads. The bead-biotinylated RNA mixture was incubated for 30 minutes at room temperature on a rotating wheel to allow the binding of the biotinylated RNA molecules to the streptavidin beads. Excess and residual unbound RNA molecules were removed by performing two washes, each with 1mL of NEB Wash buffer for 1 minute, on a rotating platform. The RNA-bead mix was then washed twice in 1mL of ice-cold Binding buffer (10mM HEPES at pH 7.9, 3mM MgCl₂, 5mM EDTA at pH 8.0, 5% glycerol, 2mM DTT, 0.5% IGEPAL, 40mM KCl) for 1 minute each.

Affinity Pulldown of RNA-RBP Complexes: A master mix of the S10 ovarian extract was prepared, to achieve a total volume of 150µL of soluble proteins was used per point

of biotinylated RNA. This S10 ovarian extract master mix was prepared with Binding buffer (refer to composition detailed above), supplemented with 3 µg/µL of heparin and 0.5µg/µL yeast tRNA to reduce non-specific binding. The resulting mix was equally divided between each of the biotinylated RNA-bead complexes (i.e., 150µL/point) before reactions were incubated for 1.5 hours at 4°C with constant mixing on a rotating wheel. After the incubation has completed, unbound fractions were recovered for western blot analysis by applying the samples to a magnetic rack and collecting the supernatants. The beads were then washed four times for 5 minutes each with 500µL of ice-cold Binding buffer lacking heparin or tRNA. The bound fractions were then collected by eluting the bound proteins in 35µL of 2× SDS-PAGE Loading buffer, following a 5-minutes incubation at 90°C. The bound fractions were transferred into clean Eppendorf tubes and kept on ice until required, or stored at -80°C for long-term storage (i.e., overnight or longer period).

Detection of RNA-RBP Interactions via Immunoblotting: A 15µL of the recovered bound fraction in 2× SDS-PAGE Loading buffer was diluted with equal volume of analytical water (final concentration of 1×) before incubating at 95°C for 5 minutes. Extracts were run on 9% SDS-polyacrylamide gel alongside a protein ladder, and were then probed for the detection of Shep and PTB proteins in the bound fraction. Similarly, the unbound fractions were run as described for the bound fractions and were probed to detect a soluble protein (dynein heavy chain; Dhc) as a loading control, ensuring equal amounts of proteins were used in each RNA pulldown reaction.

2.4.2 Co-Immunoprecipitation (Co-IP) Assay

Co-immunoprecipitation (Co-IP) is a widely used technique in molecular biology to investigate protein-protein interactions providing key insights into the molecular mechanisms of biological processes critical for cellular function. Briefly, antibodies that bind the proteins of interest are incubated with a protein extract derived from ovarian tissues to capture and precipitate any associated binding partners. The samples are then subjected to series of washes to remove any non-specific interactions, and the remaining proteins that are pulled down because of their interaction with the target protein are typically identified by western blot (see **Figure 2.8**).

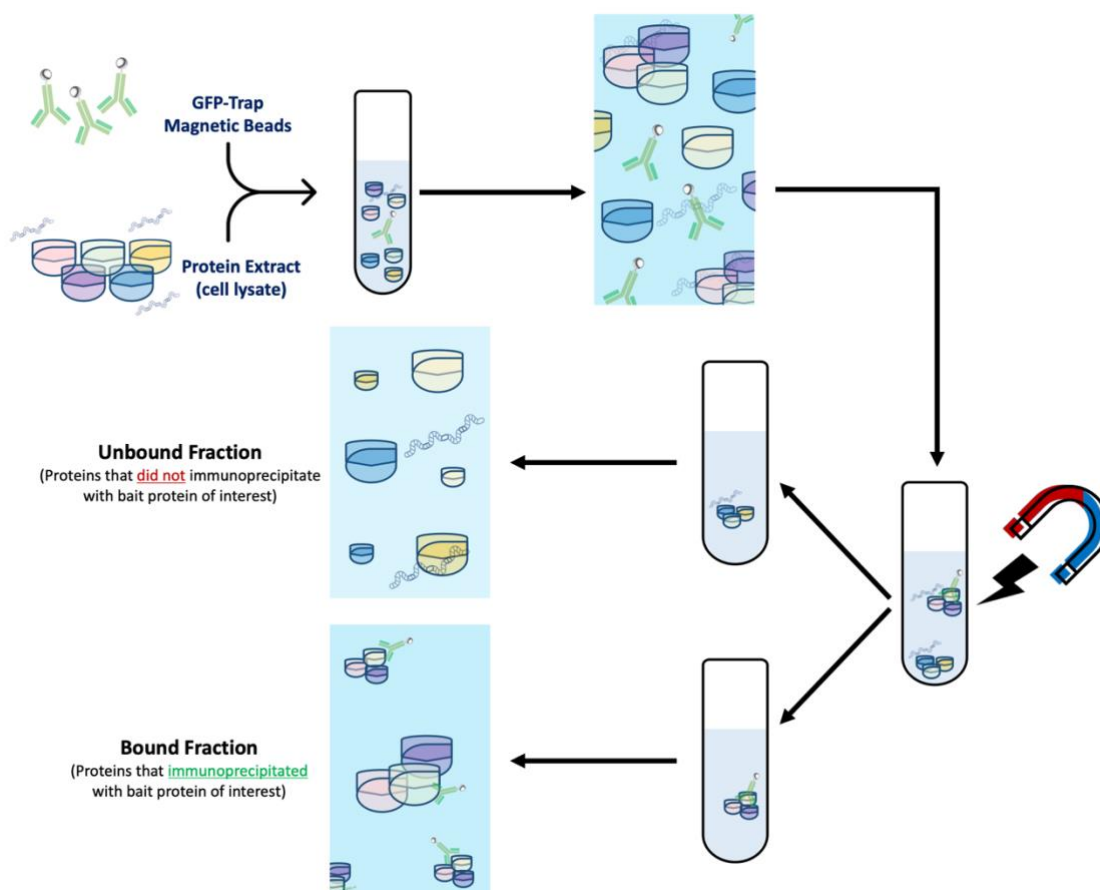


Figure 2.8: Co-Immunoprecipitation assay for the detection of protein complexes. The assay allows for the characterization of protein complexes in a cell-free system by incubating a cell lysate with antibodies against a protein of interest. The complexes containing the target protein are then immunoprecipitated using the antibody-coupled to the magnetic beads (bound fraction). Protein components in the pulled down complexes are typically detected by Western blot analysis.

Preparation of Ovarian Lysates: Protein extracts were prepared from ovaries of the Shep GFP protein-trap line, or UAS-driven eGFP control females. Instead of wildtype ovaries, eGFP was used as a negative control. Females were prepared for dissection according to the standard method described in Section 2.5.4. The ovarian extract was prepared using 75-150 pairs of ovaries per co-immunoprecipitation point. After dissection, ovaries were rinsed once with ice-cold PBS and once more in Lysis buffer (10mM Tris-HCl^{pH 7.5}, 150 mM NaCl, 0.5mM EDTA^{pH 8.0}, protease inhibitor cocktail (Roche), 1 mM PMSF, 0.5% NP-40 aka. IGEPAL). Washed ovarian tissues were then resuspended in 100 μ L of ice-cold Lysis buffer and homogenized on ice using 30 strokes of a Dounce homogenization pestle. The resulting lysate was cleared by centrifugation at 11,000 rpm for 15 minutes at 4°C to remove cellular debris before the supernatant was collected. An appropriate volume of the Dilution buffer (10mM Tris-HCl^{pH 7.5}, 150

mM NaCl, 0.5mM EDTA^{pH 8.0}, protease inhibitor cocktail, 1 mM PMSF), lacking the NP-40 detergent, was added to the supernatant up to 550µL. A 10% portion of the resulting diluted lysate (i.e., 50µL) was retained as input fraction for western blot analysis. Typically, ovaries of the same genotype were pooled together to make a master ovarian protein extract. The extracts were always kept on ice and processed immediately by proceeding to the following step of the assay, the ‘Binding of Beads to Protein Extract’ section.

A modified version of the CoIP protocol was used to determine if the detected protein-protein interactions were RNA-dependent. In these experiments, the extracts were pre-treated with RNases before protein complexes were immunoprecipitated with GFP-trap nanobodies. A test was initially performed to determine the optimal concentration of nucleases and digestion conditions to ensure efficient digestion of endogenous RNA in the ovarian extract. After determining the optimal conditions needed for the RNase treatment, extracts were obtained as previously described (see above). Following the homogenization and centrifugation steps, the clear lysate was treated with RNases by adding 1/10 of RNase A (10 mg/mL; Thermo Fisher, EN0531) and 1/20 of diluted RNase T1 (100,000 U/mL; Thermo Fisher, EN0541) directly to the extracts. RNases were then incubated with the extracts for 1 hour at 4°C. After the completion of the RNase treatment, the extracts were processed as described above by adding Dilution buffer up to 550µL.

Equilibration & Preparation of Nano-Trap Affinity Beads: An aliquot of 25µL GFP-Trap® Magnetic Agarose beads (ChromoTek, gtma) was used with each immunoprecipitation point. The beads were rinsed three times in 1mL of ice-cold Dilution buffer by applying a magnet (DynaMag™-2) for approximately 30 seconds. The magnetic beads were then resuspended in 500µL of Dilution buffer per point. The equilibrated beads were kept on ice until required or processed immediately by proceeding to the next step of the assay, the ‘Binding of Beads to Protein Extract’ section. Once the beads were ready for incubation with soluble proteins, the Dilution buffer in which the beads had been resuspended was removed.

Binding of Beads to Protein Extract: The remaining volume of protein extracts (i.e., 500 μ L) were incubated with equilibrated beads for 1 hour at 4°C with constant mixing on a rotating platform. Following the incubation step, the unbound fraction (i.e., flowthrough) was collected for western blot analysis, and the beads were washed 3 times with Wash buffer (10mM Tris-HCl^{pH 7.5}, 150 mM NaCl, 0.5mM EDTA^{pH 8.0}, 1× PI Cocktail (Roche), 1 mM PMSF, 0.05% NP-40 aka. IGEPAL) by gently inverting the tube 10 times at 4°C. The pellets (i.e., bound fractions) were resuspended in 32 μ L of 2× SDS-PAGE Loading buffer and boiled at 95°C for 10 minutes to recover immunoprecipitated protein complexes for western blotting analyses. The bound fractions were transferred into clean microcentrifuge tubes and kept on ice until required or stored at -80°C for long-term storage.

Detection of Protein-Protein Interactions via Immunoblotting: 7.5-15 μ L of input or flowthrough fractions (approximately 1.5-3% of total ovarian starting material) were prepared in SDS-PAGE Loading buffer, whereas 100% of the recovered bound fractions (i.e., 32 μ L) were used in the 2× undiluted SDS-PAGE Loading buffer. The samples were incubated at 95°C for 5 minutes before loading into the wells. Along with a protein ladder, the samples were run on 10% SDS-polyacrylamide gel and probed for the detection of Shep and PTB proteins first. Then, membrane was stripped and reprobed for the detection of GFP.

2.4.3 RNA Immunoprecipitation Assay

RNA Immunoprecipitation (RIP) assay is a powerful technique widely used in molecular biology to investigate the interactions between RNA molecules and RNA-binding proteins (**Figure 2.9**). This methodology allows for the identification and characterization of RNA targets that interact with the RBP of interest in a cell-specific context, providing valuable insights into post-transcriptional gene regulation and RNA metabolism within a cell or tissue. This information is crucial for understanding the roles of RBPs in various biological processes that contribute to development. By shedding light on the interplay between RNA and RBPs, RIP assay has emerged as an essential tool in unravelling the intricate landscape of post-transcriptional gene regulation, ultimately paving the way for understanding the underlying mechanisms that govern cell function.

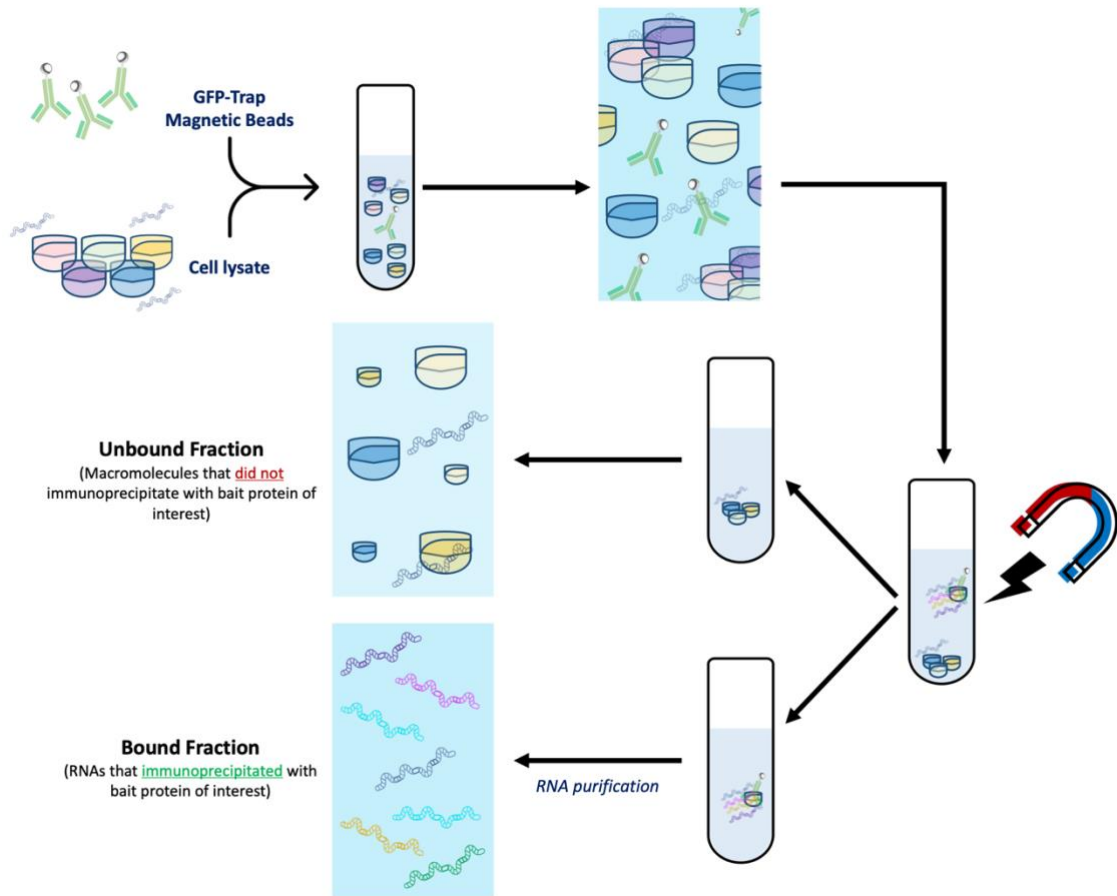


Figure 2.9: A Schematic of the RNA Immunoprecipitation Assay for the Identification of RNA Targets. The assay allows for the identification of protein-RNA interactions in a cell-free system by incubating antibodies against an RNA-binding protein of interest with a cell lysate. Complexes containing the RBP bait and RNA molecules are then immunoprecipitated using antibody-coupled magnetic beads (bound fraction). RNA targets in the pulled down fraction are typically identified by Sequencing (RIP-seq) or reverse transcription polymerase chain reaction (RIP-RT-PCR).

To identify candidate mRNAs bound to Shep-GFP in *Drosophila* ovaries, RNA immunoprecipitation was carried out as described for the CoIP assay with slight modifications to maintain an RNase-free environment. This was achieved by cleaning the workspace, equipment, and homogenizing pestle with either RNaseZap Solution or an alternative, RNaseAWAY surface decontaminant. Around 150 pairs of fresh ovaries were dissected for each genotype and homogenised in 100 μ L of Lysis buffer containing 50 Units of RiboShield™ RNase Inhibitor (PCR Biosystems, PB30.23) to work in a final concentration of 0.5U/ μ L. After centrifugation of the lysate to remove cellular debris and collection of the supernatant, a Dilution buffer supplemented with RiboShield™ RNase Inhibitor (0.5 U/ μ L) was added to reach a final volume of up to

550 μL . Subsequent sample preparation, nanobodies preparation, and immunoprecipitation procedure were carried out as described for CoIP assay. Following immunoprecipitation of the GFP-containing complexes, containing the bound RNAs, the bound fraction was resuspended in 150 μL of RNA Lysis buffer (provided by RNA purification kit) at 65°C for 10 minutes, instead of the 2 \times SDS-PAGE Loading buffer used in the CoIP protocol. RNA extraction was performed using the RNeasy[®] Micro Total RNA Isolation kit (Qiagen, AM1931) following the manufacturer's instructions. The purified RNAs were eluted in 25 μL of the elution solution provided by the kit. To ensure RNA integrity in the samples, and to prevent potential degradation from RNase contamination, 1 μL of RiboShield[™] RNase Inhibitor was added to the eluted RNA sample as a precautionary measure. RNA samples were kept at -20°C for long-term storage.

DNase Treatment and cDNA Synthesis: After RNA quantification using either a NanoDrop or Qubit, 100ng of purified RNA from each sample was treated with the DNase I supplied by the RNA purification kit to remove traces of gDNA in a 10 μL reaction. Subsequently, first-strand cDNA templates were synthesized using UltraScript 2.0 cDNA Synthesis kit in a 20 μL reaction containing 50ng of RNA as template (i.e., 5 μL of DNaseI-treated RNA).

Semi-Quantitative RT-PCR: The RIP-PCR enrichment analysis was performed in a 12.5 μL PCR reactions using redTaq DNA Polymerase (PCR Biosystems, PB10.13), containing 1.25ng of cDNA template and primers designed for qPCR (refer to Appendix 1 for primer sequences). Semi-quantitative RT-PCR analysis was carried out using equivalent amounts of cDNA template to ensure comparability when investigating the enrichment of specific mRNA candidates in the different immunoprecipitation samples. The target mRNA candidates analysed for enrichment included *oskar*, *gurken*, *bicoid*, *nanos*, *shotgun* (Drosophila DE-Cad), *delta*, and *notch*.

2.5 Fly Husbandry & Routine Fly Work

2.5.1 General Fly Maintenance

D. melanogaster was used as the *Drosophila* fruit fly stock throughout this study. Stocks were cultured on a standard maize-based food medium containing dextrose, autolysed yeast, agar, and anti-microbial agents such as propionic acid and Nipagin. The recipe used for culturing fruit fly stocks was adapted from the 'Cornmeal Food' recipe provided by the Bloomington Drosophila Stock Centre (BDSC), with modifications to the composition of ingredients. The BDSC recipe was obtained from their website (<https://bdsc.indiana.edu/information/recipes/bloomfood.html>) and Table 2.10 provides a summary of the modifications made to the recipe. Flies were grown at room temperature between 21-24°C and were placed on shelves next to a window to expose them to natural day-night cycles (Cardiff, Wales, UK). Established stocks were maintained by regularly “flipping” (i.e., transferring) adult flies to new vials with fresh food every three weeks.

Table 2.10: Recipe to prepare 1L of cornmeal-based fly food. *: Nipagin, also referred to as Tegosept, is chemically known as p-hydroxybenzoic acid methyl ester. Nipagin is dissolved in ethanol to prepare a 10g/L stock solution. **: The fly food was cooked in bulk by our brilliant lab technician, Sabrina Williams.

Component(s)	Composition per 1L of Fly Food **	
	Quantity	Percentage
Fine Maize Meal (Southern Milling, 20050488)	72.5 g	7.25%
Dextrose Monohydrate (Flystuff, FLY1156)	75 g	7.5%
Autolysed Yeast (Kerry, 20050488)	35 g	3.5%
Agar (supplied by BTP Drewitt)	7 g	0.7%
Propionic Acid (Thermo Scientific™, 149300025)	3.5 mL	0.35%
Nipagin * (Flystuff, FLY1136)	22 mL*	2.2%
Water	Up to 1L	78.5%

Fly Husbandry: Fly stocks were maintained under standard lab conditions at room temperature on cornmeal-based fly food, and regularly transferred to fresh medium, as previously described. However, the specific conditions that flies were reared at varied, depending on the experimental design. A summary of the conditions is described:

- ↳ For extraction of ovarian lysate used in biochemical assays, young adult females (1-3 days post eclosion, DPE) were collected and kept at 25°C with males and yeast for at least three days prior to dissection.
- ↳ For immunostaining using *w¹¹⁸* flies, young adult females were collected, mated with males in the presence of yeast, and maintained at 25°C for at least three days prior to dissection.
- ↳ For ovarian phenotypic analysis using the Shep protein-trap line, young adult females were collected and kept at 25°C with males and yeast for a minimum of three days prior to dissection or downstream experimentation.
- ↳ For RNAi experiments, flies carrying tissue-specific Gal4 drivers were crossed with those harbouring RNAi line of interest. The F0 generation of the crosses were kept at 25°C for three days before a copy of the cross was made and F1 embryo/larvae were shifted to 29°C to increase RNA interference efficiency. Here, the F1 generation was permanently kept at 29°C. Young adult F1 females were collected, fed, mated with males in the presence of yeast, and maintained at 29°C for at least seven days prior to dissection. The only exception to these conditions applies to RNAi experiments targeting the Notch signalling pathway. Due to the adverse effects on oogenesis, the crosses and their F1 progeny were maintained at temperatures between 18°C and room temperature.
- ↳ For *shep* loss-of-function analysis where flies carrying deficiencies were crossed with those carrying various *shep* P-element insertions. The crosses were maintained at room temperature. Subsequently, young adult F1 females of the correct genotype were collected and kept at 25°C with males and yeast for a minimum of three days prior to dissection.

- ↳ For overexpression experiments, flies carrying tissue-specific Gal4 drivers were crossed with the designated overexpression stock. The crosses were kept at either room temperature or 25°C for three days before a copy of the cross was made and F1 embryo/larvae were maintained at the same temperature. Subsequently, young adult F1 females were collected and kept at 25°C with males and yeast for a minimum of three days prior to dissection.

With all experiments, particularly those for kept at 25°C, the F0 flies were regularly flipped into fresh vials to prevent them from sticking to wet food.

Quarantining Procedure: All new incoming *Drosophila* stocks were kept isolated from the main stock collection to ensure *Drosophila* cultures remain 'mite free'. Any newly arrived stocks will undergo quarantine where vials are closely monitored at room temperature for the presence of mites for 3 full generations. Stocks infested with mites or their eggs were disinfected according to the 'ridding fly cultures of mites' method as described by Ashburner *et al.* (2005; Chapter 34 Section 11.4) and Ashburner and Roote (2007) which includes daily transfer, washing, and chemical treatment. To briefly describe the procedure, adult flies are transferred to vials with fresh medium daily for few weeks. Alternatively, dechorionated fly eggs with diluted bleach or pupae washed in 70% ethanol are transferred to vials with fresh medium to establish mite-free cultures.

Depending on when adult flies were needed for experimentation, the temperature at which the flies were kept was adjusted to either accelerate or delay their progression through the different developmental stages (**Figure 2.10**).

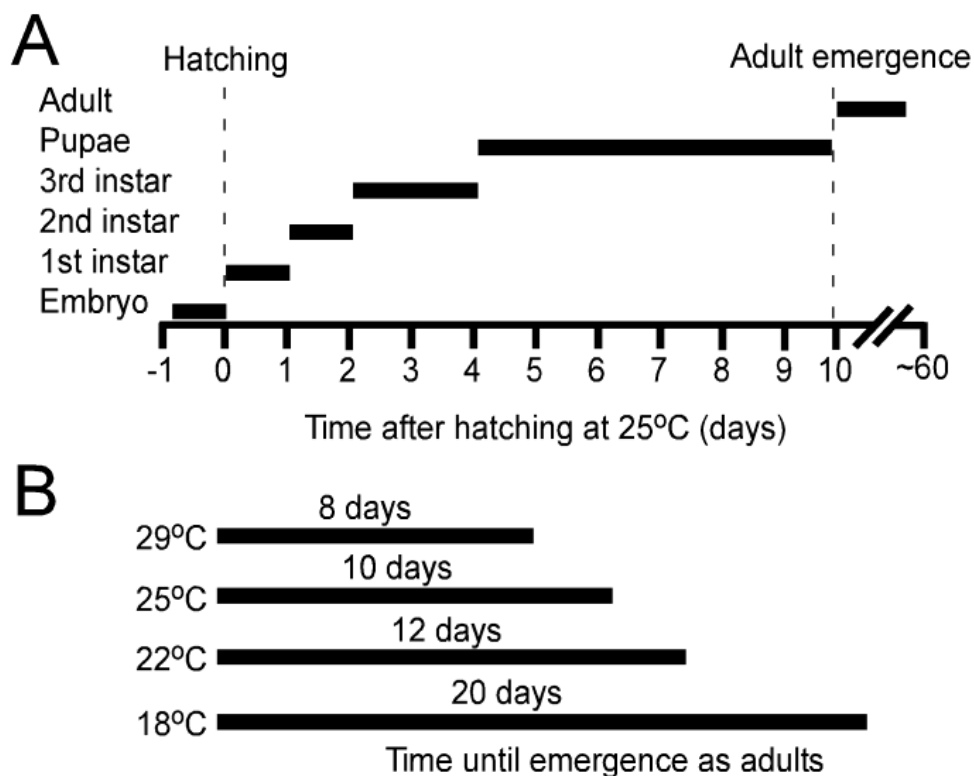


Figure 2.10: The impact of ambient temperature on the developmental progression of *Drosophila melanogaster*. Throughout a holometabolous insect's life cycle, the insect goes through different developmental stages, including the embryonic, larval, pupal, and adult stages. The progression through the life cycle of a fruit fly is temperature dependent as it influences the metabolism, therefore affecting the rate of development and duration of each stage. At lower ambient temperatures, the metabolic rate slows-down, leading to a prolonged duration of each developmental stage and a delayed emergence of the adult flies. Conversely, an increase in ambient temperatures results in speeding-up metabolism and ultimately shortening the developmental duration. Diagram obtained from Pulver and Berni (2012) with permission.

2.5.2 Fly Stocks

A variety of *D. melanogaster* strains were utilized in this project for different applications, as summarized in Appendix 6. These strains encompass stocks of wildtype, Protein-Trap, Gal4 drivers, UAS constructs, RNAi constructs, P-element insertion mutants, and deficiencies. Apart from the transgenic stocks newly generated in this project (see Appendix 4.3), others already available stocks were obtained from various sources, including those generated by the SLQ lab, gifted to us, provided by collaborators, or acquired from *Drosophila* stock centers such as the Bloomington *Drosophila* Stock Centre (BDSC), The Vienna *Drosophila* Resource Center (VDRC), and KYOTO Stock Center (KSC). Most of those strains were used in the UAS-Gal4 system to achieve conditional gene expression, as described in Section 1.3.1 of the

Introduction. Briefly, this system enables tissue-specific induction of a transgene of interest. Furthermore, the w^{118} fly strain was used as a wildtype control throughout the project.

Balancing of Fly Stocks: Balancer chromosomes are rearranged chromosomes carrying multiple inversions that serve as an essential tool in fly genetics. They allow for the maintenance of deleterious mutations in stable stocks by preventing homologous recombination during meiosis (meiotic recombination). Balancing of stock's chromosomes is also essential when combining different stocks to make a new stock as they enable researchers to track and follow the inheritance of chromosomes in genetic mating schemes. This is possible because balancer chromosomes are lethal in the homozygous state, and are marked with dominant visible mutations to help with their identification. In this project, frequently used balancer chromosomes and other dominant markers, include those on the 2nd and 3rd chromosome, were used to balance fly stocks. This was achieved by crossing virgin females of the stock that needs balancing with males of a multiple-balancer stock for one or two generations in a row. The multiple-balancer stock used contains an allele of Kruppel (Irregulate facet, If) and the balancer Curly of Oster (CyO, marked with Cy) on the second chromosome, and TM3 marked with Stubble (Sb) and TM6B marked with Humeral (Hu) on the third chromosome.

2.5.3 Mating Scheme & Experimental Crosses

Virgin Collection Procedure: To ensure purity and reliability of a genetic cross, virgin female flies should be used when setting up cross schemes. Virgin females of the desired genotype were collected following a rigorous procedure and best standard practices described by Greenspan (2004), Roote and Prokop (2013), and Markstein (2018). To maximize virgin collection, vials containing dark lat-stage pupae from which adult flies are expected to eclose (i.e., end of pupal stage), were shifted to 18°C overnight to slow development and eclosure. The selection procedure starts by emptying vials thoroughly from any adult flies on food or vial walls. This is generally done first thing in the morning. Virgins were routinely collected twice a day, in the morning and late afternoon, until an adequate number for flies were obtained to set up a cross. During morning collections, only young females with multiple visible virgin markers were

selected. Visual markers are often present during the first few hours from eclosion and they include folded wings, pale coloration of the cuticle, and presence of flyconium (the fly equivalent of meconium). To maximize the number of virgins collected, vials were kept at room temperature (21-24°C) throughout the day but were shifted to 18°C overnight.

Setting Up Cross: Crosses were set up by matting 6-8 virgin female flies (♀) of a particular genotype with 3 male flies (♂) of a different genotype in fresh food vials. A small amount of dried instant yeast was sprinkled on the food before adding the males and virgin females together to increase the likelihood of a successful mating. Crosses were generally kept at 25°C unless otherwise specified and depending on the experimental design the F1 larvae were kept at 25°C or shifted at 29°C. For a detailed description of the different rearing conditions refer to the ‘Fly Husbandry’ section. To establish more copies of the cross, the parent flies (i.e., the F0 generation) were frequently transferred into fresh food vials supplemented with dry yeast every 3 days. This ensures that enough sample size of F1 flies to be obtained for experimental analysis. Appendix 6 summarizes the *Drosophila* stocks used in this study for experimentation purposes.

Combining of Stocks: Several stocks were generated by combining two transgenes together in the same fly stock, including: *oskar-nanos* driver (*w-; osk-Gal4:VP16/CyO; nos-Gal4:VP16/TM3,Sb*), *oskar* Gal4-driver in a Shep protein-trap background (*w-; osk-Gal4:VP16/CyO; Shep:GFP^{CC00236}*), *nanos* Gal4-driver in a Shep protein-trap background (*nos-Gal4;; Shep:GFP^{CC00236}*), *oskar* Gal4-driver in a *shep* deletion background (*w-; osk-Gal4:VP16/CyO; Df(3L)ED210/TM3,Sb*), *traffic jam* Gal4-driver in a Shep protein-trap background (*w-; tj-Gal4/CyO; Shep:GFP^{CC00236}/TM3,Ser*), *Notch* RNAi Valium20 constructs in a tubulin driven temperature sensitive Gal80 background for somatic expression (*w-; tub-Gal80ts/CyO; Valium20-N/TM3,Sb*), *Delta* RNAi constructs in a tubulin driven temperature sensitive Gal80 background for somatic expression (*w-; tub-Gal80ts/CyO; Valium20-D/TM3,Sb*), double *shep* RNAi construct for germline expression (*w-/7; Valium22-shep/CyO; Walium22-shep/TM3,Sb*), double *shep* RNAi construct for somatic expression (*w-; Walium20-shep/CyO; Valium20-shep/TM3,Sb*), *kinesin heavy chain* RNAi construct with UAS

mCherry tagged Shep-E construct (*w*-; UAS-mCh:Shep-E/CyO; Valium22-*khc*/TM3,Sb), and *oskar-bam* driver (*w*-; *osk*-Gal4:VP16/CyO; *bam*-Gal4:VP16/TM3,Sb). A summary of the individual fly stocks used to combine these different transgenes can be found in Appendix 6.

RNA Interference (RNAi) Analysis of *shep* in the Female Germline: To investigate the role of Shep in the female germline, the effects of *shep* gene knockdown were assessed using a short hairpin RNA interference system from the TRiP collection. To ensure maximum gene knockdown efficiency specifically in the germline, several crosses were established using virgin females of different germline-specific Gal4 drivers and males containing a double copy of UAS-driven RNAi hairpin. The different germline-specific Gal4 drivers include the double *oskar* Gal4-driver (*w*-; *osk*-Gal4:VP16/CyO; *osk*-Gal4:VP16/TM3,Sb), homozygous maternal triple driver (*ott*-Gal4:VP16; *nos*-Gal4; *nos*-Gal4:VP16), *nanos-oskar* driver (*w*-/*nos*-Gal4; *osk*-Gal4:VP16/TM3,Sb), *oskar-nanos* driver (*w*-; *osk*-Gal4:VP16/CyO; *nos*-Gal4:VP16/TM3,Sb), *oskar* Gal4-driver in a Shep protein-trap background (*w*-; *osk*-Gal4:VP16/CyO; Shep:GFP^{CC00236}), *nanos* Gal4-driver in a Shep protein-trap background (*nos*-Gal4; Shep:GFP^{CC00236}), *oskar* Gal4-driver in a *shep* deletion background (*w*-; *osk*-Gal4:VP16/CyO; Df(3L)ED210/TM3,Sb), and *oskar-bam* driver (*w*-; *osk*-Gal4:VP16/CyO; *bam*-Gal4:VP16/TM3,Sb). The genotype of male flies were *w*-/7; Valium22-*shep*/CyO; Valium22-*shep*/TM3,Sb for *shep* and *w*-/7; Valium22-*eGFP*/CyO; Valium22-*eGFP*/TM3,Sb for eGFP. The Gal4-drivers were crossed with RNAi hairpins of *shep* and *eGFP* to knockdown the expression of endogenous or recombinant Shep in the female germline. The Valium22-*eGFP* also served as a control for crosses lacking the Shep protein trap in the background. Ovaries were dissected from the resulting F1 female flies carrying both the Gal4-driver and RNAi hairpin constructs for subsequent FISH and confocal analysis. A list of transgenic fly stocks can be found in Appendix 6.

Testing *shep* RNAi Hairpins: To investigate whether the newly generated RNAi construct is capable of downregulating *shep* levels in the female germline, we set up two crosses. Virgin females from the *oskar* Gal4-driver in a Shep protein-trap background (*w*-; *osk*-Gal4:VP16/CyO; Shep:GFP^{CC00236}) were crossed with males containing two copies of the UAS-driven RNAi hairpin (*w*-/7; Valium22-*shep*/CyO; Valium22-

shep/TM3,Sb). For control, the same germline-specific Gal4 driver was crossed with *w¹¹⁸* control flies. Ovaries from the F1 females of the right genotype were dissected and fixed to visualize Shep-GFP expression.

RNAi Analysis of Notch Signalling Pathway Components: To investigate the involvement of the Notch signalling pathway in regulating Shep expression in the ovarian tissue of flies, male flies carrying UAS RNAi constructs targeting *Notch* and *Delta* mRNAs were mated with germline- and soma-specific Gal4 drivers in a Shep protein trap background. To achieve germline-specific knockdown, flies carrying the *oskar* Gal4-driver in a Shep protein-trap background (*w*-; *osk*-Gal4:VP16/CyO; Shep:GFP^{CC00236}/TM3,Sb) were crossed with flies carrying RNAi constructs targeting *Notch* mRNA (*w*+/*7*; +; Valium22-*Notch*/Valium22-*Notch*) and *Delta* mRNA (*w*+/*7*; Valium22-*Delta*/CyO; +). For RNAi expression in soma, flies carrying the *traffic jam* Gal4-driver in a Shep protein trap background (*w*-; *tj*-Gal4/CyO; Shep:GFP/TM3,Ser) were mated to male flies carrying RNAi construct against *Notch* or *Delta* in a temperature-sensitive Gal80 background driven by *tubulin*: *w*-/*7*; *tub*-Gal80ts/CyO; Valium20-*N*/TM3,Sb or *w*-/*7*; *tub*-Gal80ts/CyO; Valium20-*D*/TM3,Sb, respectively. Ovaries were dissected from the resulting F1 female flies with both the Gal4-driver and RNAi hairpin constructs to analyze the expression and localization of Shep using confocal microscopy.

Shep Overexpression Analysis: To investigate the effects of overexpressing different Shep isoforms in ovarian tissues of flies, male flies carrying UAS constructs driving mCherry-tagged Shep were crossed with virgin female flies carrying three different Gal4-drivers: the *oskar* Gal4-driver (*w*-; +; *osk*-Gal4:VP16/TM3,Sb) was used for germline expression, the *traffic jam* Gal4-driver under the control of a temperature-sensitive Gal80 was used for somatic cell expression (*w*-; *tj*-Gal4/CyO; *tub*-Gal80ts/TM3,Ser), and the border cell specific Gal4-driver (*w*-; *slbo*-Gal4/CyO; *slbo*-lifeAct:GFP/TM3,Sb) was used for expression in border cells. The genotype of the male flies were *w*-/*7*; UAS-mCh:Shep/CyO; + for Shep isoforms -A, -C, and -E. Ovaries from the F1 female flies with the correct genotype combination were dissected and analysed using FISH, immunostaining, confocal and biochemical assays. Additionally, fertility status and mature egg examination were performed on the F1 female flies.

Subcellular Localization of Shep: To visualize the effects of *kinesin heavy chain* gene downregulation on Shep-E subcellular localization pattern in the oocyte, female flies carrying double *oskar* Gal4-driver (w^- ; *osk*-Gal4:VP16/CyO; *osk*-Gal4:VP16/TM3,Sb) were crossed with males carrying *khc* RNAi construct and UAS mCherry tagged Shep-E construct ($w^-/7$; UAS-mCh:Shep-E/CyO; Valium22-*khc*/TM3,Sb). Ovaries were dissected from the F1 female flies for analysis using confocal microscopy.

Shep Loss-of-function Analysis: To investigate the role of Shep in the post-transcriptional regulation of *oskar* mRNA, we analysed various loss-of-function genetic backgrounds by combining different deficiencies and P-element insertions to each other. Virgin female flies carrying the Df(3L)Exel6103 or Df(3L)Exel6104 deficiencies were crossed with males of different genetic backgrounds. These males were either from w^{118} controls, carrying the Df(3L)ED210 deficiency, or bearing one of the following P-element insertions in the *shep* gene: *shep*^{C522}, *shep*^{BG02468}, *shep*^{KG10149}, and *shep*^{BG00836}. Ovaries from the resulting F1 female flies with the desired genotype were dissected and co-immunostained for Oskar and Gurken proteins.

2.5.4 Preparation of Female Flies for Ovarian Dissection, Egg Collection, or Fertility Assay

Ovarian development and egg chamber progression are dependent on various factors including genotype, age, and ambient temperature. As a general rule, for a healthy egg chamber production, adult females of a particular genotype 1–4 days old post-eclosion were selected and kept in vials with fresh food. Depending on the age of the fly, females were kept for 2–3 consecutive days at room temperature or at 25°C in the presence of males (at a 3:1 sex ratio) and yeast powder. Supplemented yeast powder encourages females to mate and lay eggs. To avoid females from sticking to the wet food, especially when kept at 25°C, flies were moved into a new vial of fresh food supplemented with yeast powder the day before dissecting or used in other downstream analyses. The terms well-fed females, conditioned females, mated females, or females from uncrowded vials are used interchangeably in the literature and throughout this section to indicate females prepared for dissection or egg collection.

Few exceptions were made to the general conditions described above, especially for females expressing RNAi constructs targeting the Notch signalling pathway. Those females were always kept at 18°C to reduce UAS-driven expression of the RNAi hairpins and thus, avoid arresting oogenesis: the Delta-Notch signalling pathway plays a crucial role in *Drosophila* oogenesis, regulating multiple aspects of ovarian development, and should not be strongly downregulated to ensure the progression of oogenesis. Moreover, females containing the temperature-sensitive Gal80, a negative regulator of Gal4 that blocks Gal4 activity at room temperature or 25°C, were shifted at 29°C to inactivate Gal80 and allows the binding of Gal4 to the upstream activating sequences to induce transgene expression.

2.5.5 Fertility Assay of *D. melanogaster* Adult Females

To assess the fertility of adult female flies of an indicated genotype, they were prepared in vials as previously described (see Section 2.5.4). The vials, each containing 6 female flies and 4 male wildtype control flies (*w¹¹¹⁸*), were allowed to mate and deposit embryos/eggs at 25°C for eight consecutive days on standard fly food containing yeast. On the 9th day, the parent flies were removed from the vials just days before adult offspring were expected to eclose. After 15 days at 25°C, the number of eclosed flies (i.e., progeny) was manually counted from each vial to determine the percentage of viable progeny produced by a genotype, which was calculated using the following equation:

$$\% \text{ of Viable Progeny} = \frac{\text{Total N}^{\circ} \text{ of adults Flies eclosed from Experimental Females}}{\text{Total N}^{\circ} \text{ of adults Flies eclosed from Control Females}} \times 100$$

For each genotype analysed, six replicates were obtained. When the parent females of the specified genotype were found to be completely sterile (i.e., 0% fertile, producing no viable progeny), vials were photographed using an iPhone 13 Pro Max camera. To establish the scale of the images, a Scienceware™ ruler was used to determine the length of the scale bar in inches.

2.5.6 Dissection of *D. melanogaster* Adult Ovaries

Under a stereomicroscope (Leica S4E) and the influence of carbon dioxide anaesthetization, ovaries from mated females were dissected in PBS (137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄, pH 7.4; Fisher Bioreagents, BP399). with a pair of forceps (no. 5, Stainless Steel 0.05×0.01 mm tip; Dumont Biology Tweezers), as described by Wong and Schedl (2006), Prasad *et al.* (2007), and Spracklen and Tootle (2013). Ovaries were subsequently cleaned-up by detaching other organs (such as the intestines and Malpighian tubules) with a pair of tweezers. For histochemistry applications, ovarioles were gently separated and carefully teased apart to allow for efficient penetration of antibodies and other reagents (including DNA stains). Ovaries were then transferred to an Eppendorf tube using a glass Pasteur pipette. The ovarian tissues were immediately processed according to the downstream procedure. Alternatively, the dissected ovaries were stored at -20°C (not in suspension, i.e., after discarding the PBS Dissection buffer). For RNA applications, roughly 10–50 pairs of ovaries were stored at -20°C in 100µL of RNA stabilization solution, RNAlater® (Sigma-Aldrich, R0901).

2.5.7 Measuring the Size of *D. melanogaster* Ovaries

Ovaries from well-fed adult female flies of a desired genotype were dissected in PBS with forceps under a stereomicroscope, as previously described in Section 2.5.4 and 2.5.6. To ensure reproducibility of measurements, three independent experiments were conducted for each genotype (N= 3 replicas), with each replicate consisting of a minimum of five females (n= a minimum of 5 pairs of ovaries). After ovaries were dissected, a glass Pasteur pipette was then used to transfer and carefully position one pair of ovaries into a drop of PBS placed on a glass microscope slide (76×26 mm, 1mm thickness; FisherBrand™, 11562203), as depicted in **Figure 2.11**. A darkfield illumination was used to image ovaries using a fluorescence upright microscope (Olympus BX50) at an overall magnification of 2×, utilizing an Olympus Plan 4×/0.10 ∞/- Objective and 0.5× C-Mount Camera Adapter. With each experiment, a stage micrometer (Watson, 0.1- & 0.01-mm scale) was also imaged using the exact same settings used to image the ovaries in order to help predefine the pixel-by-pixel size for measuring ovarian area from the digital images (see below). The digital images were

captured by JVC C-Mount Camera (3-CCD, KYF75U) and saved as TIFF files. To obtain ovarian area measurements in μm^2 , the outline of individual ovaries was traced in ImageJ (version 1.8.0, NIH) after predefining the pixel-by-pixel size in the software settings. A box plot was generated by Graphpad, showing ovarian area in μm^2 per single ovary. The Shapiro-Wilk normality test was performed on data before analysing the statistical significance between experimental groups. An unpaired two-sample t-test (aka. the student's t-test) was used to compare the means between different genotypes.

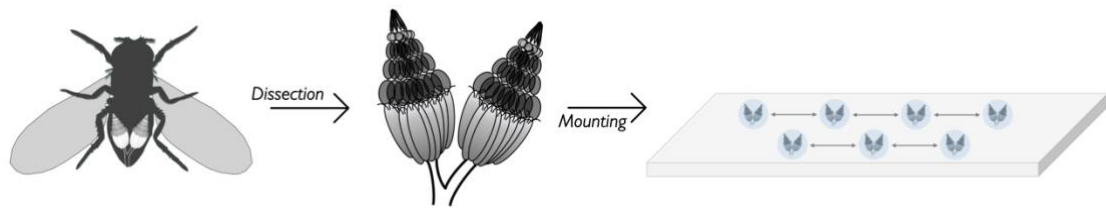


Figure 2.11: Layout of placing *Drosophila* ovaries on glass slide for imaging. The dissected ovaries were transferred onto a glass microscope slide by carefully placing a pair of ovaries per drop of PBS using a glass Pasteur pipette. This helps with the imaging and area measuring process by preventing the ovaries from crowding each other.

2.5.8 Quantification of the Number of Ovarioles Within *D. melanogaster* Ovaries

Ovaries from well-fed adult female flies of a desired genotype were dissected in PBS with forceps under a stereomicroscope, as previously described. To ensure results reproducibility, four independent experiments were conducted for each genotype (N= 4 replicas), with each replicate consisting of a minimum of five females (n= a minimum of 5 pairs of ovaries). The number of ovarioles per ovary was manually counted by carefully teasing apart the ovarioles using a pair of tungsten needles, as depicted in **Figure 2.12** and described by Wong and Schedl (2006). Experimental data is presented as a box plot generated by Graphpad, showing the ovariole count per ovary. The Shapiro-Wilk normality test was performed to check if the data follows a normal distribution before analysing the statistical significance between experimental groups. As the data was not normally distributed, a non-parametric test, specifically the two-tailed Mann-Whitney test, was employed to compare the means between genotypes.

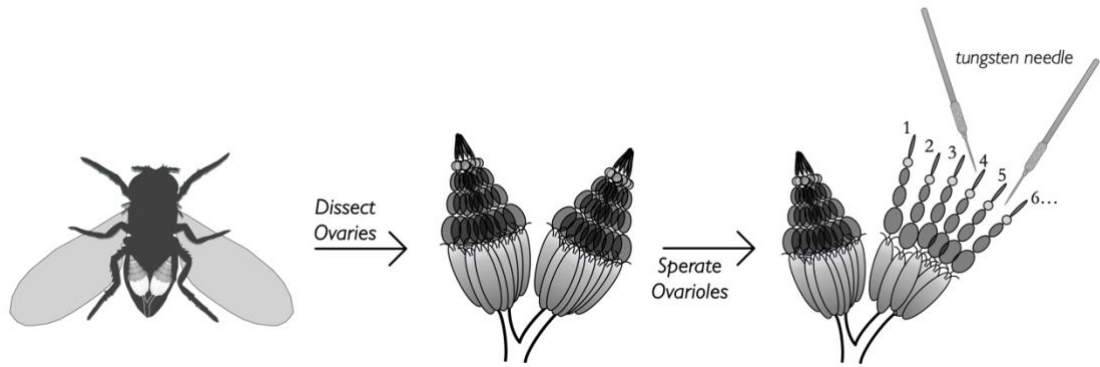


Figure 2.12: Quantification of ovarioles within *D. melanogaster* ovaries. To quantify the number of ovarioles per ovary, the dissected ovaries in PBS were gently teased apart using tungsten needles to isolate individual ovarioles.

2.5.9 Oviposition Assays – Collection of *D. melanogaster* Mature Eggs

Fruit fly agar plates supplemented with apple juice were used for the collection, examination, and quantification of eggs laid by adult female flies of various genotypes over a 24-hour period.

Preparation of Apple Juice Oviposition Plates: Apple juice agar plates for egg collection were made by dissolving 6g of glucose (Formedium, GLU01) in 60mL of apple juice, and boiled in a microwave oven on a medium setting for approximately 5 minutes. Once the sugar-rich juice mixture cools down, 3mL of 10% Nipagen M (dissolved in ethanol) was added to the mixture. In a different container, 6g of agar (Formedium, AGA01) was dissolved in 200mL of distilled water and boiled in a microwave oven at medium power for approximately 10 minutes. The molten-agar solution was allowed to cool down at room temperature until it was just comfortable to handle. The two mixtures were then combined and properly mixed by swirling. In a laminar flow cabinet, the media was carefully poured into sterile 60mm diameter Petri dishes to ensure even distribution on the bottom of the plates and avoid air bubble formation. The agar plates were left to solidify at room temperature and stored inverted at 4°C for no more than three weeks. In the literature, these plates are also sometimes referred to as egg-laying plates, embryo collection plates, and oviposition plates.

Preparation of Yeast Paste: Yeast paste was prepared by mixing instant dry yeast (e.g., Fermipan Red, Allinson's Easy Bake, or Saf-Instant® Red) with an equal volume of sterile distilled water. The mixture was thoroughly mixed until a paste consistency was

achieved. The yeast paste was prepared in small quantities and stored at 4°C for a maximum duration of two weeks to maintain freshness.

Preparation of Egg Collection Cage Apparatus: The ‘Egg Collection Cage Apparatus’ setup consists of an acrylic fly cage placed over an apple juice agar plate, which is secured in place with an elastic band. The fly cage is an acrylic cylinder (Flystuff, 59-100) designed to have openings at either side. The bottom opening fits a 60mm Petri dishes (i.e., apple juice agar plates), while the top end is covered with a 97µm stainless-steel mesh that allows for aeration and prevents condensation. This is depicted in **Figure 2.13**.

Oviposition Assay: Egg collection was conducted using young female adult flies of the desired genotypes that were 1-3 days old post-eclosion. To ensure flies had mated before starting the egg collection procedure, four female flies were allowed to mate in vials with two male flies at 25°C for 2-3 consecutive days, as described in Section 2.5.4. The flies were then transferred to the ‘Cage Apparatus’ to begin the egg collection procedure at 25°C. To further stimulate egg laying, a small quantity of freshly prepared yeast paste was placed at the center of the apple juice agar plate (as shown in the **Figure 2.13**). An acclimatization period in the cage apparatus is necessary to ensure consistent and reliable results, especially during the quantification assay. Therefore, the flies were allowed to acclimate for 24 hours before commencing the oviposition assay to assess the egg-laying capacity of female flies and to examine the morphology of mature eggs. After the 24-hour acclimatization period, female flies were allowed to deposit fertilized mature eggs on agar plates with yeast paste for another 24 hours at 25°C. After the egg collection period was completed, the fertilized mature eggs were either quantified to assess the egg laying capacity of females flies or examined for morphological defects.

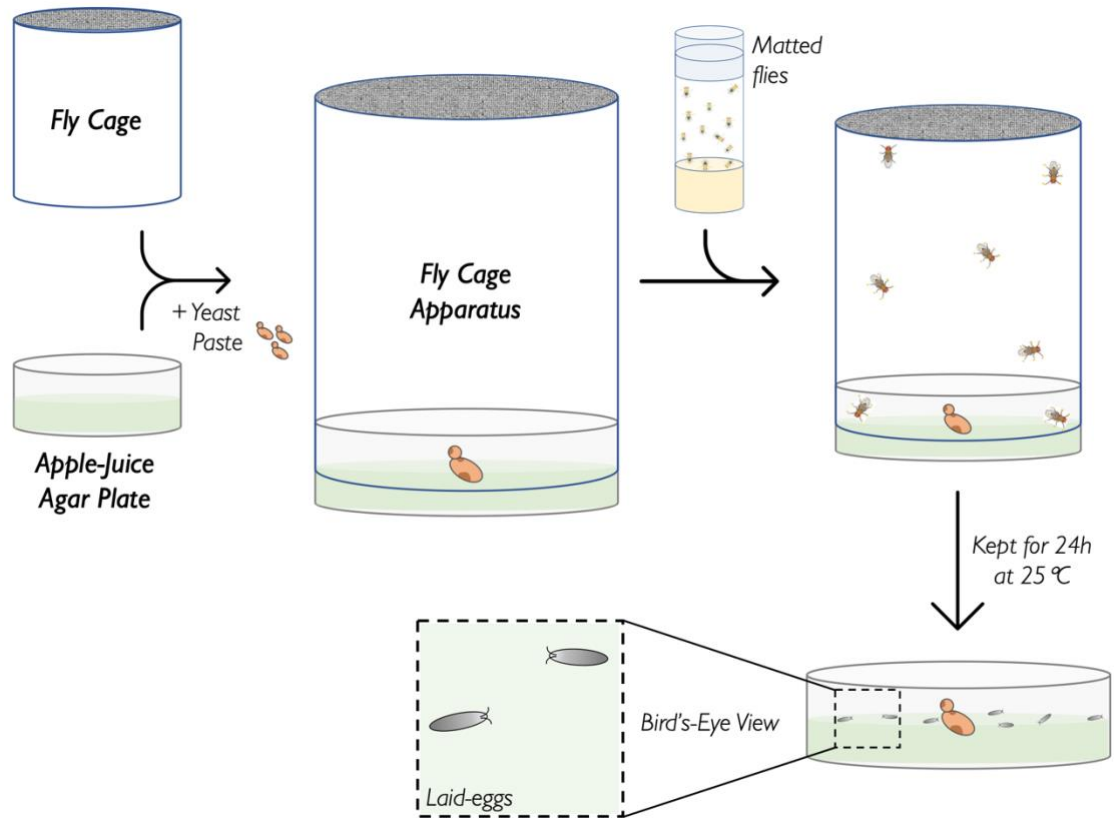


Figure 2.13: Workflow for collecting *Drosophila* eggs. The workflow of the egg collection procedure can be broken down into five phases. First, the components needed to set up the egg collection cage apparatus are gathered, which includes the fly cage and the apple juice agar plate. The second phase of the workflow involves preparing the female flies for oviposition by allowing them to mate with male flies in a fresh food vial with yeast for a period of 2-3 consecutive days at 25°C. In the third phase, the egg collection cage apparatus is set up by placing a small quantity of freshly prepared yeast paste at the center of the apple juice agar plate and securing the cage with the mated female and male flies. The fourth phase involves allowing the flies to acclimatize to the cage apparatus for 24 hours at 25°C. In the final phase, the eggs laid on agar plates were analysed to assess the egg-laying capacity of female flies and imaged to examine the morphology of mature eggs.

Quantification Assay: After the egg collection period was completed (see above), the apple juice plates were placed on a dark background (preferably black), and the eggs were manually counted under a light microscope (Leica). Eggs laid on the agar plate were first counted directly off the plates, whereas eggs laid in the yeast paste were processed differently before counting, in order to obtain a reliable number of eggs laid per female fly. For eggs laid in the yeast paste, the yeast paste containing the eggs was first transferred into an empty 90mm diameter Petri dish. The yeast paste was then carefully washed several times in distilled water. A wet paintbrush was used to disturb the paste, dissolving the yeast and releasing the eggs. To obtain the average egg count

per fly over the 24-hour period, the total count of eggs laid per plate was divided by the number of females per cage, which was calculated using the following equation:

$$\text{Eggs Laid per ♀ Fly} = \frac{\text{Total N}^{\circ} \text{ of eggs laid per plate}}{\text{N}^{\circ} \text{ of ♀ flies per cage}}$$

For each genotype analysed, laid eggs were scored from three independent experiments were obtained (n= 3 plates per genotype laid by three different sets of parental flies). The experimental data is presented as a bar graph generated by Graphpad, showing the number of eggs laid per fly for each genotype over a 24-hour period. The Shapiro-Wilk normality test was performed on the data before analysing the statistical significance between experimental groups. An unpaired two-sample t-test (aka. the student's t-test) was used to compare the means between different *D. melanogaster* genotypes.

Examination of Dorsal Chorionic Appendages: After the egg collection period was completed, eggs laid on the agar were gently transferred onto a slide in a drop of mounting medium (80% Glycerol, 2% N-Propyl Gallate, 4% Ethanol, in PBS) diluted in dH₂O. Eggs laid in yeast were first washed with PBS and then mounted on a slide. To examine the dorsal appendages, the mature eggs were orientated with the ventral side contacting the slide and the dorsal side facing upwards, using a wet paint brush (Daler-Rowney Graduate Round Brush #1). Examination of mature eggs was conducted, using phase-contrast microscopy, under both darkfield and brightfield illumination.

A darkfield Illumination of mature eggs dorsal appendages were Imaged using a fluorescence upright microscope (Olympus BX50) at an overall magnification of 4×, utilizing an Olympus Plan 4×/0.10 ∞/- Objective and 1× C-Mount Camera Adapter. The digital images were captured by a Hamamatsu ORCA-05G Camera using the Hamamatsu Corporation software (version 2.1.1.0) and saved as TIFF files. With each experiment, a stage micrometer (Watson, 0.1- & 0.01-mm scale) was also imaged under the exact same settings used to image the eggs in order to help predefine the pixel-by-pixel size on ImageJ. A scale bar in μm was determined using ImageJ (version 1.8.0, NIH).

For brightfield examination of the dorsal appendages, the eggs were examined using a fluorescence upright microscope (Olympus BX53) at an overall magnification of 4×, utilizing an Olympus UplanXApo 4×/0.16 ∞/-/OFN26.5 objective. The digital images were captured by a Hamamatsu ORCA-spark camera with a U-TV1×-2+U-CMAD3 T7 camera adapter, operating through the Olympus cellSens Standard software (version 2.3), and saved as TIFF files.

2.5.10 Whole-Mount Immunostaining of Adult *Drosophila* Ovaries

DAPI Staining: To visualize the overall egg chamber morphology of transgenic flies expressing fluorescently tagged proteins, ovaries obtained from well-fed female flies were fixed with 4% Formaldehyde solution (FA; Polysciences, 04018) prepared in PBS for 15 min at room temperature with constant orbital shaking on a rocking platform. To preserve the fluorescence of fluorescently tagged proteins, ovarian tissues were protected from the light throughout the procedure. After sample fixation, the ovarian tissues were permeabilized in PBST (PBS supplemented with 0.1% Tween® 20) for 3 hours at room temperature with constant shaking on a rocking platform and the wash solution was regularly changed every hour. The nuclei of ovarian cells were labelled by incubating the samples with 3 µg/mL of 4',6-diamidino-2-phenylindole dilactate (DAPI; Biotium, 40043) in 0.1% PBST for 5-10 minutes. The ovaries were then washed once in 0.1% PBST for 10 minutes, followed by the addition of Vectashield® antifade mounting medium (Vector Laboratories, H-1000). After this, the ovaries were kept overnight at 4°C before being mounted for confocal imaging.

Immunofluorescence Staining: To visualize the expression pattern and subcellular localization of proteins of interest in the female germline, ovaries obtained from well-fed female flies were fixed in 4% Formaldehyde solution for 25 min at room temperature with continuous agitation on a rocking platform. The ovarian tissues were washed twice in PBST (PBS supplemented with 0.1% Triton™ X-100) for 10 minutes each before being permeabilized in PBST (PBS supplemented with 1% Triton™ X-100) for 1 hour. The ovaries were then blocked in PBSTB blocking solution [0.5% Bovine Serum Albumin (BSA; Sigma-Aldrich, A2153) in 0.1% PBST] for 2 hours at room temperature. After blocking, the samples were incubated with primary antibody

diluted in PBSTB blocking solution overnight at 4°C. Refer to **Table 2.11** for primary antibody dilutions used for immunostainings. The following day, ovarian tissues were washed twice in PBSTB blocking solution at room temperature for 20 minutes each. The samples were blocked in PBSTN blocking solution [10% Normal Goat Serum (NGS; Gibco™, 16210064) in 0.1% PBST] at room temperature for 2 hours before overnight incubation at 4°C with the appropriate fluorescently conjugated secondary antibody prepared in 0.1% PBST. From this point onward, the samples were shielded from light to preserve fluorescent signals. After the overnight incubation, the ovaries were washed once in 0.1% PBST for 10 minutes at room temperature. This was followed by DAPI staining, another wash in 0.1% PBST, and then incubation in Vectashield® mounting medium. The samples were then mounted for confocal imaging.

Table 2.11: List of primary & secondary antibodies used for fluorescent immunohistochemistry. * From an aliquot of a preabsorbed antibody. Key – DSHB: Developmental Studies Hybridoma Bank.

Target Protein	Antibody	Host Organism	Dilution	Reference
Primary Antibody:				
Shep	Polyclonal	Rabbit	1:200	Matzat <i>et al.</i> (2012)
Shep	Polyclonal	Guinee pig	1:400	Matzat <i>et al.</i> (2012)
Shep	Polyclonal	Rabbit	1:200	Chen <i>et al.</i> (2019)
Oskar	Polyclonal	Rabbit	1:15,000*	Vanzo and Ephrussi (2002)
Gurken	Monoclonal	Mouse	1:40	DSHB (1D12, Supernatant)
Staufen	Polyclonal	Rabbit	1:500	Johnston <i>et al.</i> 1991
Orb1	Monoclonal	Mouse	1:300	DSHB (4H8, Concentrate)
Lamin C	Monoclonal	Mouse	1:125	DSHB (LC28.26, Supernatant)
Fluorophore-Conjugated Secondary Antibody:				
Anti-Rabbit Alexa Fluor 488	Polyclonal	Goat	1:500	Invitrogen (A11034)
Anti-Mouse Alexa Fluor 488	Polyclonal	Goat	1:500	Invitrogen (A11029)
Anti-Rabbit DyLight 594	Polyclonal	Goat	1:500	Invitrogen (35560)
Anti-Guinea Pig DyLight 594	Polyclonal	Goat	1:500	Invitrogen (SA510096)
Anti-Digoxigenin DyLight 488	Polyclonal	Goat	1:150	Vector Laboratories (DI7488)

Anti-Digoxigenin DyLight 594	Polyclonal	Goat	1:150	Vector Laboratories (DI7594)
---------------------------------	------------	------	-------	---------------------------------

A modified version of the Immunofluorescence staining protocol was used to visualize the expression pattern of Gurken protein. The modified protocol differs from the standard protocol by three main aspects. Firstly, ovaries were fixed in a 4% FA fixative solution containing 75% Heptane (Fisher Chemical, H/0160/15) and 0.25% Nonidet P-40 (Sigma-Aldrich, I8896) at room temperature for 20 minutes. Secondly, the modified protocol substitutes Triton-based Washing buffers for Tween-based buffers, which is to be used at a final concentration of 0.2% Tween-20. Nonetheless, ovaries were still permeabilized in a Triton-based 1% PBST. Finally, ovarian tissues were blocked in PBSTB blocking solution (1% BSA in 0.2% PBST) at room temperature for 1 hour only. For primary antibody incubation, samples were incubated overnight at 4°C with 1:40 dilution of mouse anti-Gurken monoclonal antibody (DSHB, 1D12) prepared in PBSTB blocking solution. The rest of the protocol remained unchanged.

RNA Fluorescence *in situ* Hybridization (FISH) Coupled to Immunostaining: To visualize the expression pattern and subcellular localization of the RNA molecules of interest in the female germline, ovaries obtained from well-fed female flies were fixed in 4% Formaldehyde solution for 25 min at room temperature with continuous shaking on a rocking platform. Ovaries were washed twice in PBST (PBS supplemented with 0.1% Triton™ X-100) for 10 minutes each, followed by a wash in 1:1 PBST and Hybridization solution (50% Formamide, 5× SSC, 0.1% Triton™ X-100, 50 µg/mL Heparin, 100 µg/mL Salmon Sperm ssDNA, in PBS) for 5 minutes at room temperature. The samples were then washed with hybridization solution at 68°C for 5 minutes using a Hybridization incubator to rotate the samples upside-down (Techne, Hybridiser HB-1D). The ovarian tissues were pre-hybridized in Hybridization buffer at 68°C for 1 hour. The probe hybridization step was carried out by incubating the ovaries with 460ng of the DIG-labeled *oskar* antisense probe in 200µL of Hybridization solution overnight at 68°C. Following the completion of probe hybridization, the ovaries underwent a series of washes at 68°C, including two washes in hybridization solution for 20 minutes each, a wash with 1:1 PBST in hybridization solution for 5 minutes, and two washes in 0.1% PBST for 20 minutes each. The samples were washed once more in 0.1% PBST for 5 minutes at room temperature. To continue with the

immunostaining procedure, the ovaries were blocked for 2 hours in PBSTB blocking solution (2% BSA in 0.1% PBST). After blocking, the samples were processed for immunostaining by continuing with the standard protocol described in the ‘Immunofluorescence Staining’ section (see above). It is worth noting that *oskar* mRNA was visualized using fluorescently labelled anti-Digoxigenin antibodies (See Table 2.11), which detect the DIG-labelled *oskar* antisense probe.

Mounting & Imaging of *Drosophila* Ovaries: Following the completion of the abovementioned staining procedures, the ovarian samples in Vectashield® medium were transferred onto glass slides (76×26 mm, 1mm thickness; FisherBrand™, 11562203) for mounting. A pair of tungsten needles were used to tease apart intact ovaries to ensure individual ovarioles and egg chambers were separated, spread out in mounting media, and suitable for imaging. A square glass cover slip (22×22 mm, 0.17 mm thickness; FisherBrand™, 12333128) was then gently lowered at an angle onto the tissue using forceps to prevent the formation of air bubbles in the mounting media. Ovarian tissues were examined using a Zeiss LSM 710 inverted confocal microscope equipped with an EC Plan-Neofluar oil-immersion 40× objective (DIC M27; N.A. 1.30). Images were acquired through the ZEISS ZEN Software (Black, Version 16.0.2.306).

2.5.11 Microscopy

Various microscopy techniques – including light, fluorescent, polarized, and confocal microscopy – were utilized for qualitative and quantitative assessments in different aspects of the project. The use of light microscopy aided with routine fly work, dissections, and quantification analysis of ovarioles and egg laying. Moreover, fluorescent microscopy was employed to visualize and examine the overall morphology of ovaries and mature eggs, respectively. Finally, confocal microscopy was utilized to obtain high-resolution images of the subcellular localization patterns of RNAs and proteins of interest in ovarian cells. A summary of the instruments used for different microscopy techniques is provided below. More specific details regarding the objectives, camera, adapter, and other relevant information used in each experiment can be found in the relevant sections of this chapter.

Light Microscope: Leica S4E stereomicroscope with L2 halogen cold fibre-optic light source.

Fluorescent Microscope: Olympus BX50 and BX53 upright fluorescence microscopes with CoolLED pE-300lite and pE-300white illumination systems, respectively.

Confocal Microscope: Zeiss LSM710 inverted confocal microscope with LASOS RMC 7812 Z2 Argon laser and LSM T-PMT detector.

2.5.12 Microscopy Image Construction:

The images used in this study were constructed to maintain a specific orientation of the specimens. Whether viewed in a transverse optical cross-section or bird's eye-view (i.e., sample viewed from the top), the orientation is indicated by an axis compass. On this compass, the letter 'A' represents anterior, 'P' for posterior, 'D' for dorsal, and 'V' for ventral.

2.6 Fruit Fly Database

FlyBase (Gramates *et al.* 2022), an online bioinformatics database for the insect family Drosophilidae, was used in this project as the primary repository of genetic and molecular data specifically for *Drosophila melanogaster*. FlyBase iterations FB2019_06 through FB2023_05 were used throughout the duration of this PhD project, spanning four years from 2019 to 2023.

2.7 Computational Analysis

2.7.1 Identification of Shep Binding Sites on Target RNA Transcripts

A computational-based sequence analysis was used to identify Shep binding sites on target mRNAs of interest. In a large-scale study that examined binding motifs for many RBPs from different species using RNAcompete, Ray *et al.* (2013) identified three 7-mer binding motifs for *Drosophila melanogaster* Shep. These binding motifs are AUAUWWD (M066_0.6; [RNCMPT00068](#)), WAUWUWD (M165_0.6; [RNCMPT00174](#)), and WUAUWWA (M166_0.6; [RNCMPT00175](#)). The binding motifs are written in nucleotide code, which is a system for representing the nucleotide base(s) in DNA or RNA using a single letter abbreviation. An IUPAC nomenclature

guide for the nucleotide code can be found in Appendix 5.6 for reference. Sequences of the untranslated regions and the coding sequence of each mRNA target were retrieved from FlyBase in FASTA format. For each binding motif, the number of predicted binding sites present in each mRNA sequence was counted using the 'Find' feature on SnapGene viewer (version used 5.2.4). The number of Shep binding sites were mapped to the different regions of the mRNA (i.e., 5'UTR, CDS, and 3'UTR), and the data is presented as a bar graph generated by Graphpad.

2.7.2 Identification of Notch Regulatory Elements in the *shep* Gene Locus

Suppressor of Hairless (Su(H)) is a transcriptional regulator that plays a central role in Notch signalling pathway/transduction. It functions by directly binding to the 'GIGRGAR' DNA consensus sequence present in the regulatory region of the genes it regulates (Morel and Schweisguth 2000; Yuan *et al.* 2016). In the absence of active Notch signalling, Su(H) acts as a transcriptional repressor when not associated with the active form of Notch intra-cellular domain (NICD). On the contrary, when associated with active NICD it acts as a transcriptional activator to activate transcription of Notch target genes.

To determine whether Notch signalling could be involved in the regulation of *Shep* expression, a sequence-based analysis was performed to identify the presence of Su(H) regulatory elements in the *shep* gene locus, specifically upstream of the transcription start sites. The binding motif (GIGRGAR) is written in nucleotide code and the IUPAC nomenclature guide for the code can be found in Appendix 5.6 for reference. The 'Extended Gene Region' sequence of *shep* was retrieved from FlyBase in FASTA format. More specifically, this region encompasses the gene region plus 1Kb upstream and downstream of the gene (FBgn0052423; Chr.3L, complement, genomic location - 5153821..5279944). The number of binding sites present in the gene locus was determined using the 'Find' feature on SnapGene viewer (version used 5.2.4). The data is presented as a graphical diagram showing the number of Su(H) binding sites mapped to the gene locus, with annotations of the transcription start sites.

2.7.3 Identification of a Putative Nuclear Localization Signal in Shep Protein

In eukaryotes, nuclear localization signal (NLS) is a short stretch of amino acid sequence that functions as an essential signal for transport into the nucleus through the nuclear pore complex. The NLS sequence is recognised by members of the importin superfamily, which are classified as nuclear transport receptors. The interaction between these proteins and the NLS of cargo proteins is key in mediating import of proteins into the nucleus.

The amino acid sequences of different Shep isoforms were retrieved from FlyBase in FASTA format, including: Shep-A (FBpp0076875), Shep-B (FBpp0076876), Shep-D (FBpp0076877), Shep-E (FBpp0290814), Shep-F (FBpp0301014), Shep-G (FBpp0305482), Shep-H (FBpp0305483), Shep-I (FBpp0305484). The *shep* gene produces 8 transcripts with 6 unique proteins. Isoforms D and B, as well as E and G are identical in amino acid residues. Differences at the transcript level suggest differential gene expression regulation for each transcript. Two computational-based sequence analysis were applied to identify a putative NLS signal for Shep nuclear transportation. In the first approach, amino acid sequences of Shep isoforms were analysed for the presence of either a functional classical or a non-canonical NLS as previously described in the literature (Romanelli and Morandi 2002; Kosugi *et al.* 2009; Freitas and Cunha 2013; J. Lu *et al.* 2021), using the 'Find' feature on SnapGene viewer (version used 5.2.4).

In the second approach, several prediction tools were utilized to identify a functional NLS signal, including: Nucleolar Localization Sequences Detector (NoD, Scott *et al.* 2010; Scott *et al.* 2011), NLStradamus (Nguyen Ba *et al.* 2009), NucPred tool (Brameier *et al.* 2007), and PSORT II (Nakai and Horton 1999). These prediction tools employ various algorithms and patterns to predict potential NLS sequences based on amino acid composition, charge distribution, secondary structure prediction, previously characterized sequences, and machine-learning technique. Amino acid sequences of the different Shep isoforms were inputted into the tools and analysed first using the default parameters. Additionally, less stringent parameters were also used with tools employing the Hidden Markov Model to aid with NLS identification. In these

instances, the posterior threshold value of 0.2 was used instead on the recommended prediction cut-off value of 0.5.

2.7.4 Molecular Weight Estimation of Electrophoresed Proteins

To estimate the apparent size of proteins in western blot analysis (i.e., molecular weight, MW), a standard curve was generated using known molecular weight markers which were run alongside the protein samples on an SDS-PAGE gel. After processing the nitrocellulose membranes by immunoblotting, images of the membranes were analysed using ImageJ (version 1.8.0, NIH). The ‘Line Selection’ tool was used to measure the migration distance of each protein standard and the dye front from the top of the resolving gel. The relative migration distance (Rf) for each protein standard was calculated by dividing the migration distance of the protein standard by the migration distance of the dye front. In Microsoft Excel, a linear standard curve was then generated by plotting the log(MW) against the Rf of each protein standards. The ‘Slope-Intercept Form’ equation ($y=mx+b$) with an R-coefficient of at least 0.9 was extrapolated from the curve. This equation was then used to estimate the apparent molecular weights of proteins in samples of interest.

2.8 Statistics

A versatile statistics software, GraphPad also known as Prism (Version 9.5.1 528), was used to generate graphs and for the statistical analysis of different experimental data sets. Before performing statistical analysis on GraphPad, the distribution pattern of a data set was assessed for normality using the Shapiro–Wilk test. The normality test was performed using an online calculator provided by Statistics Kingdom (accessed via <https://www.statskingdom.com/shapiro-wilk-test-calculator.html>). Different t-test analyses were performed depending on whether the data followed a normal distribution. The specific statistical test used for each experiment to compare the significance between the means of experimental groups can be found in the relevant sections of this chapter.

In graphs within the results section, asterisks (*) or ‘n.s.’ (not significant) were used to indicate the statistical significance of a p-value. The abbreviation ‘n.s.’ is used when the p-value is higher than 0.05 ($p > 0.05$) to indicate that the difference between groups is

not considered statistically significant. Conversely, a statistically significant difference between the means of experimental groups is typically defined by a p-value less than or equal to 0.05 ($p \leq 0.05$), which indicates a low probability of the observed difference occurring by chance. The number of designated asterisks corresponds to the increasing degree of statistical significance between the groups being compared: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), and **** ($p < 0.0001$). Moreover, the data in graphs are presented as the mean values \pm SD (standard deviation) or SEM (standard error of mean), with the number of experimental replicas denoted by an uppercase 'N' and the number of observations within each experiment indicated by a lowercase 'n'.

2.9 Experimental Replicas

To ensure the reproducibility and representativeness of the data presented here, a minimum of two/three independent biological replicates were obtained for each experimental observation.

⌘ Chapter III ⌘

Shep Expression Pattern During
D. melanogaster Oogenesis

3 Chapter III – Shep Expression Pattern During *D. melanogaster* Oogenesis:

In the mid-2000s, *Drosophila shep* was identified in a forward genetic screen as a gene involved in geotaxis (Armstrong *et al.* 2006). Since then, research has primarily focused on investigating its role in the nervous system. However, to date, our understanding of Shep's expression patterns, regulation, and function within the ovary remains largely unexplored. We employed a combination of fly genetics, molecular biology techniques, biochemical assays, and computational analysis to address its potential involvement in *Drosophila* oogenesis.

Aims
1) Characterize Shep expression & subcellular distribution pattern within ovarian tissues.
2) Analyse Shep expression at the mRNA and protein levels in the ovary.
3) Understand the regulatory mechanisms driving Shep expression in the ovary.

3.1 Preliminary Data

Since PTB was identified as a regulatory component of the *oskar* mRNP complex (Besse and López de Quinto *et al.* 2009), our research group has focused on characterizing and investigating the multifunctionality of PTB as an RNA-binding protein in fruit fly ovaries, especially its contribution to *oskar* mRNA regulation during oogenesis. To achieve this, a yeast two-hybrid screen was performed by a company called Hybrigenics (France), using full-length PTB as a bait against an ovarian prey library to identify potential PTB-interacting partners in *D. melanogaster* ovaries. The screen analyzed a total of 62.3 million interactions and identified 19 unique candidates from 73 independent positive clones (López de Quinto, unpublished). Among the identified proteins, the majority belonged to the RNA-binding protein class (15%, N° = 11), followed by members of the DNA-binding protein class.

Notably, PTB and Yps were identified as potential PTB-interacting candidates. Because PTB can exist as a homodimer (Pérez *et al.* 1997; Gong *et al.* 2021), it was reassuring that the screen was able to detect this dimerization event, adding credibility to the screen's findings. In addition, Yps binds to *oskar* mRNA, regulating its localization and translation during *Drosophila* oogenesis (Mansfield *et al.* 2002). It is interesting to find that PTB interacts with other RBPs within the *oskar* mRNP, which highlights the complexity of interactions within the *oskar* mRNP. Given all this, we decided to examine and compare the expression and localization patterns of these candidates with the PTB pattern in the ovary.

3.2 Shep as a Potential Interacting Partner of PTB

To investigate the localization pattern of the candidate proteins identified through the yeast two-hybrid screen and their relationship to the PTB pattern during oogenesis, we examined their subcellular distribution using protein trap lines. We paid particular attention to patterns that might implicate candidates' involvement in *oskar* mRNP regulation. In fact, protein traps have previously been used to screen for and identify novel proteins involved in *oskar* mRNA post-transcriptional regulation (Besse and López de Quinto *et al.* 2009). This initial screening was possible for a subset of PTB-interacting partners that had available protein-trap lines (N^o = 10). From this screening, the subcellular distribution pattern of the RNA-binding protein Shep was of particular interest.

During mid-oogenesis, Shep exhibits a cortical subcellular localization within the oocyte, with pronounced enrichment at the posterior pole of the oocyte (**Figure 3.1C**). Shep localization at the oocyte's posterior cortex appears to coincide with that of the *oskar* mRNP complex (**Figure 3.1D**), a pattern that is referred to in the field as a posterior crescent (Kim-Ha *et al.* 1991; Vanzo and Ephrussi 2002; Bose *et al.* 2022). This localization pattern is a characteristic feature associated with proteins that are key players in the *oskar* mRNP post-transcriptional regulation, such as PTB and Yps. Unlike Shep, which also showed enrichment across the oocyte cortex, PTB and Yps predominantly localized to the posterior cortex of the oocyte, as demonstrated in panels A-C of **Figure 3.1**. Among other proteins, Hrp48, Staufén, Kinesin-1, and Par-1 display

a similar pattern, as shown in **Figure 3.1**, and are known to be key components of the *oskar* RNP regulatory complex (Brendza *et al.* 2000; Shulman *et al.* 2000; Huynh *et al.* 2004; Yano *et al.* 2004). This distinctive pattern suggests that Shep could be a novel component of the *oskar* regulatory mRNP complex. In addition, the enrichment of Shep protein at the oocyte cortex also imply that Shep could be part of other mRNP complexes.

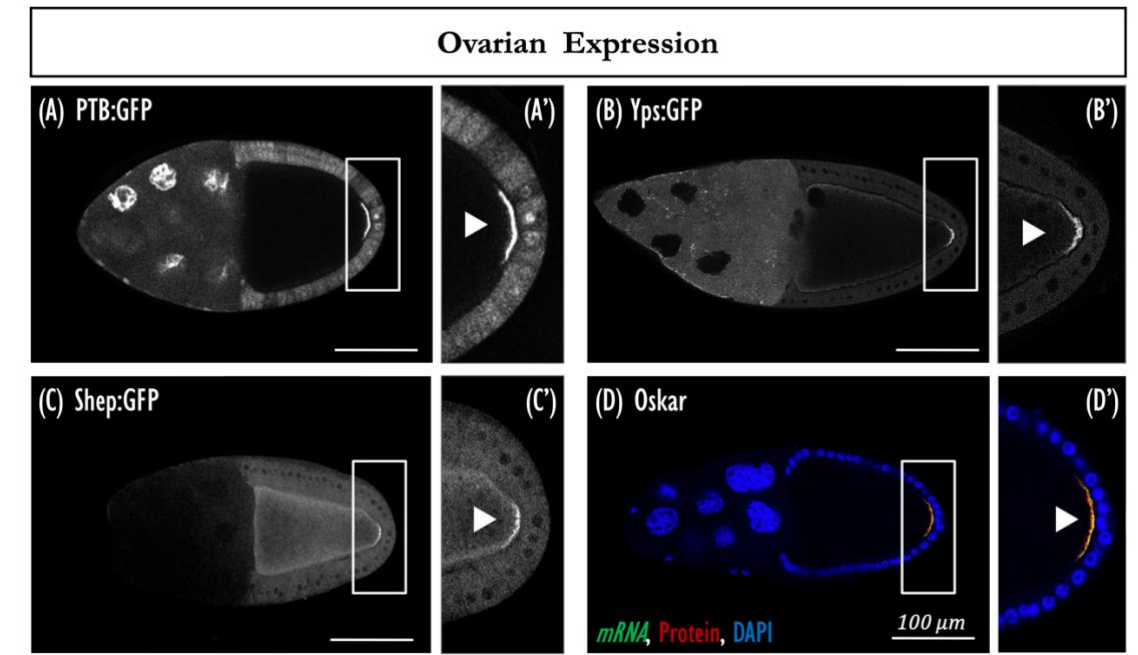


Figure 3.1: Shep enrichment at the posterior pole of the oocyte. Subcellular distribution of GFP-trapped PTB (A), Yps (B), and Shep (C) in stage-10 egg chamber. The expression of PTB, Yps, and Shep protein-trap from their respective genes is driven and under the control of endogenous regulatory elements (e.g., promoter, enhancers, etc.). (D) Fluorescence in situ hybridization combined with immunostaining of wild-type stage 10 egg chamber showing the distribution of *oskar* mRNA (in green) and Oskar protein (in red). (D–D’) Cell nuclei stained with DAPI and visualized in blue. Colocalization of both *oskar* mRNA and protein signals appear yellow. Similar to *oskar* mRNA, Oskar protein, and other proteins of the *oskar* mRNP complex, Shep displays a posterior enrichment, as indicated by the white arrowhead. Regions outlined by the white box indicate the magnified posterior portion of the oocyte shown in panels a’–d’. Scale bar, 100 μ m. Fly genotype(s) – Panel A (*w¹¹¹⁸*; PTB:GFP^{dsRed}/TM6B,Tb,Hu), B (*y,w⁻*; Yps:GFP/TM3,Sb), C (*w⁻*; Shep:GFP^{CC00236}/Shep:GFP^{CC00236}), and D (*w¹¹¹⁸*).

3.3 Ovaries of Shep GFP Protein-Trap Flies are Small

During the preparation of female flies from the Shep protein-trap line (CC00236), we made an interesting observation: these females appear to exhibit smaller abdomens and ovaries than control females.

The CC00236 protein-trap line is a P-element insertion of a GFP-containing mobile cassette into the genomic locus of *shep* (see green triangle in **Figure 3.2A**). Given the specific location of the GFP-trap cassette insertion at the beginning of the 5'UTR of the small *Shep* isoforms (namely *Shep* -H, -I, and -E), it is possible that these isoforms will not be trapped with GFP. Additionally, this raises the possibility that the insertion might interfere with the transcription or translation of these isoforms, potentially leading to a loss of function. Consequently, this protein line has the potential to serve as a null mutant for the small isoforms of *Shep*. This hypothesis will be investigated later in this chapter (see Section 3.4.1).

To further characterize our initial observations, we decided to systematically compare the size of *Shep*-GFP ovaries with that of the control flies. Ovaries of mated adult female flies of similar ages from the *Shep* protein-trap line were carefully prepared for dissections as described in the 'Materials & Methods' Chapter. For the control, we used white-mutant females, which were specifically chosen for this analysis because they represent the genetic background into which the GFP-trap cassette was inserted. The experimental setup was standardized to ensure consistency and to enable a reliable comparison between the *Shep* protein-trap line and the control group; variables such as the age of females post-eclosion, the number of flies per vial, and the ambient temperature were carefully controlled. *Shep* protein-trap females exhibited noticeably smaller abdominal stature than their control counterparts (data not shown). After dissection, we observed a significant reduction in the size of the *Shep* protein-trap ovaries compared with the control group, with a decrease of approximately 20% (Student's t-tests, $P < 0.001$). This notable reduction in ovary size between both groups is demonstrated in panels B and C of **Figure 3.2**.

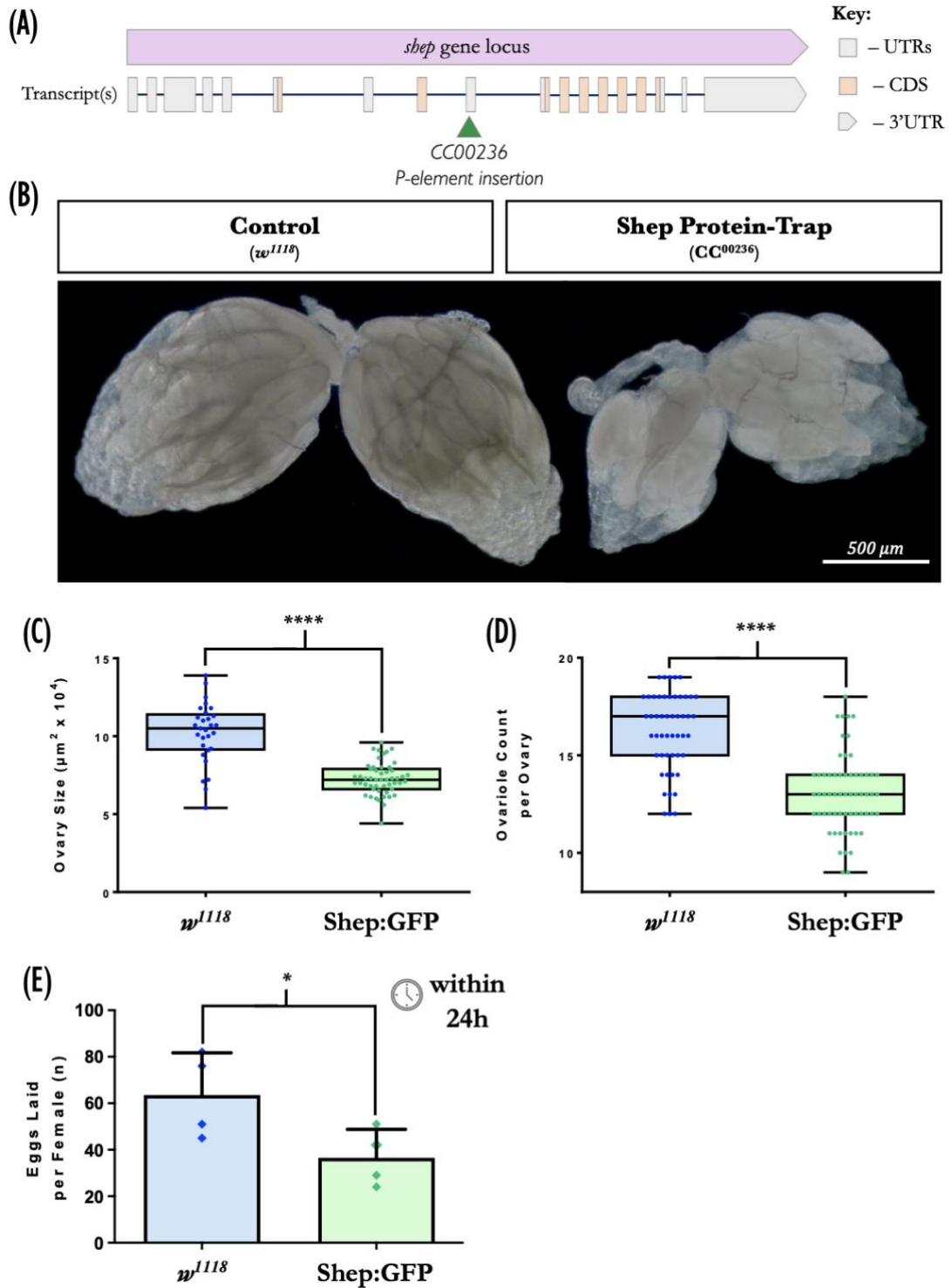


Figure 3.2: Genetic manipulation of *shep* affects female reproductive organs. (A) Schematic representation of *shep* gene (purple) organization with depictions of its corresponding coding (orange) and noncoding exons (gray) for all *shep* splice variants, namely transcripts A, B, D, E, F, G, H, and I. The gray bar with the pointed end represents the 3'UTR end of the transcripts. The green triangle denotes the insertion site of the p-element GFP cassette within the CC00236 protein-trap line. The diagram is not drawn to scale or orientation. (B) Representative images of ovaries from adult *D. melanogaster* control (*w¹¹¹⁸*) and homozygous *shep* mutant (CC00236 protein-trap line) females. Scale bar, 500 μ m. (C–D) Graphs showing the reduction in ovary size (in $\mu\text{m}^2 \times 10^4$; total n = 15–25 pairs of ovaries per experimental group), number of ovarioles per a single

ovary (n = 25–30 females per experimental group in total), and average number of eggs laid per day per female (N = 4 independent experimental replicas; total n = 12 females per experimental group) between the indicated genotypes. The whiskers of box plots represent the minimum and maximum values of the dataset. Error bars indicate mean \pm SD. Statistical significance was analysed using Student's t-test and Mann–Whitney test to compare the differences in means between experimental groups. Student's t-test was performed to obtain a p-value for both ovary size and number of eggs laid datasets, while the Mann–Whitney test was performed for the ovariole count dataset. The significance between experimental groups is indicated by the number of asterisks (*) above the plots: * (P<0.05) and **** (P<0.001). Fly genotype(s) – control female flies (*w¹¹¹⁸*;;) and Shep GFP protein-trap line (*w*;; Shep:GFP^{CC00236}/Shep:GFP^{CC00236}).

Because ovary size has been shown to correlate with the number of ovarioles and rate of egg production (Boulétreau-Merle *et al.* 1982), we wanted to address the underlying factors contributing to the observed reduction in ovarian area. Specifically, we sought to determine whether this reduction could be attributed to a decrease in the number of productive ovarioles per ovary, a decrease in the number of eggs laid per female, or a combination of both factors.

To count the number of ovarioles per ovary, female flies were prepared and dissected following a procedure similar to the assessment of ovary size. However, in this case, the ovaries were carefully torn apart using a tungsten needle to facilitate the quantification of ovarioles. Ovariole count varies depending on the species or strain of *Drosophila*, but a *D. melanogaster* ovary typically contains an average of 16-23 of these egg-producing functional units called ovarioles (Sarikaya *et al.* 2012). As anticipated, the number of ovarioles in the ovaries from the protein-trap females was significantly lower than that in the control group, with an average of approximately 13 ovarioles per ovary (see **Figure 3.2D**; Mann–Whitney test, P<0.001). This decrease in the number of ovarioles is in agreement with previous literature (Sarikaya *et al.* 2012; Sarikaya and Extavour 2015; Zhao *et al.* 2022), suggesting a positive correlation between egg size and the number of ovarioles in certain insects.

To examine the impact of the reduced number of ovarioles on the capacity of Shep protein-trap females to lay eggs, we performed oviposition assays. Female flies were prepared following a procedure similar to that used in previous ovarian assessments. However, instead of dissection, the flies were transferred to an egg collection apparatus, and females were allowed to lay eggs for a period of 24-hours (see Section 2.5.9 of the 'Materials & Methods' Chapter for the exact conditions used). We quantified the total

number of eggs per vial and then calculated the average egg count per female within the 24-hour timeframe by dividing the total number of eggs by the number of females in each vial. Upon closer examination of the oviposition plates, we observed a significant discrepancy in the daily egg deposition between both experimental groups (Student's t-tests, $P < 0.05$). *Shep* protein-trap females laid significantly fewer eggs than the control, with an approximate decrease of 43% (**Figure 3.2E**). This finding aligns with our previous observations of reduced ovary size and decreased number of ovarioles.

Taken together, these results demonstrate that *Shep* can influence the development and capacity of adult female reproductive organs. Furthermore, these findings strengthen our hypothesis that *Shep* could play a pivotal role in oogenesis, making it a compelling candidate for further investigations.

3.4 Expression Pattern of *Shep* in *D. melanogaster* Ovaries

Given the limited knowledge about the expression and localization pattern of *Shep* in the ovaries, we decided to conduct a comprehensive characterization of *shep* gene expression in this tissue. Our analysis encompassed: (i) *Shep* protein expression, (ii) *shep* transcript expression, and (iii) *Shep* protein distribution pattern. Our aim was to elucidate the spatial and temporal expression patterns of *Shep*, thereby improving our understanding of its role in ovarian biology.

3.4.1 Characterization of the *Shep* Isoforms Expressed in *Drosophila* Ovarian Tissues

In *Drosophila*, the *shep* gene can produce 8 transcripts (A, B, D, E, F, G, H, and I) as annotated in FlyBase (Gramates *et al.* 2022). However, the coding sequences of transcripts B and D, as well as E and G, are identical, resulting in only six unique translated isoforms. Variations in the untranslated regions of these transcripts, despite encoding the same amino acid sequences, suggest regulation of gene expression at the mRNA level. These isoforms can be categorized into two groups on the basis of their predicted molecular weights: the ‘large/long’ and ‘small/short’ isoforms. The large isoforms, including *Shep* -A, -B, -D, and -F, have molecular weights ranging from 53 to 62 kDa. Conversely, the small isoforms, including *Shep* -E, -G, -H, and -I, have molecular weights of approximately 40 kDa.

To identify the isoforms of Shep expressed in *Drosophila* ovarian tissues, we performed immunoblotting analysis using a crude ovarian extract from adult ovaries of both the wild-type control (w^{118}) and homozygous Shep protein-trap (CC⁰⁰²³⁶). For each genotype, a crude ovarian extract equivalent to one pair of ovaries was electrophoresed on a 9% SDS-PAGE gel. Western-blot analysis using anti-Shep antibodies raised against the *Drosophila* protein (a generous gift from Elissa Lei, Matzat *et al.* 2012), detected the endogenously expressed large and small isoforms in the wild-type ovarian extract. As depicted in **Figure 3.3**, the large isoforms observed in the control ovarian extracts at around 70 kDa shifted to approximately 100 kDa in the protein-trap line because of the GFP insertion upstream of the RRM1 domain. Notably, the small isoforms seemed to be absent in the Shep protein-trap extract as we failed to detect either the un-tagged isoforms or their GFP-tagged polypeptides at an expected size of approx. 65 kDa (**Figure 3.3**). It is worth noting that the apparent molecular weights of these proteins were slightly higher than their predicted molecular weights, possibly because of post-translation modifications. These results not only shed light on the potential isoforms expressed within the fruit fly ovary but also support our initial assumption that the chromosomal position of the P-element insertion could hinder the expression of the smaller isoforms.

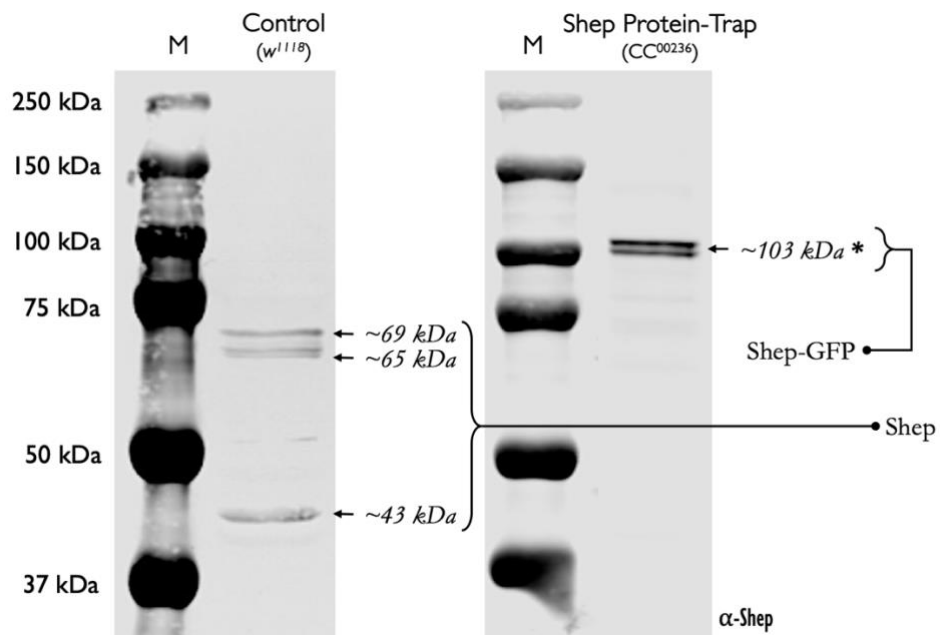


Figure 3.3: Different Shep protein isoforms are expressed in the *Drosophila* ovary. Ovarian crude extracts, equivalent to one pair of ovaries, from w^{118} and homozygous Shep protein-trap (CC00236) females were electrophoresed on a 9% SDS-PAGE gel. Shep was detected using

Western blotting by probing with rabbit anti-Shep antibodies. Protein markers were run in the lane labelled with 'M' (Precision Plus Protein Dual Color Standards, BioRad). In the Shep protein-trap, the insertion of GFP upstream of the RRM1 domain results in an upward shift in the apparent molecular weight of the larger Shep isoforms to approximately 100 kDa, as indicated by an asterisk (*). Fly genotype(s) – wild-type control ($w^{1118};$) and Shep protein-trap line ($w; ;$ Shep:GFP^{CC00236}/Shep:GFP^{CC00236}).

3.4.2 *Drosophila* Ovaries Express Multiple *shep* Transcripts

To confirm the identity of the Shep isoforms detected in our Western blot analysis, we aimed to identify the transcripts expressed in wild-type adult ovaries by performing a series of reverse transcription polymerase chain reactions (RT-PCR). From w^{1118} ovaries, total ovarian RNA was first purified and treated with DNaseI before being converted into cDNA. In cases where a particular transcript was not expressed in the ovaries, then we used total RNA obtained from different *Drosophila* tissues as positive controls.

The primer design was a critical aspect of this analysis because primers should be transcript-specific to selectively amplify and fish out particular transcripts. As mentioned previously, in *Drosophila*, the *shep* gene produces eight transcripts according to the FlyBase annotation: A, B, D, E, F, G, H, and I. These transcripts share several common exons, although some exons are unique to either a specific transcript or a subset of transcripts (**Figure 3.4A**). For instance, the 5'UTR regions of transcripts A and F are unique to these two transcripts. Transcript F, however, possesses a transcript-specific region within its coding sequence, distinguishing it from the others. Similarly, transcripts B, D, and G also have unique 5'UTR sequences that are transcript-specific, thereby simplifying primer design. Moreover, while the 3'UTR varies in length across different transcripts, its initial region is shared across all transcripts. Furthermore, using an endpoint RT-PCR identification method to assess the expression of all transcripts – apart from transcript I – was relatively straightforward because of their unique, transcript-specific exons. However, distinguishing transcript I from transcripts E and H, all of which encode small Shep isoforms, proved to be challenging. To overcome this, we employed an approach that included a series of endpoint PCR amplifications and DNA gel purifications, as detailed in Section 2.1.3.3 of Chapter 2. This enabled us to dilute the cDNA template library, specifically enriching transcripts H and I. We then

conducted a nested RT-PCR to ensure precise amplification and identification of transcript I.

Following this, our results confirmed the expression of transcripts A, B, F, E, H, and I within ovarian tissues, as demonstrated by the PCR amplifications shown in panel B of **Figure 3.4**. Interestingly, transcripts D and G were absent in the ovary, especially compared with their expression in *Drosophila* testes. In the second round of our nested RT-PCR analysis, when examining the expression of transcript I, we included a control reaction to test for single-stranded DNA (ssDNA) amplification. This control, consisting of all PCR components except for the forward primer, indicated a relatively weak amplification of ssDNA alongside that of transcript I (**Figure 3.4C**).

Altogether, these results provide further insights into the identities of the Shep proteins detected in our western blot analysis (**Figure 3.3**), with the large isoforms being isoforms A, B, and F, and the low molecular weight proteins corresponding to isoforms E, H, and I. The data also hint at a tissue-specific regulation of *shep* transcripts expression, suggesting that different Shep protein isoforms may play different roles in various tissue types. This nuanced regulation of *shep* expression could potentially contribute to the broad functionality of this protein across different biological contexts and stages of development, adding layers of complexity to our understanding of the functions of Shep protein in *Drosophila melanogaster*.

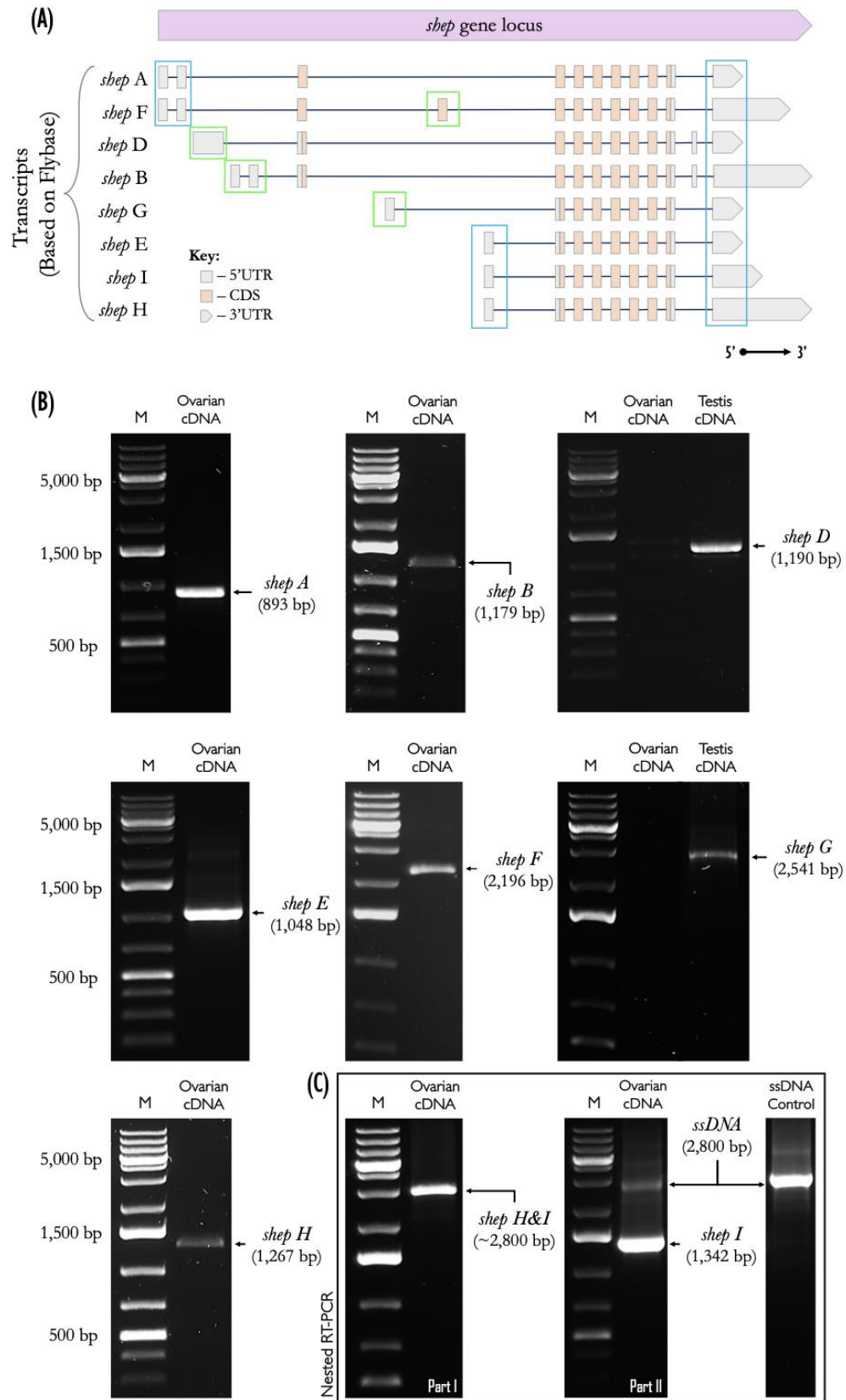


Figure 3.4: Analysis of the *shep* transcripts expressed in *Drosophila* ovaries. (A) Gene model of *shep* showing the eight transcripts (A, B, D, E, F, G, H, and I) produced during gene transcription, based on FlyBase annotations. The diagram, which is not drawn to scale or orientation, uses coloured shapes to represent different regions of an mRNA: light orange for the

coding sequence and gray for untranslated regions. Regions common across all transcripts or unique to a subset of transcripts are highlighted in blue, whereas green indicates transcript-specific regions. **(B)** Endpoint RT-PCR analysis of *shep* gene expression. Gene expression analysis was performed using qualitative endpoint RT-PCR with a cDNA library generated from the total RNA of *w¹¹¹⁸* ovaries or testis. cDNA was amplified using 30 cycles of PCR and visualized on a 1% agarose gel. The qualitative and semi-quantitative endpoint RT-PCR, using transcript-specific primer pairs, generated amplicons of varying lengths: transcript A (893 bp), transcript B (1,179 bp), transcript D (1,190 bp), transcript E (1,048 bp), transcript F (2,196 bp), transcript G (2,541 bp), transcript H (1,267 bp), and transcript I (1,342 bp). **(C)** Nested RT-PCR was used for the identification of transcript I, which comprised two parts (I and II). In the first part, both transcripts H and I were amplified from the pool of cDNA templates in the library. This was followed by a series of PCR amplifications and DNA gel purifications before the second part of nested RT-PCR. To test for the amplification of ssDNA in the second part of the nested RT-PCR reaction (Part II), a control reaction was included, which consisted of all PCR components, excluding the forward primer. Abbreviations – ssDNA: single-stranded DNA. Fly genotype – wild-type control (*w¹¹¹⁸;+*).

3.4.3 Characterization of the Subcellular Distribution Pattern of Shep Protein in Ovaries

To thoroughly investigate the cellular expression and subcellular localization patterns of Shep protein within the ovarian tissue, we employed a multifaceted approach including protein-trap lines, fluorescently tagged fusion proteins, and immunocytochemistry of the endogenous protein. This strategy provides valuable insights into its spatiotemporal expression and potentially elucidates its biological functional roles.

Initially, we achieved this by fixing ovaries from the Shep protein-trap line to examine the GFP fluorescence from the trapped Shep isoforms, reflecting their endogenous protein expression. In general, our findings showed that the Shep protein is expressed in both germline and somatic cell types of the ovary (**Figure 3.5B**). Contrary to its expression in the nervous system, where it displayed both nuclear and cytoplasmic localization (Matzat *et al.* 2012; Chen *et al.* 2018), we found Shep in the ovary to be exclusively localized to the cytoplasm. This differential expression pattern suggests that Shep may play different biological roles and be subject to distinct regulatory mechanisms in the ovaries compared with the nervous system.

In the germline, Shep protein was expressed as early as in the germarium, specifically in the cytoplasm of the germline stem cells and the dividing cells of the cyst (see panel Bi of **Figure 3.5**). Here, the Shep protein was evenly distributed among the sibling cells

of the cyst until the oocyte is specified (**Figure 3.5B**). In region 3 of the germarium, Shep begins to specifically accumulate within the oocyte and continues to get enriched within the oocyte throughout oogenesis. This early and dynamic expression of Shep in the germline suggests its potential involvement in germline stem cell maintenance, as well as oocyte development and maturation.

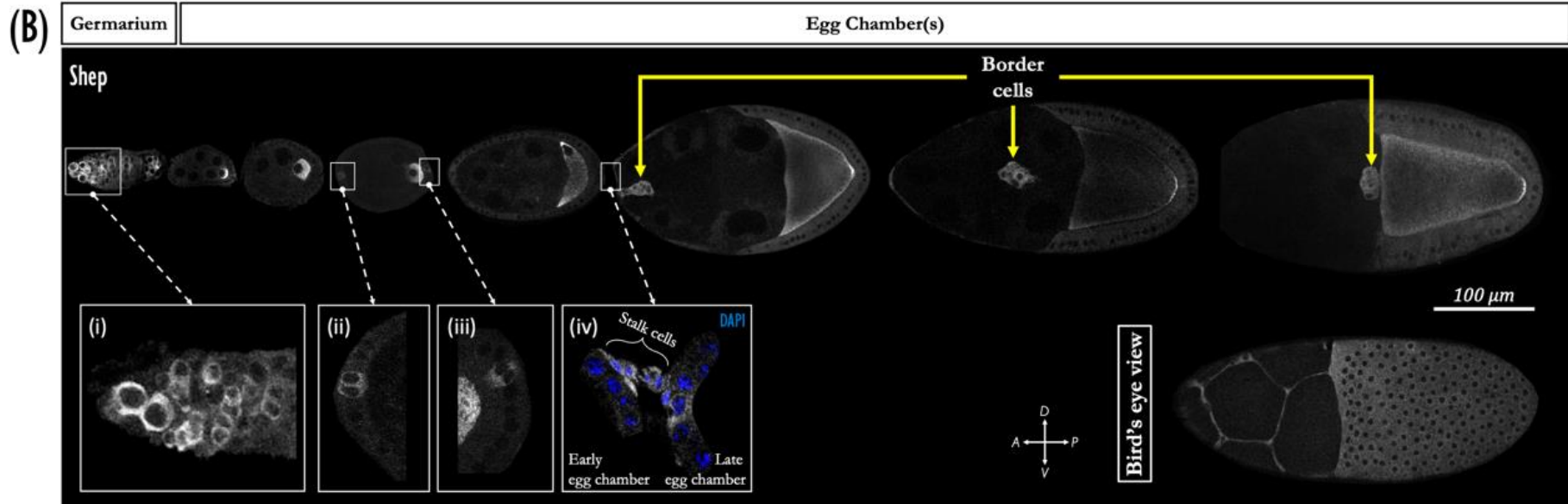
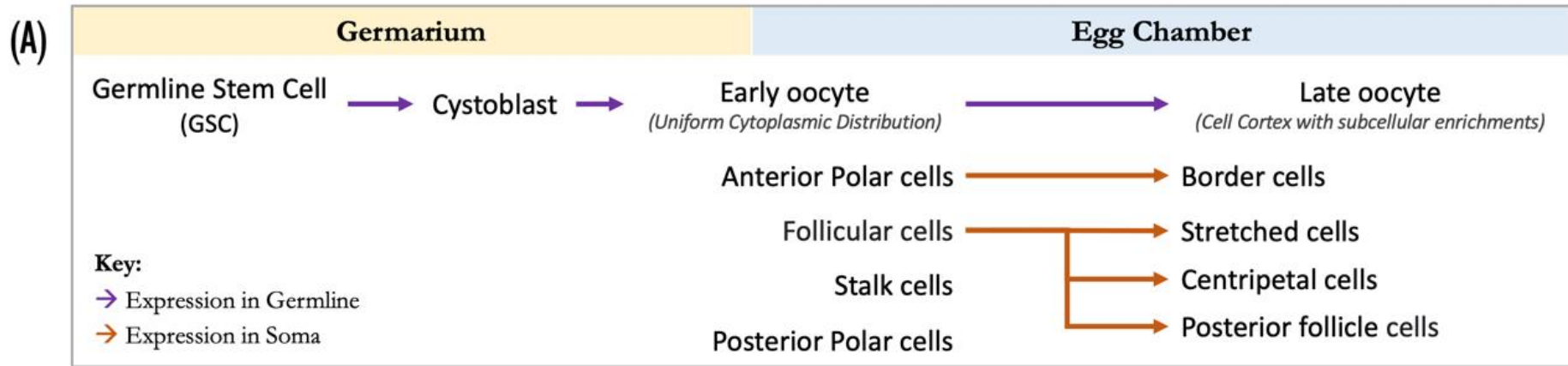
Similarly to other maternally loaded mRNAs and proteins, Shep transiently localizes to the anterior region of the oocyte upon entering through the ring canals. As the oocyte develops and matures, Shep eventually translocates to the cortex. Specifically, during mid-oogenesis, the Shep protein concentrates subcellularly at both the posterior and anterior poles of the oocyte cortex (**Figure 3.5**). These specific regions are where key maternally loaded mRNAs are localized and are important for embryonic development. For instance, *oskar* localizes to the posterior pole of the oocyte, playing a critical role in specifying the posterior fate of the developing embryo and pole cells (Lehmann and Nüsslein-Volhard 1986; Vanzo and Ephrussi 2002). On the other hand, *gurken* localizes to the dorsal-anterior corner, defining the dorsal fate of the embryo (Cáceres and Nilson 2005). This potential association of Shep with various mRNP complexes suggests a role for Shep in regulating the expression or spatial distribution of these key mRNAs, thereby influencing developmental outcomes. Moreover, if the Shep protein happens to be a regulatory component of both the *oskar* and *gurken* mRNP complexes, its localization at the oocyte cortex suggests that Shep could also be part of other mRNP complexes for mRNAs located there. These may include those involved in signalling pathways and the cortical cytoskeleton, both of which are crucial for intercellular communication and positioning of the oocyte within the egg chamber.

Shep protein expression in the ovarian somatic cells surrounding the germline syncytium starts at a later stage in oogenesis than that in the germline. Notably, Shep is expressed in the cytoplasm of all follicle cell types in the ovary, with the earliest and most pronounced expression been observed in both the anterior and posterior polar cells of a stage 5 egg chamber (see panels Bii and Biii of **Figure 3.5**). This pronounced expression of Shep persists in polar cells throughout oogenesis and it remains specifically enriched within the anterior polar cells as they transition into border cells and start their migration towards the oocyte's anterior at around stage 8 (**Figure 3.5B**).

Despite the gene expression profile changes that border cells undergo during their specification and migration, Shep expression consistently remains in the cytoplasm of of the border cell cluster. Shep was also expressed in the cells of the anterior structure known as the stalk (**Figure 3.5Biv**), which serves as a physical connexion between adjacent egg chambers. Furthermore, Shep protein expression in the follicular epithelium starts at a low level around stages 4-5 and gradually intensifies as the egg chamber progressively matures. At later stages of oogenesis, Shep expression intensified in all subpopulations derived from the precursor main body follicle epithelium encapsulating the germline, namely stretched cells, centripetal cells, and posterior follicular epithelial cells.

<<< See Figure Next Page >>>

Figure 3.5: Distribution of Shep protein during *Drosophila* oogenesis. (A) Overview of Shep protein expression in distinct cell populations of the ovary. The progression of Shep protein expression in the germline and somatic lineages throughout *Drosophila* oogenesis is indicated by purple and orange arrows, respectively. (B) The expression pattern of Shep protein in an ovariole, from the germarium up to a stage 10 egg chamber. A complete ovariole of the Shep expression pattern was constructed using homozygous Shep-GFP protein-trap females, except for panel 'i' that was obtained after immunostaining with anti-Shep antibody. A *Drosophila* ovariole consists of two morphologically distinct regions (refer to **Figure 1.8** for reference): the germarium, which houses stem cells and the germline cyst, and a chain of egg chambers at different developmental stages, from stage 1 at the end of the germarium to stage 14 (not shown). In *Drosophila*, the developing oocyte exists as a part of a functional unit known as an egg chamber, with 15 nurse cells supporting its development and a monolayer of follicle cells around the germline syncytium. A monolayer of follicular epithelium surrounds the entire germline syncytium for the majority of oogenesis. At stage 8, the follicle cells begin differentiation and migration to encapsulate the oocyte only. The oocyte, whose size dramatically increases as oogenesis proceeds, is the most posterior cell of the germline cyst. Panels i–iv represent expansions of the highlighted regions, captured from different optical sections. Scale bar, 100 μm . Fly genotype(s) – wild-type control ($w^{1118};$) and Shep protein-trap CC00236 line ($w;$; Shep:GFP^{CC00236}/Shep:GFP^{CC00236}).



Because the characterization of Shep expression in the ovary was primarily based on homozygous Shep GFP protein-trap females, which selectively trap a subset of isoforms, particularly the large isoforms (see bands around 100 kDa in **Figure 3.3**), it was important to ensure that the expression pattern observed using the Shep CC00236 protein-trap line accurately represents all isoforms expressed in the ovary. Thus, to validate our findings we decided to confirm the endogenous expression patterns using immunostaining and *in vivo* fluorescence tagging. To detect endogenous Shep proteins in *w¹¹¹⁸* ovaries, we used two different polyclonal antibodies generously provided by Elissa Lei (Matzat *et al.* 2012) and Kathleen Beckingham (Chen *et al.* 2014). Our immunostaining experiments detected the expression of Shep protein during the early stages of ovarian development (**Figure 3.5Bi**). However, during later stages, the antibodies could not penetrate the follicular epithelium of the egg chambers (data not shown). As a result, the detection of Shep expression was compromised in the later stages. It is important to note that, based on the immunostaining results from the early stages, the expression of Shep protein in the ovary is localized to the cytoplasm and not to the nucleus.

Isoform-Specific Subcellular Distribution Pattern of Shep: To visualize Shep proteins without the use of antibodies, we generated three UAS constructs of different Shep protein isoforms, namely A, C and E. These constructs were designed with an mCherry coding sequence fused to the N-terminus of each of the Shep polypeptides, allowing us to visualize the distribution of three Shep isoforms using the fluorescent signal of mCherry. The rationale for selecting these specific isoforms is detailed in Section 4.3 of Chapter 4. In brief, they represent one of each of the large and small isoforms expressed in the ovary, as confirmed by western blotting and RT-PCR analyses (sections 3.4.1 and 3.4.2). Following the generation of these transgenic flies, they were crossed with various tissue-specific Gal4 driver stocks to drive the expression of mCherry-tagged Shep proteins in different ovarian tissues, including the germline syncytium, the main body follicle epithelium, and the border cell cluster (**Figure 3.7**). After the crosses were set up, the ovaries from F1 females were dissected and fixed. Subsequently, the ovarian expression patterns of Shep isoforms were compared to identify any differential localization patterns between the germline and soma.

In the germline, the distribution pattern of different Shep isoforms is comparable to that of the endogenous or GFP-trapped proteins, as previously shown in **Figure 3.5B**. Shep is primarily enriched in the oocyte cytoplasm, with lower expression levels observed in nurse cells. The differences in localization patterns between Shep isoforms became more apparent during later stages of oogenesis, particularly at stages 8 to 11. At these stages, isoforms C and E displayed a clear posterior crescent that was absent in Shep-A (**Figure 3.6**). In contrast to the consistent enrichment of Shep-E at the oocyte posterior pole, isoform C was only occasionally enriched at the posterior cortex of the oocyte (compare panels B-D of **Figure 3.6**). It is noteworthy that this variation in the localization pattern of Shep-C at the posterior varied considerably within the same ovary and between different experimental replicates.

The differential localization of Shep isoforms at these specific stages suggests distinct functional roles and potentially different regulatory mechanisms governing their subcellular localization. Additionally, due to the consistent subcellular localization of Shep-E at the posterior pole of the oocyte, it seems highly likely that it could be the isoform associated with the *oskar* mRNP complex and potentially involved in the post-transcriptional regulation of *oskar* expression.

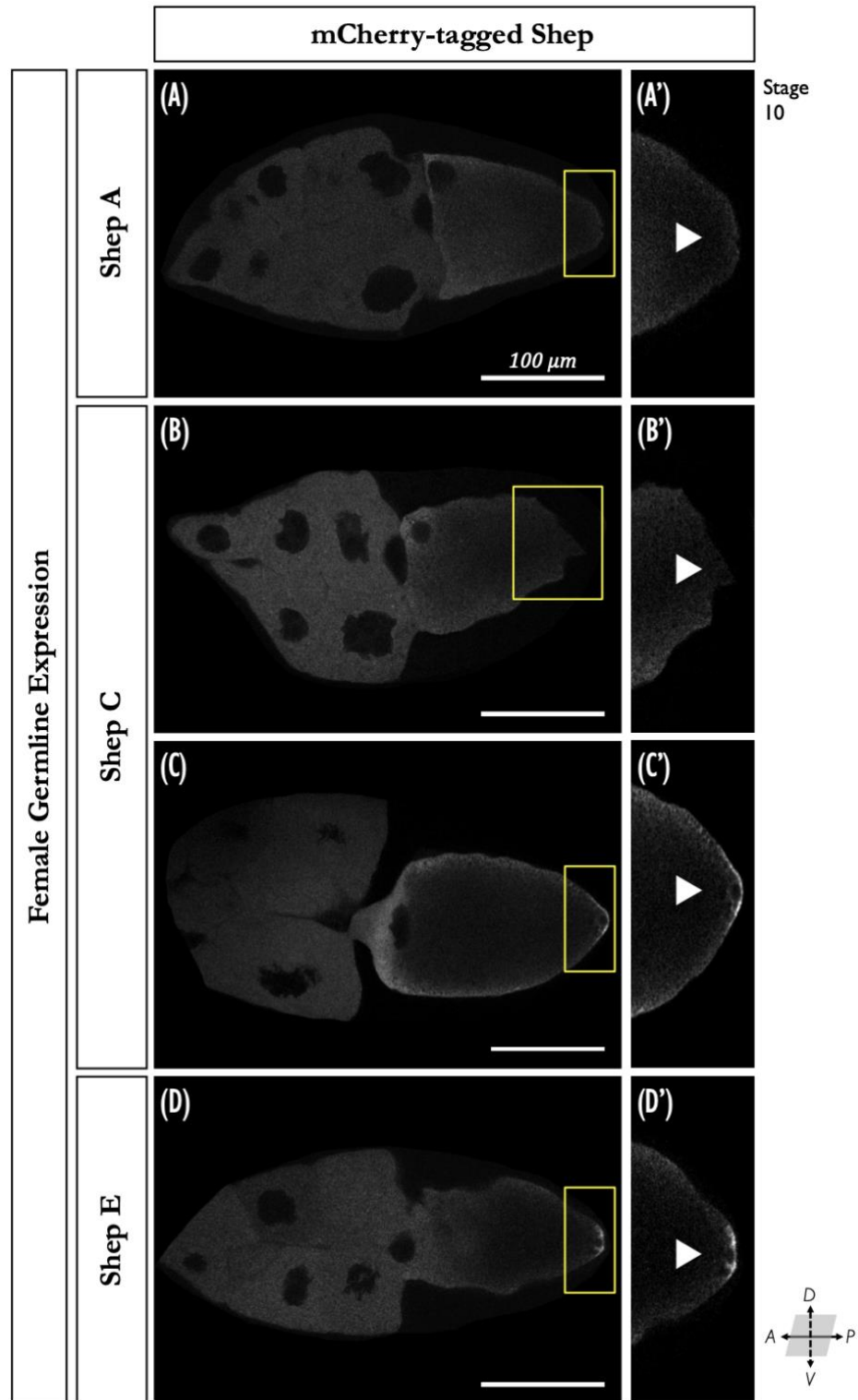


Figure 3.6: Different Shep isoforms display distinct localization patterns in late-stage oocytes. In stage 10 oocytes, the various mCherry-tagged isoforms of Shep showed different enrichment patterns at the posterior crescent. **(A)** While mCh:Shep-A is absent from the posterior pole of the oocyte, **(B–C)** mCh:Shep-C occasionally localizes there. **(D)** In contrast, mCh:Shep-E is consistently present at the posterior pole of the oocyte. Regions outlined by the yellow box indicate the magnified posterior portion of the oocyte shown in panels A'–D'. Scale bar, 100 μm. White arrowheads in the magnified panels (A'–D') point towards the posterior cortex of the oocyte where Shep enrichment is expected. Fly genotype(s) – Panel A (*m⁻; osk-Gal4:VP16/UASp-*

mCh:ShepA;), **B–C** (*w⁻*; *osk*-Gal4:VP16/UASp-mCh:ShepC;), and D (*w⁻*; *osk*-Gal4:VP16/UASp-mCh:ShepE;).

Shep Protein is Not Nuclear in Any of the *Drosophila* Ovarian Cell Types: Strikingly, our results have shown that Shep protein seems to localize exclusively to the cytoplasm of various ovarian cell types, which is in contrast to what has been reported in the central nervous system (Chen *et al.* 2017; Chen *et al.* 2018), where Shep protein is detected in both the cytoplasm and nucleus of neurons and glia.

Next, we investigated whether the nuclear localization of the Shep protein is dependent on the isoform, cell type, concentration, or a combination of these factors. To address this, we employed the UAS-Gal4 system to overexpress various mCherry-tagged Shep isoforms (namely A, C, and E) in different cell types of the ovary. Transgenic flies were crossed to obtain F1 progeny, so that each isoform was overexpressed in the germline syncytium (using *osk*-Gal4:VP16 driver), follicle epithelium (using *lj*-Gal4 driver), or border cell cluster (using *slbo*-Gal4 driver). Irrespective of the isoform or ovarian cell type, overexpression of Shep in the ovary generally did not result in its localization to the nucleus (**Figure 3.7**). When overexpressed in the 16-cell germline syncytium (i.e., the oocyte and nurse cells), Shep isoforms remained localized to the cytoplasm (**Figure 3.7A**). Similar to observations within the germline, Shep showed no nuclear localization when overexpressed in either the follicle epithelium or border cells (**Figure 3.7B-C**). Our results clearly demonstrated that even when overexpressed in the ovary, Shep does not translocate to the nucleus.

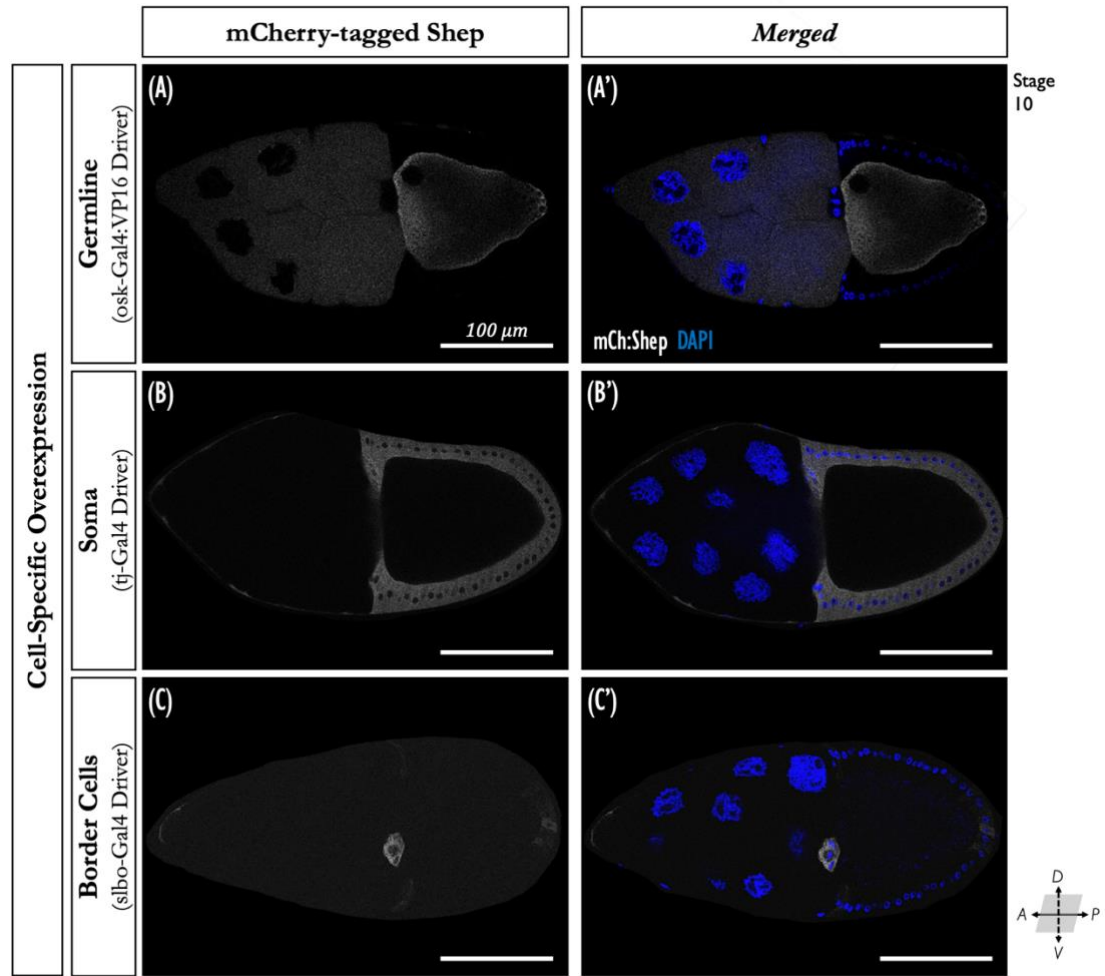


Figure 3.7: mCherry-tagged Shep isoform distribution in various ovarian tissues. Stage 10 egg chambers showing representative localization patterns of different mCherry-tagged Shep isoforms (namely, -A, -C, and -E) in the germline syncytium (panel A), follicular epithelium (panel B), and border cell cluster (panel C). For expression in distinct tissues of the ovary, different Gal4 drivers, including a single copy of the *osk-Gal4:VP16* driver for germline expression, a single copy of the *tj-Gal4* driver for expression in soma, and a single copy of the *slbo-Gal4* driver for expression in border cell clusters. These drivers were crossed with transgenic flies carrying the UAS-mCh:Shep -A, -C, and -E constructs. By combining the drivers and UAS-constructs in the F1 progeny, the UAS-Gal4 system enabled precise spatial and temporal control over the expression of mCherry-tagged Shep isoforms in these tissues. (A'–C') Overlay of both mCherry-tagged Shep and cell nuclei, which were stained with DAPI and visualized in blue. Scale bar, 100 μ m. Fly genotype(s) – Panel A is a representative of (*w*; *osk-Gal4:VP16/UASp-mCh:ShepA*); (*w*; *osk-Gal4:VP16/UASp-mCh:ShepC*); and (*w*; *osk-Gal4:VP16/UASp-mCh:ShepE*). Panel B is a representative of (*w*; *tj-Gal4/UASp-mCh:ShepA; tub-Gal80ts/+*), (*w*; *tj-Gal4/UASp-mCh:ShepC; tub-Gal80ts/+*), and (*w*; *tj-Gal4/UASp-mCh:ShepE; tub-Gal80ts/+*). Panel C is representative of (*w*; *slbo-Gal4/UASp-mCh:ShepA; slbo-LifeAct:GFP/+*), (*w*; *slbo-Gal4/UASp-mCh:ShepC; slbo-LifeAct:GFP/+*), and (*w*; *slbo-Gal4/UASp-mCh:ShepE; slbo-LifeAct:GFP/+*).

Kinesin-Dependent Localization of Shep Protein to the Oocyte Posterior Pole: The transport of maternal mRNA and proteins from the nurse cells to the oocyte is an active process that depends on the proper formation and organization of the microtubule cytoskeletal network. Specifically, this selective mode of transportation within the egg chamber depends on the polarization of the microtubule network. Before stage 6, when the minus-end accumulates in the distal-side of the oocyte, the transport of maternal factors from the nurse cells into the oocyte is mediated by the minus-end-directed motor protein, dynein (Kato and Nakamura 2012). However, during mid-oogenesis, gurken signalling between the oocyte's posterior and the adjacent overlying follicle cells triggers the reorganization of the microtubule network within the oocyte, thereby restricting plus-end microtubules to the oocyte's posterior cortex (Steinhauer and Kalderon 2006). Consequently, the transport mechanism switches to the plus-end-directed motor protein, kinesin-1, facilitating the movement of cargos destined for posterior localization to the oocyte's posterior. In light of this, and because the Shep protein is subcellularly enriched at the oocyte posterior cortex, we investigated whether its targeted enrichment at the posterior pole of the oocyte relies on the plus-end-directed kinesin machinery.

To do this, we assessed the effects on Shep-GFP subcellular localization using an RNAi-mediated knockdown approach in a Shep protein-trap background, specifically targeting *khc* levels in the germline. Initially, we generated a stock by combining flies carrying Valium22 *kinesin heavy chain* (*khc*), the force generating subunit responsible for the motor activity of kinesin-1, with flies carrying the Shep protein-trap transgene. These flies were then crossed with a double *osk-Gal4* driver to achieve a strong knockdown specifically in the germline. For the controls, we crossed the same germline-specific driver with either those without any RNAi construct (*w¹¹¹⁸* flies) or those carrying unrelated RNAi. Ovaries from females of the F1 generation were dissected and fixed for visualization of Shep-GFP fluorescence. While Shep still travelled into the oocyte, our results showed that it failed to enrich as a crescent at the posterior pole of the oocyte when compared to the control (**Figure 3.8**). In these *khc* depleted flies, Shep appears to accumulate anteriorly towards the centre of the oocyte, as shown in **Figure 3.8B-C**. The loss of the distinctive posterior crescent pattern suggests that the localization of Shep protein to the oocyte posterior cortex is kinesin dependent.

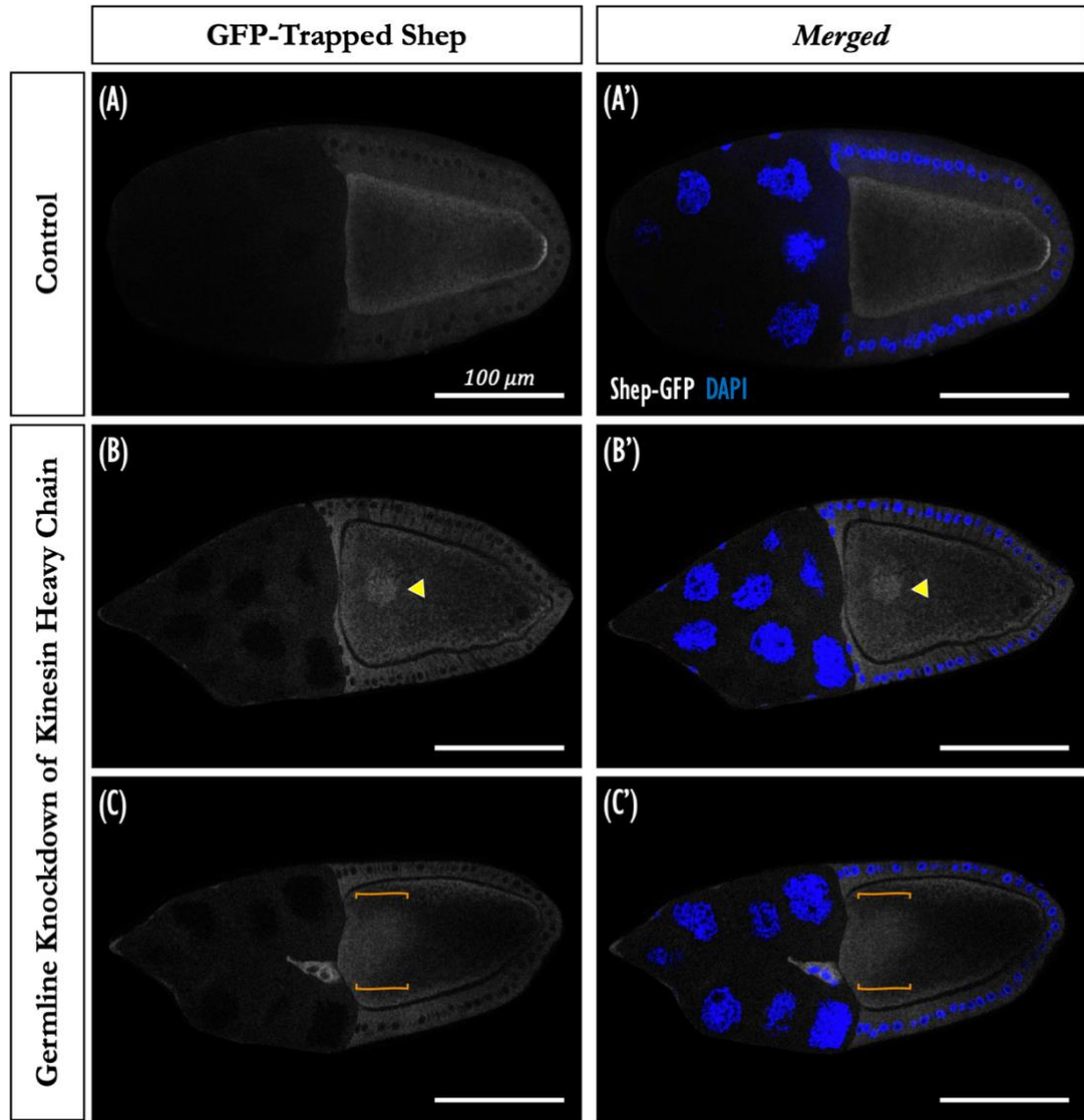


Figure 3.8: Shep subcellular localization to the oocyte posterior pole is Kinesin-1 dependent. (A) Under normal conditions, Shep protein in the germline is enriched in the oocyte and localizes to the posterior of the oocyte in stage 10. (B–C) However, RNAi knockdown of the kinesin-1 heavy chain in the germline using the double *osk*-Gal4:VP16 driver in a Shep protein-trap background results in Shep failing to enrich at the oocyte's posterior pole. Instead of forming a posterior crescent as observed under control conditions, Shep-GFP accumulates at the anterior end of the oocyte, as indicated by the yellow triangle or orange bracket. Nuclei are marked by DAPI staining in blue. (A'–C') Overlay of both mCherry-tagged Shep and cell nuclei, which were stained with DAPI and visualized in blue. Scale bar, 100 μm. Fly genotype(s) – Panel A (*w*⁻/*w*¹¹¹⁸; *osk*-Gal4:VP16/+; Shep:GFP^{CC00236}/+) and B–C (*w*; *osk*-Gal4:VP16/Valium22-*kbc*; *osk*-Gal4:VP16/ Shep:GFP^{CC00236}).

3.5 Notch Signalling Regulates Expression of Shep in the Ovary

During *D. melanogaster* oogenesis, the Delta-Notch signalling pathway is crucial for the differentiation of various cell types in the follicular epithelium along the egg chamber's

anterior-posterior axis (Xu and Gridley 2012), as detailed in Section 1.4.3 of the ‘Introduction’ Chapter. Briefly, polar and stalk cells are among the first follicle cell types to differentiate through the Delta-Notch signalling, while the remaining cells of the follicular epithelium remain undifferentiated until stage 6 (Roth and Lynch 2009), when a second round of Delta-Notch signalling induces their differentiation. Interestingly, our initial results showed that the first obvious signals of Shep protein expression in the ovarian somatic cells were detected in the polar and stalk cells (**Figure 3.4Bii-iv**). Additionally, from mid-oogenesis onwards, Shep expression becomes evident across the entire follicular epithelium. This expression gradually intensifies, especially as the follicle cells differentiate and progress through the later stages of oogenesis. Given this temporal correlation between Shep expression in the soma, follicle cell differentiation, and the activity of the Notch signalling pathway, we investigated the potential role of the Delta-Notch signalling pathway in regulating Shep expression within ovarian somatic cells. To address this, we performed *in vivo* induced RNAi-mediated knockdown of Notch signalling pathway components in a Shep protein-trap background and examined the effects on the expression of Shep-GFP proteins.

3.5.1 Downregulation of Delta Levels in the Female Germline

In the 16-cell germline syncytium, the transmembrane ligand Delta is localized on the cell surface and activates Notch receptors on the surrounding follicle cells (López-Schier and Johnston 2001), as depicted in **Figure 3.9**. Given the directionality of the Notch signalling pathway, we therefore decided to cross flies carrying the Valium22 *Delta* RNAi transgene with those carrying the germline *oskar*-Gal4:VP16 driver in a Shep protein-trap background. As controls, we mated flies bearing the same germline-specific driver to either *w¹¹⁸* flies (lacking any RNAi construct) or to those carrying unrelated RNAi constructs (targeting the Notch receptor or PTB). We used these different control conditions in RNAi experiments to mitigate unwanted artifacts stemming from either the genetic background of the Gal4 driver flies, or from overloading both the endogenous transcriptional and RNA-induced silencing complex (RISC) machineries when expressing the RNAi construct. Ovaries from F1 generation females, carrying all transgenes, were dissected and fixed to visualize Shep-GFP expression.

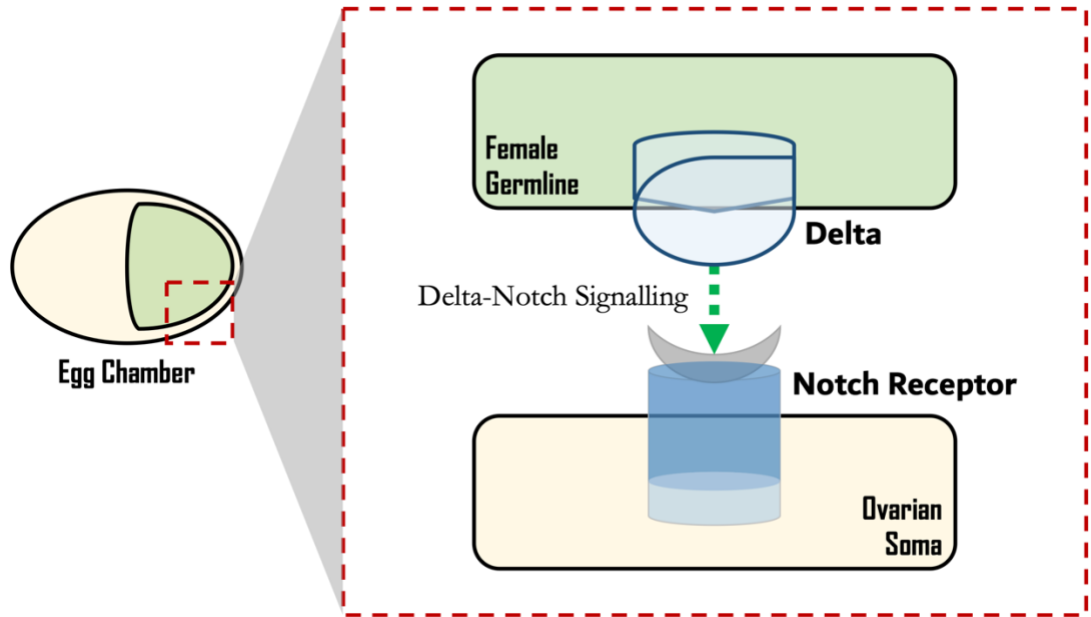


Figure 3.9: Direction of Notch signalling pathway in the *Drosophila* egg chamber. Notch signalling is required for numerous important aspects of *Drosophila* development (Xu and Gridley 2012). During oogenesis in *D. melanogaster*, the core components of this canonical pathway are the Notch ligand, which is encoded by the Delta genes, and the Notch receptor. Within the egg chamber, cells of different embryonic origins communicate intercellularly through the Delta-Notch signalling pathway. Specifically, the Notch ligand Delta, which is present on the cell surface of the germline cells (depicted as the upper cell in green), interacts with neighbouring somatic cells expressing the Notch receptor (represented as the lower cell in green). Consequently, the directionality of the Delta signals is from the germline to the soma, as denoted by the green arrow.

Given the importance of the Delta-Notch signalling pathway in oogenesis, our initial attempts to maximize RNAi knockdown by keeping the cross at higher temperatures, 29°C and 25°C, led to an arrest of oogenesis. To mitigate this, we therefore decided to maintain the crosses at a lower temperature (18°C). Consequently, oogenesis partially resumed. In these ovaries with partial Delta knockdown, we observed a mixture of phenotypes, with some ovarioles appearing wild-type like, whereas others presented with organizationally and developmentally defective egg chambers. In ovarioles that displayed a wild-type like appearance, Shep expression in both the germline and follicular epithelium appeared normal (data not shown), with an expression pattern similar to that observed in the controls (see panels A-B of **Figure 3.10**).

On the other hand, in ovarioles where Delta was effectively knocked down in the germline using RNAi hairpins, leading to defective egg chambers, the proper morphology and organization of the egg chambers were compromised (compare panels

A-D of **Figure 3.10**). In these instances, the oocyte was no longer positioned posteriorly, and the germline syncytium was no longer fully encapsulated by a layer of follicular epithelium. This is attributed to the dysregulation of intracellular communication and premature termination of follicle cell proliferation, as previously discussed. Additionally, these aberrant follicle cells, characterized by their cuboidal shape and small nuclei, exhibit cellular features indicative of an undifferentiated state (see highlighted region in panel C of **Figure 3.10**; Wu *et al.* 2008; Yu *et al.* 2008; Jia *et al.* 2015). This emphasises the pivotal role that the Notch signalling pathway plays in establishing both the posterior positioning of the oocyte and the distinct follicle cell fates during oogenesis, and shows that our RNAi experimental conditions recapitulate phenotypes observed in Delta germline cyst mutants (Torres *et al.* 2003).

Interestingly, although the oocyte was no longer correctly positioned at the posterior of the egg chamber, and the germline syncytium is not fully encapsulated by a layer of follicular epithelium, the expression of Shep in the female germline and its specific enrichment in the oocyte remains unaffected upon downregulation of Delta in the germline (**Figure 3.10C**). Nonetheless, the Shep protein no longer localizes subcellularly to the posterior pole of the oocyte. This is likely due to the disruption of oocyte positioning and repolarization of the microtubule cytoskeleton, as evidenced by the mis-localization of the oocyte within the egg chamber and the mis-positioning of the oocyte nucleus (see asterisks and lowercase letter ‘n’ in panel C of **Figure 3.10**). This observation highlights the importance of oocyte polarity, which is determined by the oocyte positioning at the egg chamber’s posterior, in ensuring the proper spatial and distinct subcellular localization of Shep.

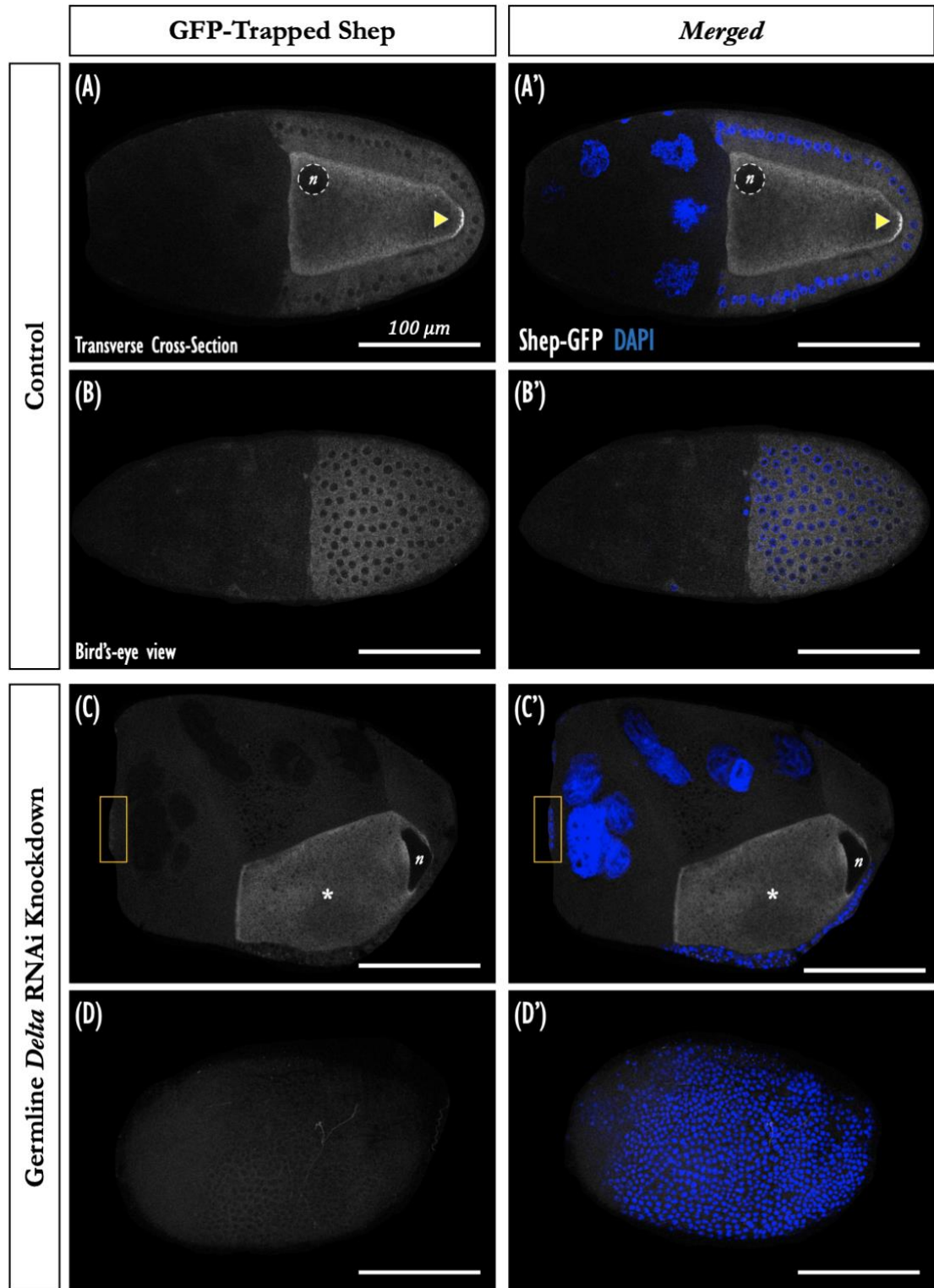


Figure 3.10: Delta RNAi knockdown in the female germline affects Shep expression in ovarian somatic cells. (A–B) Shep expression pattern in stage 10 egg chambers of control is expressed in the germline and follicular epithelium. In the germline, the Shep protein is enriched at the posterior pole of the oocyte, as indicated by the yellow triangle. **(C–D)** Knockdown of Delta in the germline via TRiP RNAi hairpins not only disrupts the spatial localization of Shep in the oocyte but also leads to a complete loss of Shep signal in the follicular epithelial cells. In ovarioles with Delta knockdown, the egg chambers exhibit morphological defects. These include the mislocalization of the oocyte within the egg chamber (indicated by the asterisk, *), the

mispositioning of the oocyte nucleus (indicated by lowercase letter ‘n’), the presence of immature follicle cells (indicated by the orange box), islands of multilayered follicular epithelium, and the absence of follicle cells surrounding the entire germline syncytium. (A’–D’) Overlay of both GFP-tagged Shep and cell nuclei, which were marked by DAPI staining and visualized in blue. For each control and experimental condition, late-stage egg chambers were examined using an optical transverse cross section (panels A and C) or a bird’s-eye view showing the egg chamber’s surface (panels B and D). The nucleus of the oocyte is indicated by the lowercase letter ‘n’. Scale bar, 100 μm . Fly genotype(s): Panels A–B (*w-/w¹¹¹⁸; osk-Gal4:VP16/+; Shep:GFP^{CC00236}/+*), and C–D (*w-; osk-Gal4:VP16/Valium22-D¹; Shep:GFP^{CC00236}/+*).

Interestingly, the knockdown of Delta in the germline led to a complete loss of Shep expression in the follicle cells (**Figure 3.10C–D**). Given the absence of Shep expression in these follicle cells that appear to be trapped in an immature state, our results suggest that proper differentiation of follicle cells and establishment of distinct follicle cell fates are crucial for Shep expression in the somatic cells of the ovary. Taken together, these findings highlight the importance of intracellular communication via the Delta-Notch signalling pathway, which governs oocyte posterior positioning and follicle cell differentiation, in enabling the proper expression and spatial localization of Shep in the ovary.

3.5.2 Knockdown of the Notch Receptor in Ovarian Somatic Cells

Notch receptors are present on the apical side of somatic follicle cells, facing the germline syncytium, which presents Delta. These receptors are activated to transduce intercellular signalling within follicle cells (**Figure 3.9**). Given the directionality of this signalling pathway, we conducted a reciprocal experiment to the one previously conducted in the germline. In this case, we examined the effects of Notch loss-of-function in ovarian somatic cells on Shep expression during oogenesis.

Here, the RNAi hairpin was expressed throughout the somatic cells under the control of the *traffic jam* Gal4 driver. Global Notch knockdown in soma often results in lethality; therefore, to avoid this, we used a temperature-sensitive Gal80 suppressor to inhibit the expression of the RNAi hairpin during development. In the days leading up to eclosion, the flies were shifted to 29°C to enable the maximum expression of the RNAi hairpin targeting the RNA encoding the Notch receptor. Similar to the Delta knockdown in the germline, the knockdown of Notch in the soma also led to an arrest of oogenesis. When the flies were maintained at a slightly lower temperature, oogenesis

partially resumed. Similar to what was observed with germline knockdown of Delta, these ovaries exhibited a mixture of phenotypes. Some ovarioles appeared wild-type like, while others displayed organizational and developmental defects in the egg chambers. In ovarioles exhibiting a wild-type like appearance, Shep protein expression in both the germline and follicular epithelium appeared normal (data not shown). This expression pattern resembled that of the control group, which used RNAi hairpins targeting delta in the soma (represented in **Figure 3.11A-B**).

On the other hand, in ovarioles displaying defective egg chambers, the normal structural organization of egg chambers is compromised, with many egg chambers no longer fully enclosed by a monolayer of follicle cells (see outlined region in panel D of **Figure 3.11**). Knockdown of the Notch receptor in the soma had no effect on the expression of Shep in the germline, although it disrupted its subcellular localization to the oocyte posterior pole (**Figure 3.11C**). This phenotype may have arisen because of a domino effect stemming from the early disruption of Notch signalling in oogenesis. Moreover, upon knocking down the Notch receptor in the follicular epithelium, we observed a range of Shep expression outcomes. Shep expression levels in follicle cells varied from complete loss to a significant reduction, and even normal levels in some cases (see panels D-E of **Figure 3.11**). The observed variability in Shep expression within the follicular epithelium may be ascribed to the incomplete and inefficient knockdown of the Notch receptor in individual follicle cells. A potential contributing factor to this variability might be the maintenance of the crosses and F1 adult female flies at suboptimal temperatures for RNAi induction, a precaution taken to prevent the arrest of oogenesis. Despite the variability observed in Shep expression, our results demonstrate the involvement of Notch signalling in regulating Shep expression in the follicular epithelium.

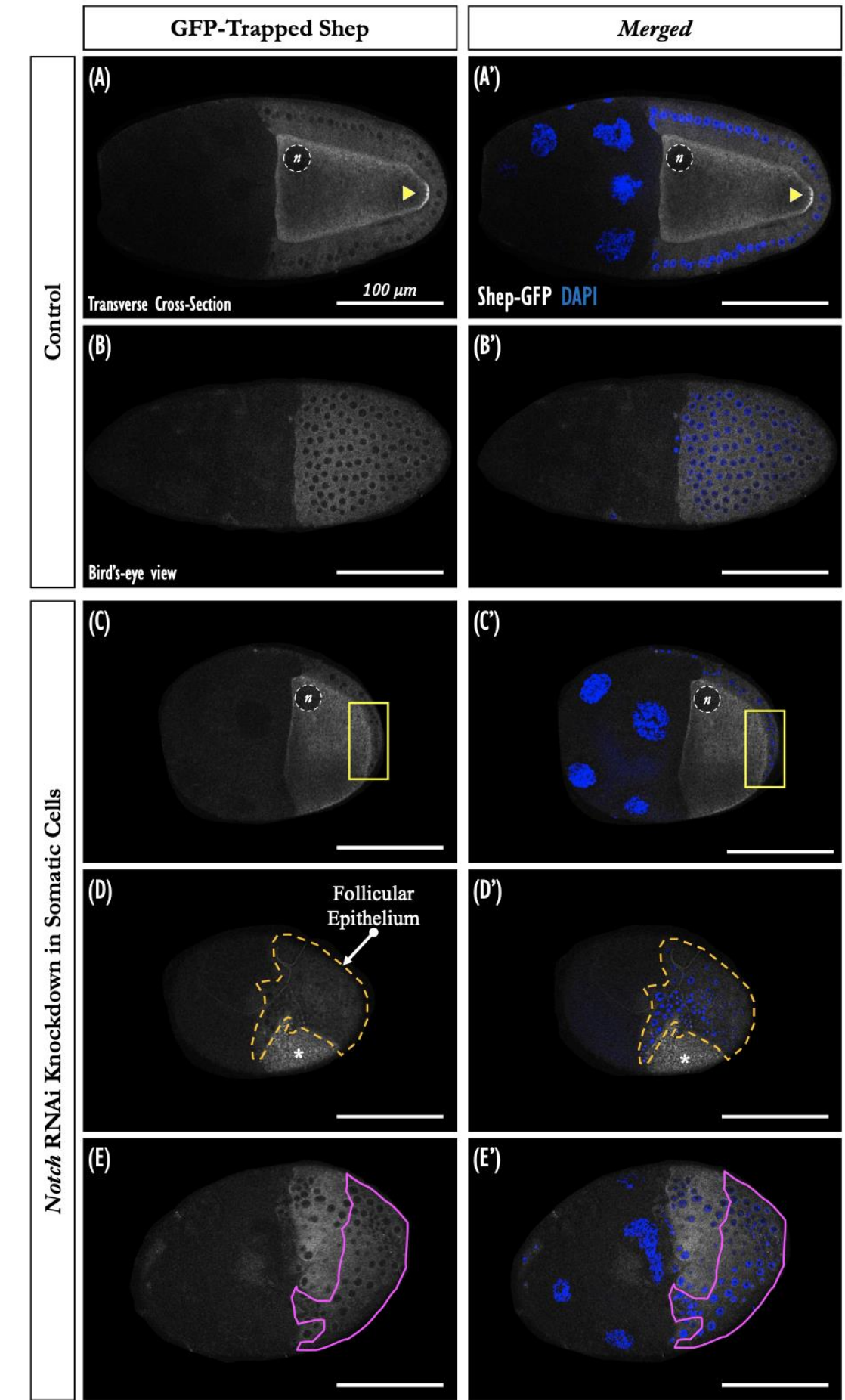


Figure 3.11: RNAi knockdown of the Notch receptor in the soma affects Shep expression in ovarian somatic cells. (A–B) In control stage 10 egg chambers, Shep expression pattern is expressed in the germline and follicular epithelium. In the germline, Shep protein is enriched at the posterior pole of the oocyte, as indicated by the yellow triangle. **(C–D)** Knockdown of the Notch receptor in ovarian somatic cells via the TRiP RNAi hairpin disrupts Shep expression in cells of the follicular epithelium. Three outcomes were observed in terms of Shep expression in the follicle cells: complete loss (depicted in panel D, outlined in orange), reduced levels (illustrated in panel E, outlined in magenta), and normal levels (shown in the un-outlined regions of panel E). In panel (D), the expression of GFP-trapped Shep from the underlying oocyte was observable, as indicated by the asterisk (*), because of the compromised follicular epithelium caused by the knockdown of the Notch receptor in the soma. In Notch receptor knockdown ovarioles, the egg chambers exhibit morphological defects, which include the following: immature follicle cells (indicated by the yellow box in panel C), and absence of follicle cells around the germline syncytium (indicated by the asterisk in panel D by the asterisk). Nuclei are marked by DAPI staining in blue (A'–E'). For each control and experimental condition, stage 10 egg chambers were examined using an optical transverse cross-section (panels A and C) or a bird's-eye view (panels B, D, and E). The nucleus of the oocyte is indicated by the lowercase letter 'n'. Scale bar, 100 μ m. Fly genotype(s) – Panels A–B (*w/w¹¹⁸; tj-Gal4/+; Shep:GFP^{CC00236}/+*) and C–E (*w; tj-Gal4/tub-Gal80ts; Shep:GFP^{CC00236}/Valium20-N*)

Interestingly, we observed a correlation between the size of the follicle cell nuclei and Shep expression levels. Follicle cells with smaller nuclei were either lacking Shep expression or expressed at lower levels. In contrast, follicle cells with larger nuclei displayed normal levels of Shep expression, as shown in panels D and E of **Figure 3.11**. This finding suggests that the differentiation state of follicle cells, as indicated by the size of their nuclei, could affect the expression of Shep. This observation aligns with our initial findings from Delta knockdown in the female germline (refer to Section 3.5.1).

Finally, the phenotypic similarities in terms of Shep expression disruption between Notch receptor knockdown in the soma and Delta knockdown in the germline (**Figures 3.11 & 3.10**) emphasize the importance of intercellular communication during oogenesis for proper Shep expression within the ovarian somatic cells. This, in turn, provides compelling evidence for the involvement of the Delta-Notch signalling pathway in the regulation of Shep expression.

3.6 Conclusions

In conclusion, the results presented here provide novel insights into the expression and distribution patterns of Shep during *Drosophila melanogaster* oogenesis, a topic previously unexplored in this specific context. Because little, if anything, is known about

Shep in the context of oogenesis, a combination of approaches was employed to characterize its gene expression pattern. These included fluorescence immunohistochemistry, *in vivo* tagging, genetic manipulation, microscopy, molecular analytical techniques, and biochemical assays.

The use of a *Shep* protein trap line, immunostaining with anti-*Shep* antibodies, and *in vivo* tagging of various *Shep* isoforms have collectively revealed that the *Shep* protein is exclusively expressed in the cytoplasm of both germline and somatic cells in the ovary. We demonstrated that, even when different isoforms were overexpressed in various tissues of the ovary, *Shep* remained cytoplasmic and did not translocate to the nucleus.

Through genetic manipulation of both the Delta ligand and Notch receptor via RNAi knockdown in the ovary, we found that *Shep* expression in the soma is under the regulation of Delta-Notch signalling pathway.

RT-PCR analyses of total ovarian RNA derived from wild-type control (*w¹¹¹⁸*) female flies have shed light on the exact transcripts of *shep*, thereby characterizing *shep* expression at the mRNA level. Six of the eight annotated *shep* transcripts are expressed in the *Drosophila* ovary, specifically transcripts A, B, E, F, H, and I. Furthermore, using protein immunodetection, we showed the expression of both small *Shep* isoforms (approximately 40kDa; E, H, and I) and large isoforms (around 65 or 70 kDa; B or A and F, respectively).

Analysis of *Shep* protein-trap CC00236 females revealed that *shep* mutants exhibited smaller ovaries, fewer ovarioles, and a reduced egg-laying capacity. Despite our primary interest in investigating the role of *Shep* in the adult female ovary –owing to its dorsoanterior and posterior enrichment in the oocyte– these findings imply that *Shep* may also play an earlier role in ovariole formation, a process that begins during larval development in the early gonad.

❧ Chapter IV ❧

Deciphering the Role of Shep
during *D. melanogaster* Oogenesis

4 Chapter IV – Deciphering the Role of Shep during *D. melanogaster* Oogenesis:

The subcellular distribution patterns described in the previous chapter suggest that Shep is a novel regulatory component of the *oskar* mRNP complex, and potentially others, such as *gurken*. This is particularly interesting because our research group identified Shep as a potential binding partner of PTB in a yeast two-hybrid screen. In the female germline, PTB plays a pivotal role as a functional component of both the *oskar* and *gurken* mRNP complexes (Besse *et al.* 2009; McDermott and Davis 2013). In fact, PTB is required for the translational repression of *oskar* mRNA and for the organization of the actin cytoskeleton at the dorsal-anterior cortex of the oocyte, which is crucial for the precise spatial regulation of Gurken signalling. Given that the role of Shep in the ovary remains largely unexplored in the existing literature, its distinct localization pattern in the oocyte, and its association with PTB, we decided to investigate the role of Shep during *D. melanogaster* oogenesis. To do this, we employed several well-established approaches for our loss-of-function and gain-of-function analyses.

Aims
4) Examine the <i>in vivo</i> effects of reducing the levels of Shep protein using an RNAi-mediated approach.
5) Characterize different <i>shep</i> mutants to identify <i>oskar</i> -related phenotypes in the absence or impairment of Shep function.
6) Study the <i>in vivo</i> effects of Shep overexpression on the regulation of key maternal mRNAs.

4.1 Loss-of-Function Analysis via RNAi-Mediated Downregulation of Shep

4.1.1 RNAi-Mediated Reduction of *shep* Levels in the Germline Resulted in No Obvious *oskar*-Related Phenotype

In the *Drosophila* nervous system, RIP-seq data from Olesnicky *et al.* (2018) identified *oskar* mRNA as a target of the Shep protein. However, they did not investigate this

further. In addition, we have previously shown that Shep displays an enriched localization at the posterior cortex of the oocyte, a pattern that coincides with the localization of the *oskar* mRNP complex (Figure 3.1 in Chapter 3). Considering these findings, *oskar* mRNA represents an ideal candidate for investigating the regulatory role of Shep in mRNA metabolism within the fly ovary. Therefore, we hypothesized that cytoplasmic Shep in the female germline binds to *oskar* mRNA not for hitchhiking its transport, but rather to regulate this mRNA at the post-transcriptional level. The Shep-*oskar* interaction could stabilize the mRNA either by preventing its targeted degradation or, in a manner similar to that of PTB (Besse *et al.* 2009), by potentially repressing the translation of the Oskar protein.

To investigate this hypothesis, we opted to deplete Shep levels in the female germline using RNAi-mediated *shep* knockdown and then examined the consequent effects on the distribution and expression of *oskar* mRNA and Oskar protein. On the basis of our hypothesis, we anticipate a range of potential *oskar* dysregulation phenotypes. These may include mRNA mislocalization, ectopic Oskar protein expression, excessive translation of Oskar protein, or even a combination thereof.

We started our analysis using a fly stock from the Bloomington *Drosophila* Stock Center, which carries a single RNAi UAS-construct on the second chromosome targeting *shep*, available as part of the TRiP collection. Because a single RNAi UAS construct against *shep* proved insufficient to induce an observable *oskar*-related phenotype (data not shown), we decided to generate a fly stock carrying a Walium22 UAS-RNAi construct on the third chromosome. The hairpin sequence from the already available TRiP Walium22-*shep* RNAi construct on the second chromosome was used to generate this new construct. Following the completion of molecular cloning and microinjection of the pWalium22-*shep* RNAi plasmid, we combined the two transgenic UAS-RNAi *shep* stocks by crossing flies from these two RNAi lines together, thereby generating a fly stock that carried both copies of the *shep* RNAi construct (i.e., the one we generated on the third chromosome with the one from the TRiP collection on the second chromosome).

To maximize the expression of the RNAi hairpins, we used flies carrying two copies of the strong germline-specific *osk*-Gal4:VP16 driver on the second and third chromosomes. These flies were crossed with those carrying two copies of the UAS-RNAi constructs targeting *shep* (refer to genotype #1 in **Table 4.1**). This allowed for the specific expression of hairpins in the germline, starting at stage 1 of oogenesis. As controls, we crossed the same germline-specific driver with wild-type (*w¹¹⁸*) flies, which did not carry any RNAi constructs, or alternatively, with those carrying RNAi UAS constructs targeting eGFP, which is not expressed in the fly. Larvae from the F1 generation were maintained at 29°C to maximize the expression of hairpin molecules in the germline, thereby ensuring a more robust knockdown.

Table 4.1: List of genotypes analysed for knockdown of *shep* in the female germline.

#	Germline Driver	Genotype	Comments
1	<i>oskar</i> Gal4:VP16 driver on 2 nd and 3 rd chromosome	<i>w</i> -; <i>osk</i> -Gal4:VP16/Valium22- <i>shep</i> ; <i>osk</i> -Gal4:VP16/Walium22- <i>shep</i>	Panel B of Figure 4.1
2	Maternal Triple Driver (MTD)	<i>w</i> -/ <i>otu</i> -Gal4:VP16; <i>nos</i> -Gal4/Valium22- <i>shep</i> ; <i>nos</i> -Gal4:VP16/Walium22- <i>shep</i> .	Panel C of Figure 4.1
3	<i>nanos</i> Gal4 driver on the X chromosome with the Shep GFP protein-trap in the background	<i>w</i> -/ <i>nos</i> -Gal4; Valium22- <i>shep</i> ; Shep-GFP ^{CC00236} /Walium22- <i>shep</i>	Data not shown
4	<i>nanos</i> Gal4 driver on the X chromosome with the <i>oskar</i> Gal4:VP16 driver on the 3 rd chromosome	<i>w</i> -/ <i>nos</i> -Gal4; Valium22- <i>shep</i> /+; <i>osk</i> -Gal4:VP16/Walium22- <i>shep</i>	
5	<i>oskar</i> Gal4:VP16 driver on 2 nd chromosome with Shep GFP protein-trap in the background	<i>w</i> -; <i>osk</i> -Gal4:VP16/Valium22- <i>shep</i> ; Shep-GFP ^{CC00236} /Walium22- <i>shep</i>	
6	<i>oskar</i> Gal4:VP16 driver on 2 nd chromosome and <i>nanos</i> Gal4: VP16 driver on 3 rd chromosome	<i>w</i> -; <i>osk</i> -Gal4:VP16/Valium22- <i>shep</i> ; <i>nos</i> -Gal4:VP16/Walium22- <i>shep</i>	
7	<i>oskar</i> Gal4:VP16 driver on 2 nd chromosome and <i>bam</i> Gal4: VP16 driver on 3 rd chromosome	<i>w</i> -; <i>osk</i> -Gal4:VP16/Valium22- <i>shep</i> ; <i>bam</i> -Gal4:VP16/Walium22- <i>shep</i>	
8	<i>oskar</i> Gal4:VP16 driver on 2 nd chromosome with <i>shep</i> deficiency	<i>w</i> -; <i>osk</i> -Gal4:VP16/Valium22- <i>shep</i> ; <i>Df</i> (3L) ^{ED210} /Walium22- <i>shep</i>	

Despite our efforts, we were not able to observe any obvious *oskar* dysregulation phenotypes following the RNAi-mediated depletion of *shep*, even when using two UAS RNAi constructs (**Figure 4.1B**). In fact, *oskar* mRNA continued to localize to the

posterior pole of the oocyte, with translation solely occurring at that location, indicating no obvious impact on its regulation.

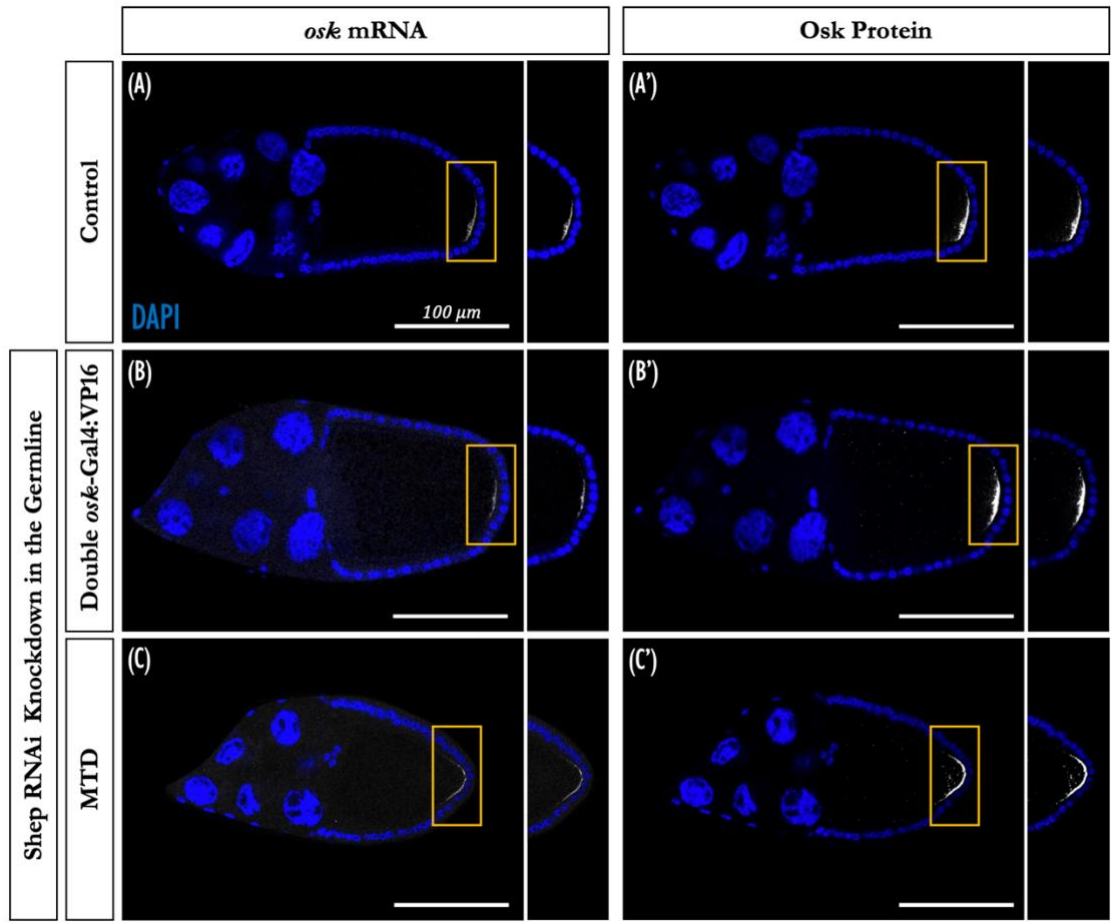


Figure 4.1: Expression of *oskar* in *shep* RNAi-mediated knockdowns in the female germline. Fluorescence *in situ* hybridization followed by immunohistochemistry to visualize *oskar* mRNA (A–C) and Oskar protein (A'–C') in stage 10 oocytes. Localization and translation of *oskar* were examined in control conditions (A) and two germline loss-of-function (B–C). Using RNAi, *shep* was downregulated using various germline-specific Gal4 drivers. At stage 10 of oogenesis, *oskar* mRNA and Oskar protein localise to the posterior pole of the oocyte. In all panels, the orange box indicates the magnified posterior portion of the oocyte. Scale bar, 100 μ m. Fly genotype(s) – Panel A (*w*; *osk*-Gal4:VP16/*Valium22-eGFP*; *osk*-Gal4:VP16/*Valium22-eGFP*), B (*w*; *osk*-Gal4:VP16/*Valium22-shep*; *osk*-Gal4:VP16/*Valium22-shep*), and C (*w*-*otu*-Gal4:VP16; *nos*-Gal4/*Valium22-shep*; *nos*-Gal4:VP16/*Valium22-shep*).

We then proceeded to induce the expression of the RNAi hairpins targeting *shep* using other germline-specific drivers. Given that *Shep* is expressed very early in oogenesis, specifically in germline stem cells, we began by choosing drivers that induce expression slightly earlier compared with the *osk*-Gal4:VP16 driver. One such driver is the maternal triple driver (see **Figure 4.2**), which induces Gal4 protein expression under

the *otu* and *nanos* gene promoters. Unfortunately, the use of the maternal triple driver, together with two copies of *shep* RNAi hairpin constructs (genotype #2 in **Table 4.1**), did not yield any obvious phenotype (**Figure 4.1C**).

We then decided to sensitize the background of our RNAi conditions using *shep* mutants. To achieve this, we individually combined *oskar*-Gal4:VP16 and *nanos*-Gal4 with the Shep GFP protein trap mutant line (see genotype #3 and 5 in **Table 4.1**). In addition, we combined the *oskar*-Gal4:VP16 driver with a null *shep* deficiency (see genotype #8 in **Table 4.1**). Subsequently, these various drivers in a sensitized background were crossed with flies carrying two copies of the *shep* UAS-hairpin constructs to systematically analyze *oskar* expression. Despite these efforts, we could not observe any clear dysregulation of *oskar* expression when Shep levels were depleted in the germline (data not shown).

In our search of a driver that would induce a robust and early *shep* depletion, potentially resulting in an *oskar*-related phenotype, we decided to broaden our approach and explore other combinations of germline drivers. Our strategy was to combine into a single stock the robust *osk*-Gal4:VP16 driver with other earlier-expressing Gal4 drivers, such as *nos*-Gal4:VP16 and *ban*-Gal4:VP16. These stocks containing different combinations of germline drivers were crossed with flies carrying the *shep* RNAi constructs (refer to genotypes #6 & 7 in **Table 4.1**). Despite the F1 females being maintained at 29°C for optimal RNAi-mediated depletion, and exhausting all possible driver combinations with the two copies of RNAi construct to achieve robust knockdown of *shep* in the female germline, we observed no clear phenotype in either *oskar* mRNA or Oskar protein expression patterns (data not shown).

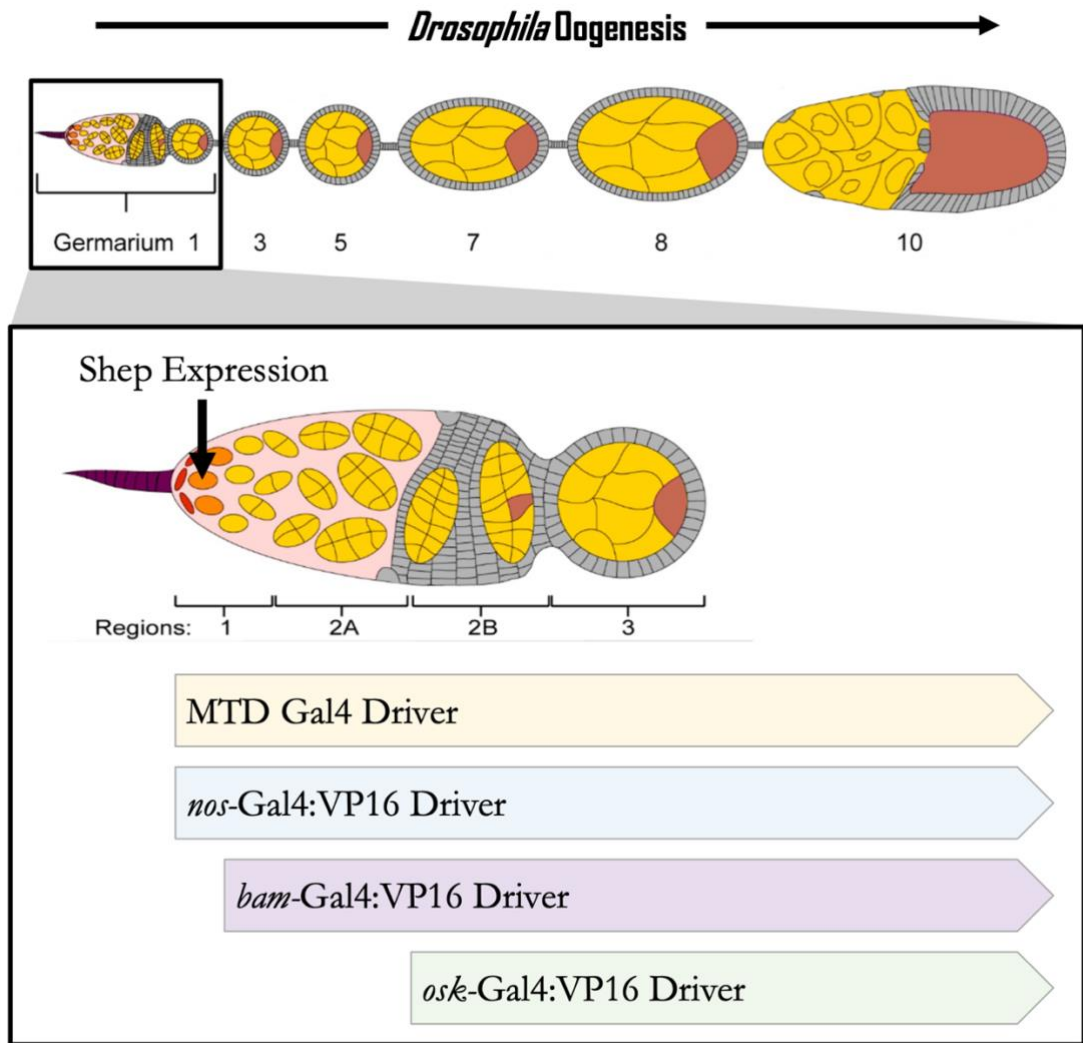


Figure 4.2: Expression time-window of various germline-specific Gal4 drivers during oogenesis. A schematic illustration of a *Drosophila* ovariole composed of several egg chambers at different developmental stages of oogenesis, from the germarium to stage 10. Egg chambers in the ovariole are color-coded: the oocyte in brown, nurse cells in yellow, and somatic cells in grey. A magnified view of the germarium showing *Shep* expression in germline stem cells (GSC, in orange) and the induction times of different Gal4 driver stocks (Chen and McKearin 2003; ElMaghraby *et al.* 2022). Both the maternal triple driver (MTD), which contains the *otu*- and *nos*-Gal4 drivers, and the *nos*-Gal4:VP16 driver are induced very early in germline stem cells. In contrast, *bam*-Gal4 is induced slightly later, specifically in the daughter cells of the GSC that are committed to differentiation, whereas *osk*-Gal4:VP16 is activated in the 16-cell germline cyst around stage 2B of the germarium. Diagrams under an open access license were obtained and modified from Lebo and McCall (2021).

Finally, the lack of an obvious *oskar*-related phenotype in our previous experiments could be the result of a poor depletion of *Shep* protein levels by the RNAi harpins used in this study. Therefore, we decided to perform a test to ensure that the RNAi hairpins were able to knockdown *shep* levels in the germline. To do this, we utilized the same RNAi line against *shep*, controlled by the germline-specific *osk*-Gal4:VP16 driver as

previously used, but this time within a Shep protein-trap background to monitor the levels of Shep protein in our RNAi knockdown experiments (genotype #5 in Table 4.1). In this context, we observed a loss of the Shep-GFP signal in the germline (Figure 4.3), which indicated an efficient RNAi downregulation under the conditions used in our study.

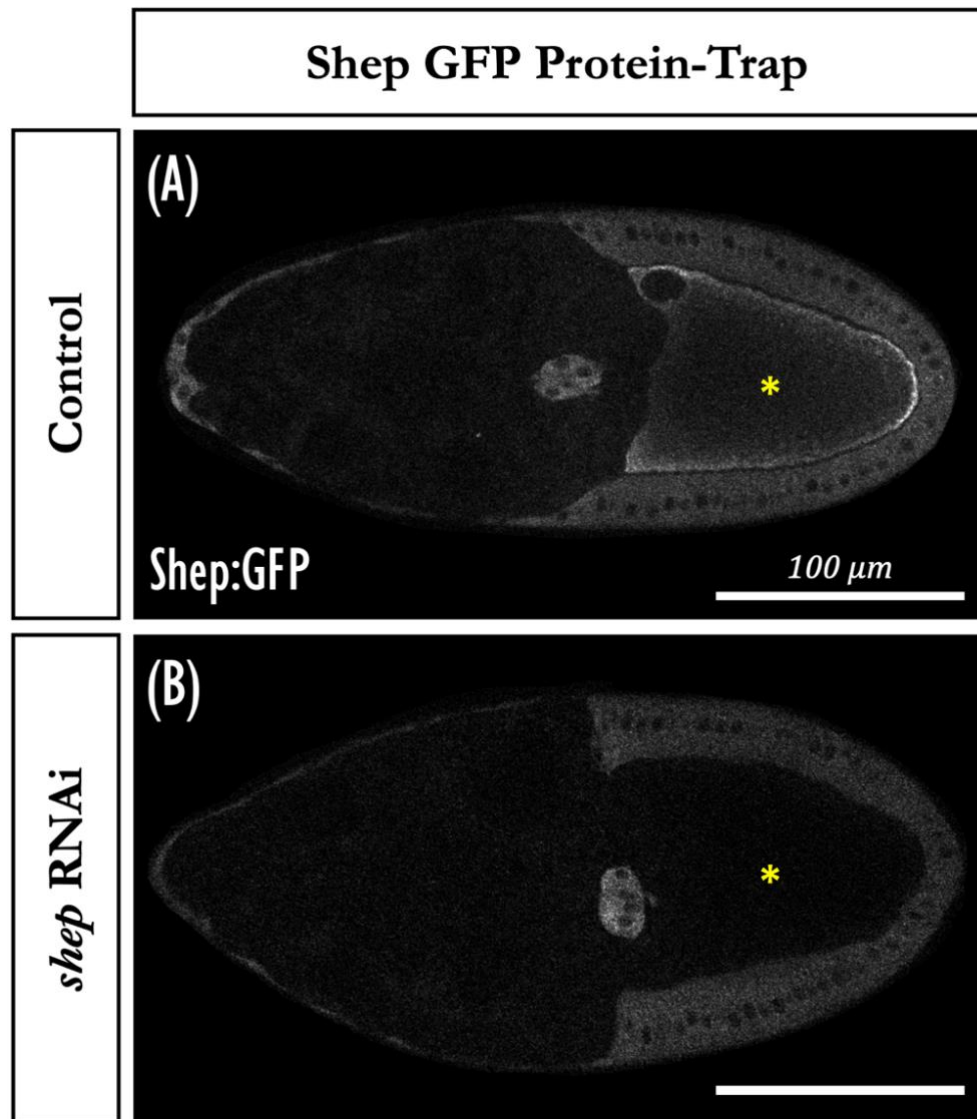


Figure 4.3: Expression of *shep* RNAi hairpins efficiently depletes Shep-GFP levels in the female germline. (A) In control ovaries, the Shep GFP protein trap is detected in both the oocyte (indicated by the yellow asterisks) and ovarian somatic follicle cells of stage 10 egg chambers. Within the oocyte, Shep localizes to the cortex with enrichment at the posterior pole and anterior corners. **(B)** RNAi hairpin expression in the female germline significantly depletes Shep-GFP levels in the oocyte. Scale bar, 100 μm. Fly genotype(s) – Panel A (*w⁻; ask-Gal4:VP16/+; Shep:GFP^{CC00236}/+*), B (*w⁻; ask-Gal4:VP16/Valium22-shep; Shep:GFP^{CC00236}/Valium22-shep*).

4.1.2 *oskar*-Unrelated Phenotypes Upon Germline Reduction of Shep Levels

Despite not observing any obvious *oskar*-related phenotypes in our previous experiment, we did notice three *oskar*-unrelated phenotypes when Shep levels in the germline were depleted via RNAi. These include morphological and structural defects in egg chambers, impaired border cell migration, and loss of Lamin C expression in stalk cells. These phenotypes were primarily investigated in the context of F1 ovaries that carried two copies of both the germline-specific *osk*-Gal4:VP16 driver and the *shep* RNAi UAS-hairpin construct. For the controls, we crossed the same germline-specific driver with *w¹¹¹⁸* flies that did not carry any RNAi constructs. Alternatively, we also used controls that carried two copies of RNAi UAS constructs targeting *eGFP*. Throughout our RNAi analysis, larvae from the F1 generation were consistently maintained at 29°C to ensure optimal expression of the hairpin molecules, which in turn would ultimately maximize the depletion of Shep levels in the germline.

4.1.2.1 *Amorphic Egg Chambers*

RNAi-mediated knockdown of Shep in the germline using two copies of the UAS-hairpin constructs led to approximately 5% of egg chambers showing a distinct morphological defect. These egg chambers appeared to have lost their organizational structure and integrity (**Figure 4.4A**). The severity of the phenotype varies, ranging from a series of multiple dysmorphic egg chambers to a single one within an ovariole (see panels A and C of **Figure 4.4**). Nonetheless, this rare yet reproducible phenotype (N= a minimum of three independent experimental replicas) is unique to *shep* RNAi knockdown and is not observed in any of the control groups. The phenotype we observed does not correspond to egg chambers undergoing programmed cell death via apoptosis, which is characterized by distinct morphological hallmarks such as chromatin condensation of the nurse cell nuclei, as described by Lebo and McCall (2021). These characteristic features of apoptosis are not obvious in the Shep RNAi egg chambers (**Figure 4.4B**). Remarkably, in these dysmorphic egg chambers, the 16-cell germline syncytium is no longer encapsulated by a monolayer of follicular epithelium (see panels A and B of **Figure 4.4**).

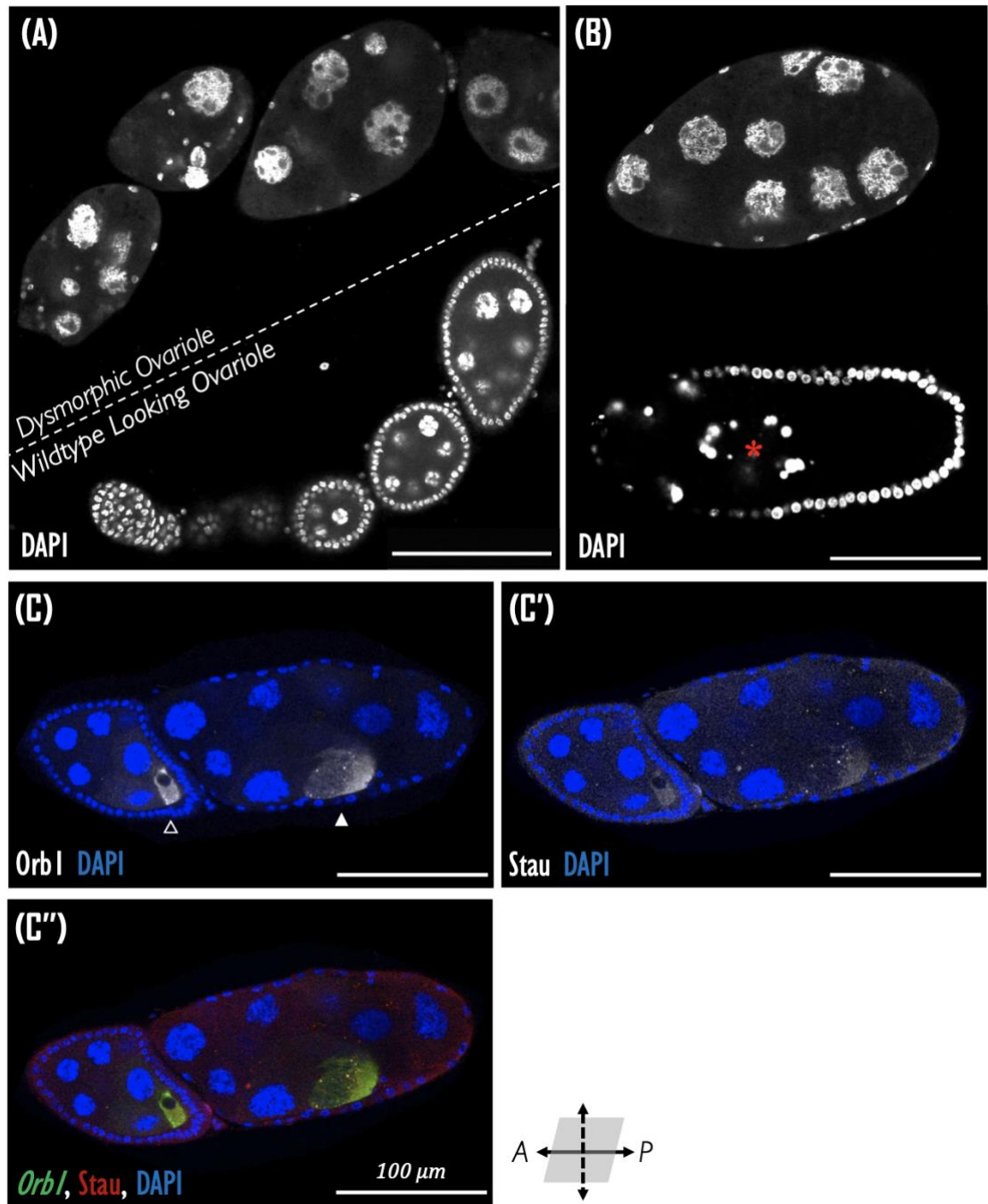


Figure 4.4: Depletion of *shep* levels in the female germline affects egg chamber architecture. (A) Ovarioles made up of egg chambers presenting dysmorphic (top) or wild-type-looking (bottom) organizational structures. Nuclei are marked by DAPI staining in white (panels A and B) or blue (panels C–C’). (B) Comparison of *Shep* RNAi-induced dysmorphic and apoptotic egg chamber nuclei. A side-by-side comparison with DAPI staining showing chromatin condensation of the nurse cell nuclei in an egg chamber undergoing apoptosis (bottom; indicated by red asterisk, *) compared with chromatin of dysmorphic egg chamber (top). (C) Egg chambers double-stained for Orb1 and Staufen, which serve as markers of the oocyte. The oocyte within a wild-type-looking or dysmorphic egg chamber is indicated by a white outlined or filled-in triangle, respectively. Scale bar, 100 µm. Fly genotype – *shep* RNAi knockdown driven by *osk*-Gal4:VP16 drivers (*w*;⁻ *osk*-Gal4:VP16/Valium22-*shep*; *osk*-Gal4:VP16/Walium22-*shep*).

Furthermore, to establish whether the oocyte was properly specified and positioned at the posterior of the egg chamber, the F1 ovaries were immunostained with two well-known oocyte markers, Orb1 and Staufen. The oocyte within these dysmorphic egg chambers seemed to be specified, as indicated by the accumulation of both Orb1 and Staufen markers within a single cell of the germline syncytium (**Figure 4.4C**). Interestingly, however, the oocyte was mislocalized and no longer positioned at the posterior end of the egg chamber (**Figure 4.4C**). Altogether, these observations not only suggest a role for *Shep* in maintaining proper egg chamber structure and oocyte positioning but also emphasize the importance of *Shep* during oogenesis.

The dysmorphic egg chambers, which result from *Shep* depletion in the germline, share phenotypic resemblance with the observations made upon RNA-mediated knockdown of *Delta* in the germline (see Section 3.5.1 of Chapter 3), particularly in terms of the lack of an intact follicular epithelium encapsulating the entire germline syncytium and the mispositioned oocyte. These observations further suggest a potential interplay between *Shep* and Notch signalling during oogenesis.

4.1.2.2 Defective Border Cell Migration

Another interesting observation derived from our germline *shep* RNAi knockdown experiments was the cell nonautonomous effect on the migration of the border cell cluster. When *shep* levels in the germline were depleted using two copies of the UAS-hairpin construct, approximately 66% of stage 10 egg chambers displayed a partial border cell migration defect (N= three independent replicas). In these egg chambers, we noticed that one or several border cells were often left behind mid-migration, while the rest of the cluster continued its migration towards the oocyte's anterior (**Figure 4.5A**). Surprisingly, this partial migration defect was also observed, albeit at a much lower prevalence, in both control groups (**Figure 4.5B**). The driver control group, which did not express RNAi, and the other control group, which expressed RNAi against eGFP, displayed migratory defects in 3% and 21% of stage 10 egg chambers, respectively (**Figure 4.5B**). Nevertheless, when compared with both controls, the frequency of this defect in stage 10 egg chambers was statistically significantly higher in RNAi-mediated *shep* knockdown samples, as determined by the one-way ANOVA ($p=$

0.0005) with Tukey’s HSD test for multiple comparisons, as shown in **Figure 4.5B** (Tukey’s HSD test: $P < 0.005$ with eGFP RNAi control and $P < 0.0005$ with Gal4-driver control). Additionally, there was no statistically significant difference between the two control groups ($p = 0.1209$).

These results highlight several critical points. First, the importance of controls: in this case, it appears that the P-element insertions of the *osk*-Gal4:VP16 germline-specific drivers impact the process of border cell migration. Second, it seems that overloading the RNAi machinery/pathway may also affect the efficiency of this migratory process. Third, despite these factors, our results suggest that Shep levels in the nurse cells may influence the migratory properties of border cells. Interestingly, adequate levels of the classical DE-Cadherin are essential in both germline nurse cells and migrating cells, providing the adhesion and traction required for migration to occur (Niewiadomska *et al.* 1999; Cai *et al.* 2014). This raises the question of whether Shep in the germline regulates DE-Cadherin levels, possibly as a regulatory component of its mRNP complex.

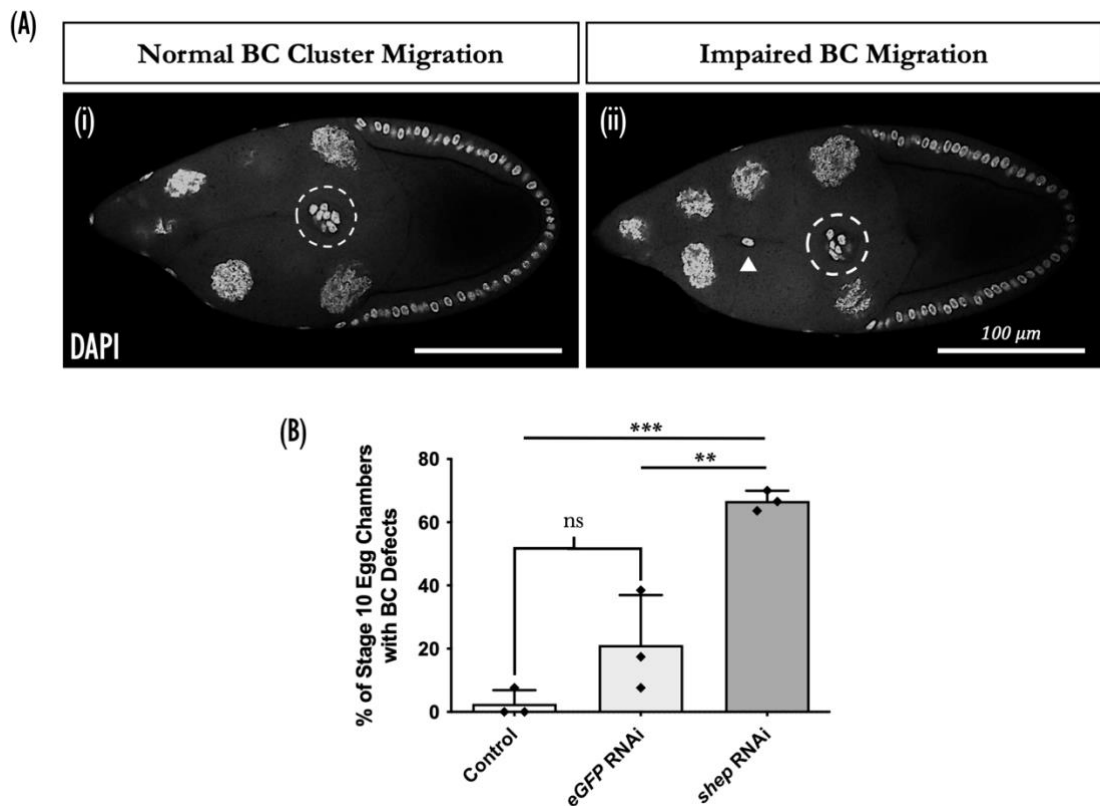


Figure 4.5: Shep levels in the female germline can impact border cell migration. (A) Examples of stage 10 egg chambers in which RNAi knockdown of Shep is driven by *osk*-Gal4:VP16 or two controls, showing normal (as seen on the left) or disrupted (as seen on the right)

migration of border cells. Fixed stage 10 egg chambers stained with DAPI (blue) to mark nuclei. Border cell clusters are indicated by the dashed circle, with white triangles pointing to border cells presenting incomplete migration. A minimum of three independent experiments were performed. Scale bar, 100 μm . **(B)** Scoring of border cell migration defects. The bar graph shows the average percentages of stage 10 egg chambers displaying border cell migration defects. Three independent experimental replicates were conducted. Error bars represent mean \pm SD. One-way ANOVA with Tukey's multiple comparison test was performed between all the group combinations (ns: $p > 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$) using GraphPad. Abbreviation – BC: border cells. Fly genotype(s) – control (*w*-/*w*¹¹¹⁸; *ask*-Gal4:VP16/+; *ask*-Gal4:VP16/+), eGFP control (*w*; *ask*-Gal4:VP16/Valium22-*eGFP*; *ask*-Gal4:VP16/Valium22-*eGFP*), and *shep* RNAi knockdown (*w*; *ask*-Gal4:VP16/Valium22-*shep*; *ask*-Gal4:VP16/Valium22-*shep*).

4.1.2.3 Differential Expression of Lamin C in Stalk Cells

Stalk cells serve as linkers between adjacent egg chambers and are known to endure mechanical stresses and tensions as they move through the ovary (Van De Bor *et al.* 2021). In *Drosophila*, Lamin C has been recognized as a critical component of the nuclear lamina involved in regulating gene expression, signal transduction, and mechanotransduction (Dialynas *et al.* 2010; González-Cruz *et al.* 2018).

Another piece of evidence pointing towards a non-cell autonomous role of Shep during oogenesis was observed when *shep* levels were depleted in the germline using two copies of the UAS-hairpin construct. The reduction of *shep* levels in the female germline indirectly affected the expression of Lamin C in neighbouring stalk cells. We inadvertently stumbled upon this phenotype during an earlier analysis, while attempting to unravel the genetic interaction between Shep and PTB. Lamin C is a unique type of nuclear intermediate filament that assembles at the nuclear periphery (González-Cruz *et al.* 2018), and is often used as a marker of differentiated stalk cells (Pearson *et al.* 2016). When *shep* was knocked-down in the germline using RNAi, Lamin C levels were decreased in approximately 77% of the examined interfollicular stalk cells compared with the control group (**Figure 4.6A-B**). Interestingly, this phenotype was specific to flies expressing the RNAi hairpins against *shep*, as Lamin C expression appeared normal in flies expressing RNAi hairpins against *ptb* (**Figure 4.6C**). Our findings demonstrate that the depletion of *shep* in the female germline indirectly affects the levels of Lamin C in interfollicular stalk cells.

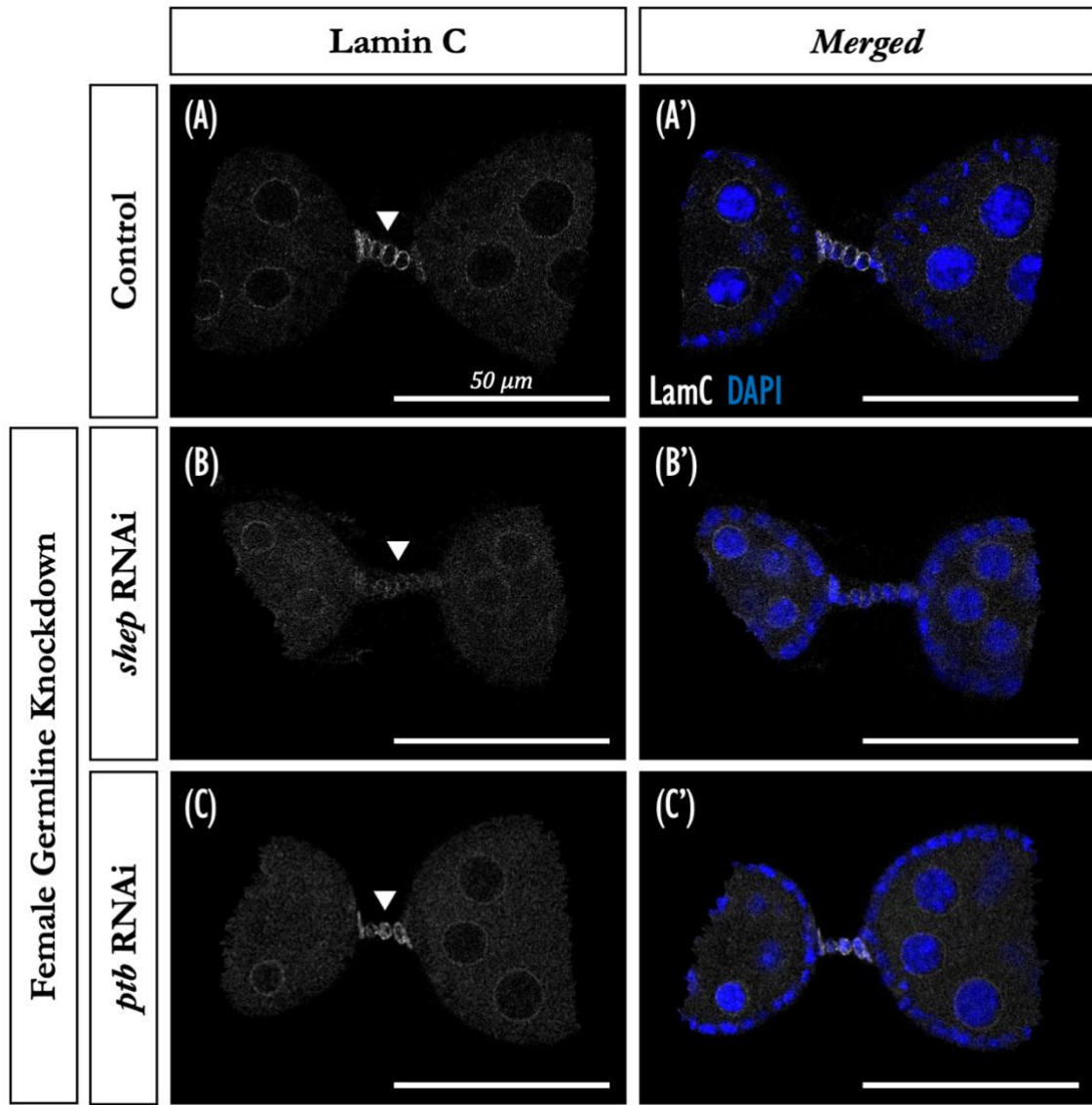


Figure 4.6: *Shep* levels in the female germline regulate Lamin C expression in stalk cells. RNAi-mediated reduction of *shep* specifically in the germline downregulates Lamin C expression in the cells that make up the stalk. Ovarioles were stained for Lamin C (white) and DNA (blue). White triangles point towards the interfollicular stalk that connects two egg chambers. Scale bar, 50 μ m. Fly genotype(s) – Panels A (*w⁻/w¹¹¹⁸; ask-Gal4:VP16/+; ask-Gal4:VP16/+*), B (*w⁻; ask-Gal4:VP16/Valium22-shep; ask-Gal4:VP16/Valium22-shep*), and C (*w⁻; ask-Gal4:VP16/Valium22-ptb; ask-Gal4:VP16/Valium22-ptb*).

The reduction of Lamin C levels, particularly in flies expressing *shep* RNAi hairpins in the germline, suggests that stalk cell differentiation might be compromised. To better understand what is happening to these cells, we hypothesized that the cells of stalk structure exhibiting reduced levels of Lamin C either remain as undifferentiated follicle cells or adopt the polar cell fate. Alternatively, the depletion of *Shep* in the germline may disrupt the specification of the adjacent polar cells, thereby inducing a domino

effect on stalk cell specification around stage 1 of oogenesis. To test these hypotheses, we analysed the expression of FasIII, a marker of polar cells, in ovaries expressing RNAi hairpins against *shep* in the germline. Immunostaining revealed that polar cells were normally specified, as indicated by the normal expression of FasIII (Figure 4.7B-C). Furthermore, contrary to our hypothesis, stalk cells did not adopt the polar cell fate, as indicated by the absence of FasIII expression (compare panels A-C of Figure 4.7).

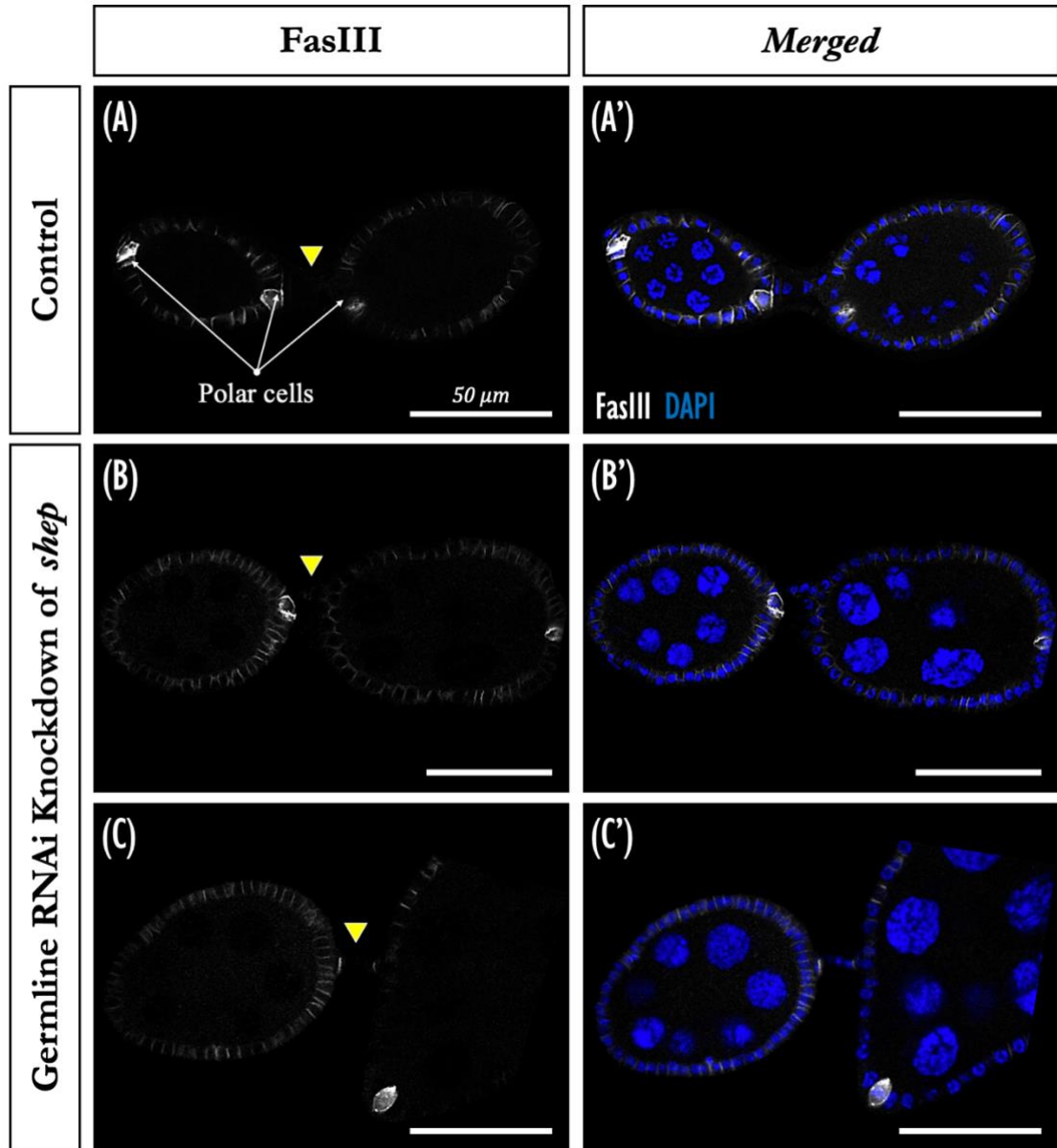


Figure 4.7: FasIII expression pattern following germline knockdown of *shep*. Depletion of *shep* specifically in the germline, achieved through RNAi, has no effect on either polar cell specification or fate transition of stalk cells to polar cells. Ovaries were stained for FasIII (white) and DNA (blue, DAPI). Yellow triangles point towards the interfollicular stalk that connects two egg chambers. Because of the 3D ovoid structure of egg chambers, the angle and placement of the specimen during mounting can affect the visibility of anterior and posterior polar cells. It is not

always possible to visualize both polar cells and stalk cells in the focal plane. Scale bar, 50 μm . Fly genotype(s) – Panels A (*w-/w¹¹¹⁸; osk-Gal4:VP16/+; osk-Gal4:VP16/+*) and B-D knockdown (*w-; osk-Gal4:VP16/Valium22-shep; osk-Gal4:VP16/Valium22-shep*).

The lack of Lamin C expression in stalk cells of ovaries expressing *shep* RNAi hairpins in the germline, together with the absence of FasIII expression within the stalk and the proper differentiation of polar cells, collectively our findings suggest that cells within the stalk remain either fully undifferentiated or partially differentiated. Interestingly, Assa-Kunik *et al.* (2007) demonstrated that Delta overexpression in the female germline affects stalk cell fate. This occurs by antagonizing JAK/STAT signalling, which is stimulated by the ligand Upd secreted by polar cells and is essential for stalk cell differentiation. Considering this, together with the results presented here, it is possible that the reduction of Shep levels in the germline may result in the upregulation of Delta, thereby leading to the phenotypes we observed.

4.2 Characterization of *shep* Deficiencies & P-element Insertions

Despite all our efforts, we failed to observe any *oskar*-related phenotype under the various RNAi conditions tested (detailed in Section 4.1). We therefore hypothesized that the pronounced expression of Shep in the cytoplasm of germline stem cells, as shown in **Figure 3.5Bi**, may play an early and crucial role in mRNA post-transcriptional regulation. Thus, this early expression of Shep, which remains unaffected by the expression of RNAi hairpins, may be sufficient to support later functions as part of the pre-assembled mRNP complexes within germline stem cells. Therefore, we decided to change our strategy and obtain several *shep* alleles from the Bloomington *Drosophila* Stock Center. Specifically, we sought stocks carrying P-element insertions in the *shep* gene and also those bearing genomic deficiencies, which result in a large deletion encompassing the *shep* gene locus. As detailed in **Table 4.2**, we crossed various *shep* stocks to combine P-element insertions with the deficiencies, and also crossed the deficiencies among themselves. We then examined Oskar protein expression and distribution in these different genetic backgrounds.

Table 4.2: Summary of the genotypes analyzed for the *shep* complementation test. * Refer to **Figure 4.8** for the relative genomic locations of deficiencies and P-element insertions within the *shep* gene.

#	Genotype(s)*	Purpose	Comments
1	<i>w¹¹¹⁸</i> ;;	Wildtype	Panel A of Figure 4.9
2	<i>w¹¹¹⁸/w</i> ;; <i>Df(3L)^{Exel6103}/+</i>	Control for Set 1	Panels C–D of Figure 4.10
3	<i>w</i> ;; <i>Df(3L)^{Exel6103}/Df(3L)^{Exel6104}</i>	Experimental genotypes (Set 1)	Panel B of Figure 4.9
4	<i>w</i> ;; <i>Df(3L)^{Exel6103}/Df(3L)^{ED210}</i>		Panel C of Figure 4.9
5	<i>w</i> ;; <i>Df(3L)^{Exel6103}/shep^{C522}</i>		Data not shown
6	<i>w</i> ;; <i>Df(3L)^{Exel6103}/shep^{BG02468}</i>		Panels C–D of Figure 4.10
7	<i>w</i> ;; <i>Df(3L)^{Exel6103}/shep^{KG10149}</i>		Panels E–F of Figure 4.10
8	<i>w</i> ;; <i>Df(3L)^{Exel6103}/shep^{BG00836}</i>		Panels G–H of Figure 4.10
9	<i>w¹¹¹⁸/w</i> ;; <i>Df(3L)^{Exel6104}/+</i>	Control for Set 2	Data not shown
10	<i>w</i> ;; <i>Df(3L)^{Exel6104}/Df(3L)^{ED210}</i>	Experimental genotypes (Set 2)	Panel C of Figure 4.9
11	<i>w</i> ;; <i>Df(3L)^{Exel6104}/shep^{C522}</i>		Data not shown
12	<i>w</i> ;; <i>Df(3L)^{Exel6104}/shep^{BG02468}</i>		
13	<i>w</i> ;; <i>Df(3L)^{Exel6104}/shep^{KG10149}</i>		
14	<i>w</i> ;; <i>Df(3L)^{Exel6104}/shep^{BG00836}</i>		

We set up two sets of fly crosses using *shep* P-element insertions. In the first set, flies carrying the P-element insertions were crossed with the Exel6103 deficiency (see genotypes #5–8 in **Table 4.2**); in the second set, the *shep* P-element insertions were crossed with another deficiency stock, Exel6104 (refer to genotypes #10–14 in **Table 4.2**). We chose these specific deficiencies for two main reasons. First, the Exel6103 deficiency has been described as a null mutant affecting all Shep isoforms, while the Exel6104 deficiency is not a null mutant and allows the expression of the small isoforms (Matzat *et al.* 2012). Second, these deficiencies not only delete a portion of the *shep* gene (**Figure 4.8**), but they also target different sets of genes on either side of the *shep* locus. Considering that only the small isoforms (i.e., E and C), not the large isoform A, localize to the oocyte’s posterior pole (**Figure 3.6** in Chapter 3), we hypothesized that the Exel6103 deficiency, which results in a null deletion of Shep, would show more profound *oskar*-related phenotypes than the Exel6104 deficiency that still expresses these isoforms. As controls, apart from using the *w¹¹¹⁸* fly stock, we also crossed the

Exel6103 and Exel6104 deficiencies with w^{1118} flies to assess the effects of the heterozygous deficiencies (i.e., on their own) on the expression of Oskar protein (see genotypes # 1, 2, & 9 in **Table 4.2**). Crosses were maintained at room temperature before analyzing the expression and subcellular distribution patterns of the Oskar protein in ovaries from various *shep* loss-of-function genetic backgrounds (genotypes listed in **Table 4.2**).

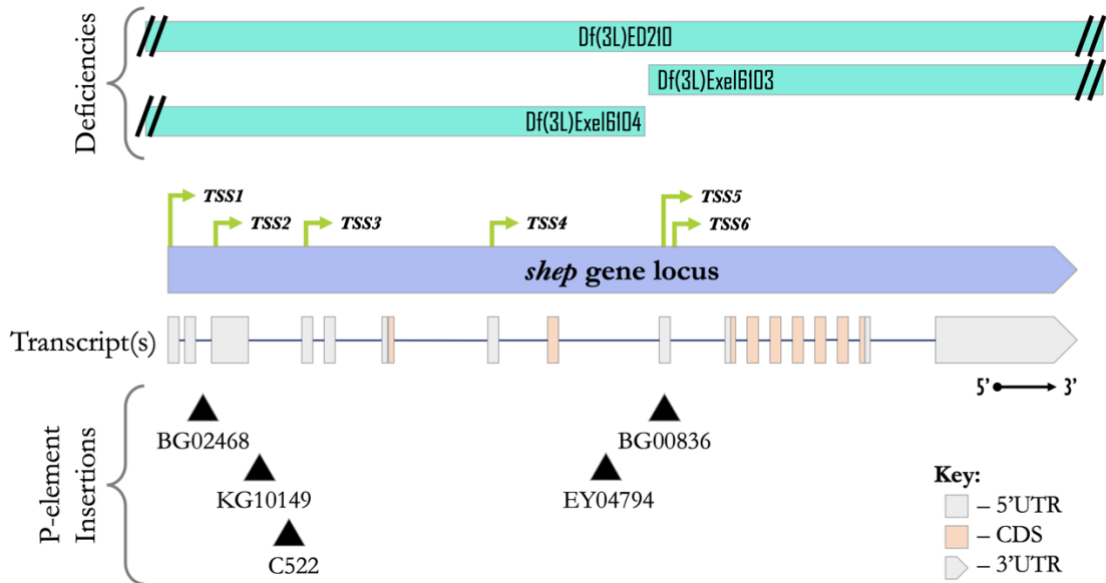


Figure 4.8: Genomic locations of deficiencies and P-element insertions in relation to the *shep* gene. A schematic representation illustrating the genomic organization of the *shep* alternative transcriptional start sites (TSS), mRNA splicing variations, P-element insertion sites, and regions deleted in the deficiencies. According to FlyBase and the embryonic modENCODE dataset, the *shep* gene can produce eight transcripts (namely A, B, D, E, F, G, H, and I) through a combination of six transcription start sites (indicated by green arrows) and alternative splicing (Gramates *et al.* 2022). In this *shep* transcript model, exons are depicted as gray and orange bars, representing untranslated and coding sequences, respectively, while lines connecting them indicate introns. Solid black triangles show the locations of P-element insertions, and the cyan bars at the top represent the regions deleted in the deficiencies (with black break symbols indicating that the deficiency deletes flanking regions that are not shown in the figure). For reference of the number of genes affected by each deficiency: ED210 (72 genes), Exel6103 (25), and Exel6104 (11). These values were obtained from the Bloomington *Drosophila* Stock Center based on the current annotations. Refer to Appendix 6.1 for more details regarding the deficiency stocks. The diagram is not drawn to scale.

We began our analysis by crossing the different deficiencies encompassing the *shep* gene with each another (see genotypes 3, 4, & 10 in **Table 4.2** and **Figure 4.8**). During oogenesis, *oskar* translation is repressed in early-stage egg chambers and is activated only when the mRNA reaches the posterior pole of the oocyte (Markussen *et al.* 1995;

Rongo *et al.* 1995). In our wild-type control females (genotype 1 in **Table 4.2**), the Oskar protein was detected only from stage 8 onwards and accumulated exclusively at the posterior cortex (**Figure 4.9A**). Remarkably, across all three of the trans-heterozygous deficiency combinations tested, we observed a distinct *oskar*-related phenotype compared with the controls. Here, the Oskar protein was detected ectopically as aggregates within the oocyte cytoplasm, as indicated by the yellow arrowheads in panels **B–D** of **Figure 4.9**. Given the developmental stage of the egg chambers (i.e., mid-oogenesis), the observed dysregulation is more likely a result of *oskar* mRNA undergoing premature translation, whereby *oskar* mRNA is ectopically translated before reaching the oocyte posterior cortex. In the Exel6103/Exel6104, Exel6103/ED210, and Exel6104/ED210 deficiency combinations, these phenotypes occurred in at least 4%, 20%, and 25% of the cases, respectively. Nonetheless, our findings from the various trans-heterozygous deficiencies hint at a role for Shep in the translational control of *oskar* mRNA.

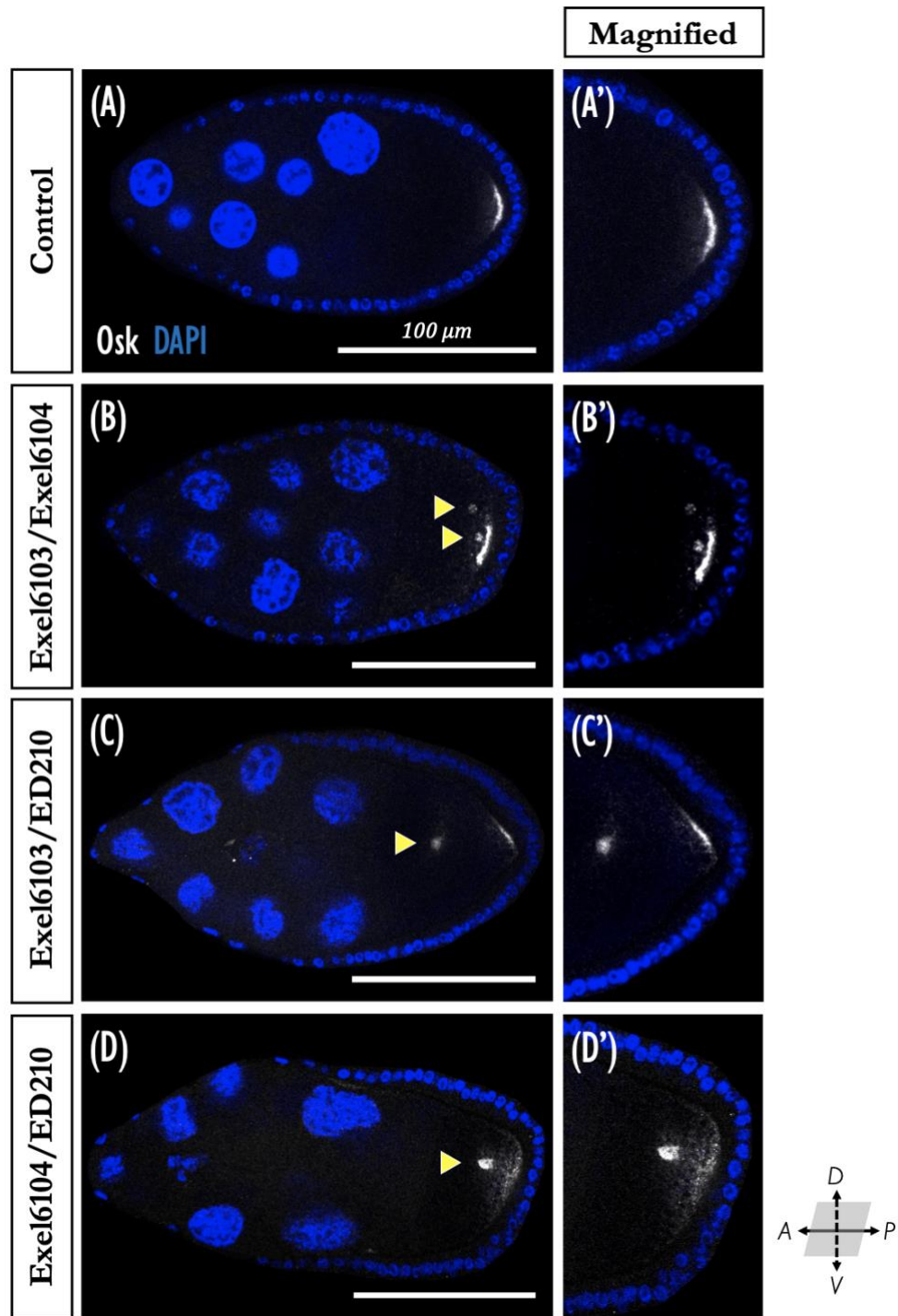


Figure 4.9: Loss-of-function of the *shep* gene dysregulates *oskar* expression in the oocyte. (A) In wild-type ovaries, the Oskar protein is first detected in the egg chamber by stage 8 and accumulates exclusively at the posterior cortex. (B–D) In ovaries with a trans-heterozygous *shep* deficiency background, the Oskar protein also appears as a dot either at the centre of the oocyte or near the cortex, as indicated by the yellow arrowheads. Nuclei are marked by DAPI staining (blue). Panels (A'–D') show a magnified view of the posterior portion of the oocyte. Scale bar, 100 μm . Fly genotype(s) – Panel A ($n^{118};$), B ($n; Df(3L)^{\text{Exel6103}} / Df(3L)^{\text{Exel6104}}$), C ($n; Df(3L)^{\text{Exel6103}} / Df(3L)^{\text{ED210}}$), and D ($n; Df(3L)^{\text{Exel6104}} / Df(3L)^{\text{ED210}}$).

Because these females carried large genomic deletions on the third chromosome, we wanted to determine if the previously observed *oskar* phenotypes were due to the loss of *shep* function as opposed to reflecting a role from some of the other genes also missing in the deficiency. Therefore, we next analyzed trans-heterozygous combinations of the deficiencies with various P-elements inserted into the *shep* gene (genotypes # 5-9 and 13-16 in **Table 4.2**). In the controls (plain heterozygous females of both deficiencies; see genotypes 2 & 9 in **Table 4.2**), Oskar protein appeared normally expressed and localized at the posterior cortex of the oocyte (**Figure 4.10 A-B**). These controls further reassured that the *oskar* phenotypes previously observed in the various trans-heterozygous *shep* deficiency mutants were not solely due to the genetic background of the single heterozygous deficiencies.

Remarkably, irrespective of the deficiency used in the experimental conditions with *shep* P-element insertions, we observed an *oskar* dysregulation phenotype that recapitulates the phenotype previously observed with trans-heterozygous *shep* deficiencies. Notably, in the second set of crosses carrying the Exel6104 deficiency, the phenotypes were rarer, occurring in less than 1% of the cases. This was anticipated because this deficiency still allows the expression of the small Shep isoforms (Matzat *et al.* 2012), which localize with *oskar* at the posterior pole of the oocyte (**Figures 3.1 & 3.6**). The *oskar* phenotypes were most prominent in the first set of crosses, specifically when the Exel6103 deficiency was combined with the BG02468, KG10149, and BG00836 *shep* P-element insertions (see genotypes #6-8 in **Table 4.2**), occurring in at least 10%, 5%, and 8% of the cases, respectively. Interestingly, ovaries of these particular genetic backgrounds displayed two distinct *oskar* dysregulation phenotypes: (i) premature *oskar* translation and (ii) lack of a tight posterior crescent (see the triangles in panels C-H of **Figure 4.10**). Here, the regulation of *oskar* mRNA is compromised at around stages 8-9 of oogenesis, resulting in the ectopic translation of Oskar, which appeared as an aggregate in the centre of the oocyte or near the posterior cortex (see yellow triangles in panels C and E-G of **Figure 4.10**). In addition, compared with the control (**Figure 4.10B**), the Oskar protein no longer forms a tight crescent at the posterior pole of the stage 9 oocyte. Instead, in these genetic backgrounds, Oskar

protein appears either as an aggregate or as a cloud near the posterior cortex of the oocyte (see green hollowed triangles in panels D and H of Figure 4.10).

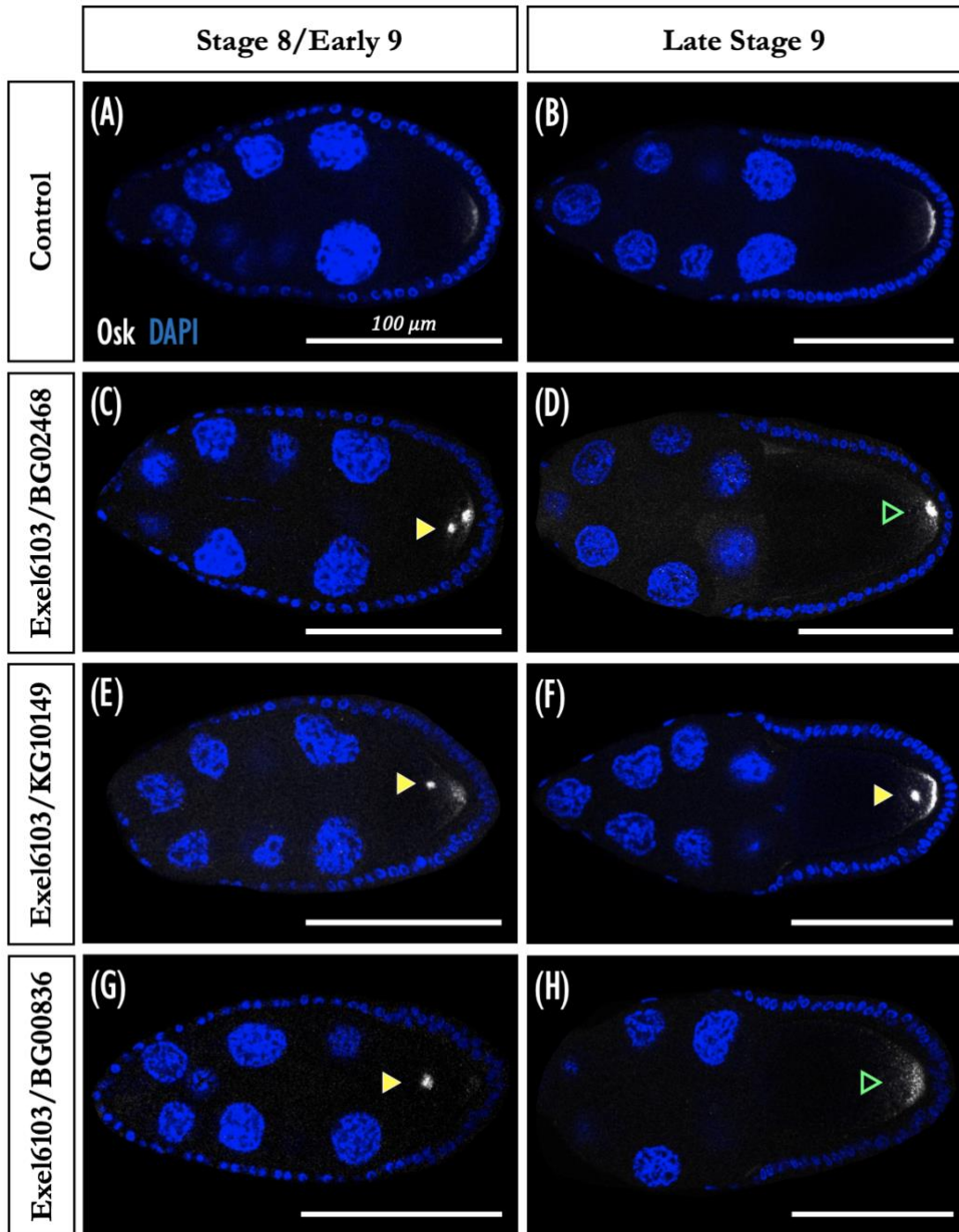


Figure 4.10: *oskar* expression is dysregulated in different *shep* mutant backgrounds. (A–B) In wild-type ovaries, the Oskar protein is first detected in the egg chamber by stage 8 and accumulates exclusively at the posterior cortex throughout oogenesis. **(C–H)** In ovaries with a *shep* P-element insertion in the Exel6103 deficiency background, Oskar appears as a dot either at the centre of the oocyte or near the cortex, as indicated by the yellow arrowheads. Additionally, in later stages, Oskar no longer retains its tight crescent shape at the posterior, as indicated by the hollowed green triangles (panels D & H). Nuclei are marked by DAPI staining (blue). Scale bar, 100 μ m. Fly genotype(s) – Panels A&B (*w*; *Df(3L)^{Exel6103}/+*), C&D (*w*; *Df(3L)^{Exel6103}/shep^{BG02468}*), E&F (*w*; *Df(3L)^{Exel6103}/shep^{KG10149}*), and G&H (*w*; *Df(3L)^{Exel6103}/shep^{BG00836}*).

Taken together, our findings indicate a regulatory role for Shep in the post-transcriptional regulation of *oskar* mRNA. When Shep's activity or levels are either compromised or insufficient, the translational repression of *oskar* mRNA seems to be prematurely lifted, resulting in ectopic expression and aggregation of the Oskar protein within the oocyte's cytoplasm. Additionally, the Oskar protein appears to no longer display its characteristic tight crescent localization pattern at the posterior cortex of the oocyte, indicating a role for Shep in the anchoring of the Oskar protein to the oocyte cortex.

Interestingly, ovaries from females with the BG00836 *shep* P-element insertions in the Exel6103 deficiency genetic background (i.e., genotypes #8) were smaller than those from the controls and other genotypes (data not shown). This could be because both the P-element insertion and the deficiency are null for the *shep* gene, leading to the most disruption in *shep* gene expression. Although we have not further analysed this observation, the small ovary phenotype here seems to mimic the small ovary phenotype we previously observed with the Shep GFP protein-trap line females (detailed in Section 3.3 of Chapter 3).

4.3 Shep Gain-of-Function Analysis

As part of the analysis conducted in Chapter 3, three isoforms of Shep were tagged with mCherry at the N-terminus, namely Shep-A, Shep-C, and Shep-E. These transgenic flies were initially generated to study their isoform-specific distribution pattern in the ovary and whether their overexpression would result in their translocation to the nucleus. Isoforms A and E were chosen for two main reasons: they represent either the large or small isoform groups, and their coding sequences were readily available from the DGRC gold cDNA. Meanwhile, Isoform C is currently unannotated in FlyBase (Gramates *et al.* 2022), but it was previously studied in Matzat *et al.* (2012). Shep-C represents a hybrid between isoforms A and E of Shep (**Figure 4.11A**). Similar to isoform E, it lacks the N-terminal extension present in isoform A. However, like isoform A, it lacks the eleven-amino-acid linker region situated between the RRM1s that is present in Shep-E and other Shep isoforms. Additionally, all three isoforms lack alternative amino acid sequences at the C-terminus, a characteristic feature of isoforms

B/D. Therefore, Shep-C is essentially a truncated version of Shep-A, lacking its N-terminal region. Furthermore, Shep-C is nearly identical to Shep-E, but without the linker region between its RNA-binding domains. Thus, comparing the results of overexpressing Shep-C with isoforms A and E would allow us to elucidate the role of the N-terminal region of isoform A, as well as the role of the eleven-amino-acid linker region located between RRM1 and RRM2 of isoform E.

In our initial analysis of the Shep isoform-specific distribution across various ovarian tissues (see Section 3.4.3 of Chapter 3), we observed a negative impact on oogenesis upon expressing the different mCherry-tagged Shep isoforms, as evidenced by the significant number of egg chambers undergoing apoptosis. Additionally, we noticed that these ovaries were particularly fragile while we dissected and mounted them: they were notably brittle, susceptible to damage even under the slightest pressure from forceps or a tungsten needle. Therefore, we decided to delve deeper into understanding the reason behind the partial arrest of oogenesis observed when Shep was overexpressed.

To do this, we crossed flies carrying a single germline-specific *osk-Gal4:VP16* driver on the third chromosome with flies containing the UASp-mCherry construct of the various Shep isoforms. As for controls, the germline-specific driver was crossed with *w¹¹¹⁸* flies, which do not overexpress any specific transgene, or with a UASp construct encoding plain eGFP. In all three analyses described below, we maintained F1 generation larvae under two conditions: either consistently at room temperature or at 25°C. After eclosion, adult F1 females of the desired genotype were allowed to mate with males at 25°C in the presence of yeast for three consecutive days before being subjected to various analyses (**Figure 4.11B**). Three sequential assessments were carried out on F1 females overexpressing Shep specifically in the female germline: (i) assessing their fertility status, (ii) inspecting the morphology of the mature eggs they laid, and (iii) examining the expression of some of the key maternal mRNAs during oogenesis.

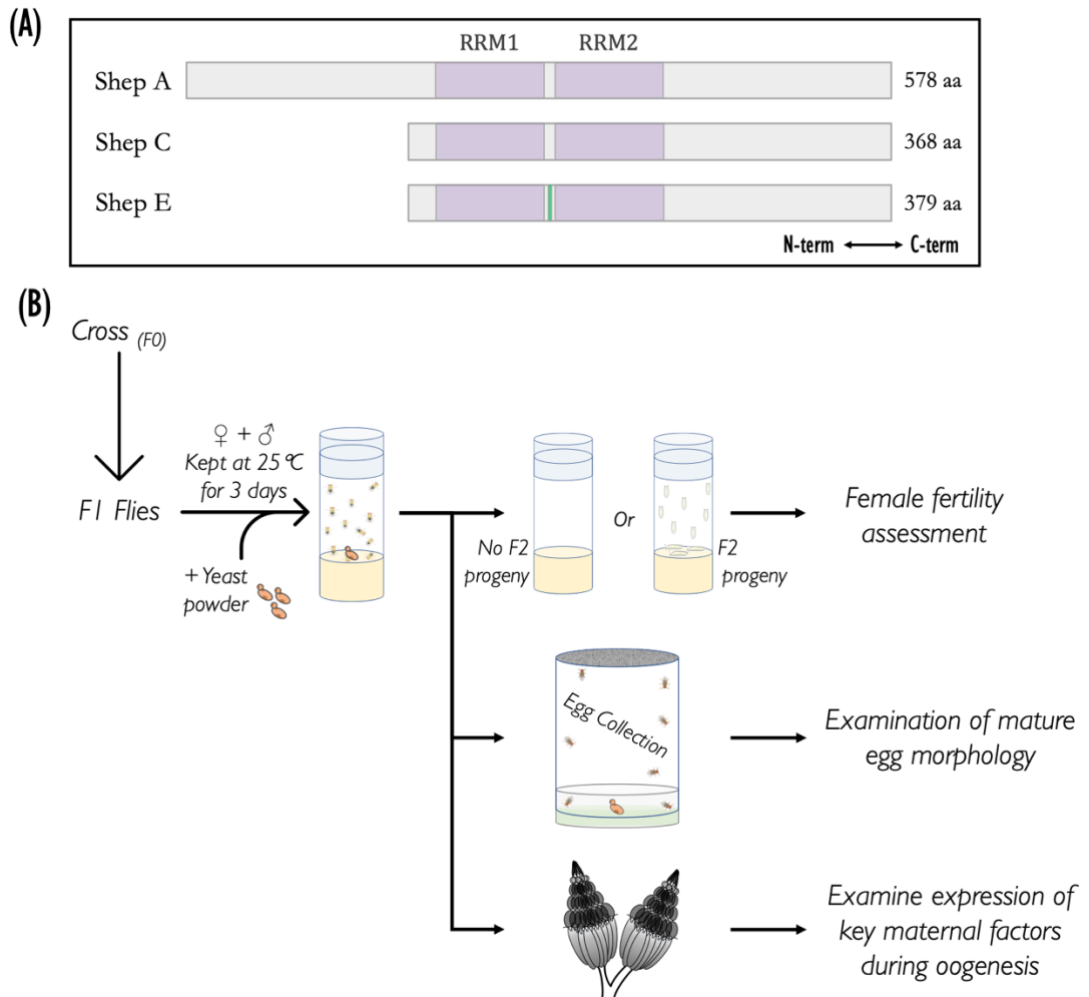


Figure 4.11: Schematic of the Shep germline overexpression analysis workflow. (A) Schematic representation of the three *Drosophila* Shep protein isoforms used in this analysis. The protein is depicted in gray and the RNA recognition motifs (RRM1 and RRM2) are shown in purple. Regions of alternative amino acid sequences between RRM1 and RRM2 are shown in green. The diagram is not drawn to scale. **(B)** To overexpress the different Shep isoforms in the female germline, a single copy of a germline-specific Gal4 driver was crossed with the UASp mCherry-tagged Shep constructs. This will cause some of the resulting F1 progeny, which possesses both the Gal4 and UAS constructs, to overexpress Shep in the female germline. Post-eclosion, F1 adult females were maintained at 25°C in the presence of males and yeast for three consecutive days. The females were then subjected to various analyses to assess their fertility, inspect the morphology of mature eggs, and investigate gene expression during oogenesis.

4.3.1 Females Overexpressing Shep in the Germline are Sterile.

To determine whether F1 females overexpressing Shep were sterile, we performed a fertility assay. Young, sexually mature female flies were kept at 25°C with male flies and dry yeast for three consecutive days, after which the adults were transferred into a new vial (see Section 2.5.5 of the ‘Material & Methods’ Chapter). The F1 females were then allowed to continue mating and laying eggs in the same vial for eight days at 25°C (**Figure**

4.12A), before being removed from the vials. The offspring were allowed to continue developing to adulthood at 25°C (i.e., giving rise to the F2 generation). During this period, we closely monitored the vials to ensure that the food did not dry out. After nine days of development, vials now containing the F2 generation were examined. Interestingly, despite the F1 mated females depositing eggs, we observed no developing offspring in vials with female flies overexpressing any of the Shep isoforms, compared with both control groups (Figure 4.12B). These results demonstrate that the overexpression of various Shep isoforms, specifically in the female germline, leads to complete sterility.

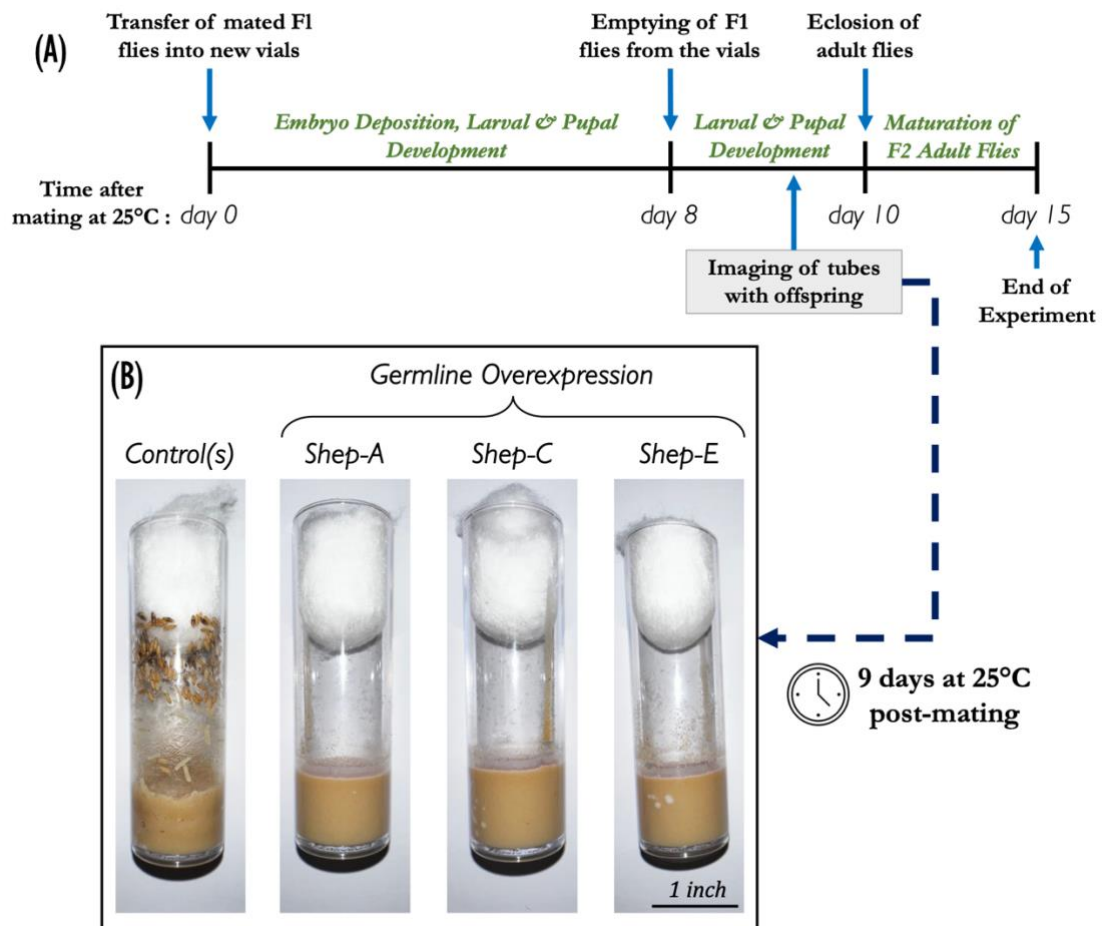


Figure 4.12: Germline overexpression of Shep causes female sterility. (A) Timeline for the experimental workflow of the *Drosophila* fertility assay. A germline-specific Gal4 driver was crossed with UASp constructs to overexpress Shep in the germline (F0). Adult F1 females derived from F0 crosses were isolated and mated with males in the presence of yeast at 25°C. After the initial mating and acclimatization period, the F1 females were transferred into new vials for embryo deposition (day 0). The offspring of the F1 females (i.e., the F2 generation) were allowed to develop for 15 days. On day 8, the F1 flies were discarded from the vial containing the developing F2 progeny. The vials were imaged on day 9, and the adult F2 offspring were counted on day 15. **(B)**

Outcome of the fertility assay. Images of vials captured on the 9th day illustrate the contrast between the fertile controls and the sterility of females overexpressing Shep in the germline, as evidenced by the presence or absence of developing progeny. Scale bar, 1 inch. Fly genotype(s) – Panel B shows vials from eGFP overexpression control (w^- ; *osk*-Gal4:VP16/+; UASp-eGFP/+), as well as Shep-A/-C/-E overexpressions (w^- ; UASp-mCh:ShepA/+; *osk*-Gal4:VP16/+), (w^- ; UASp-mCh:ShepC/+; *osk*-Gal4:VP16/+), and (w^- ; UASp-mCh:ShepE/+; *osk*-Gal4:VP16/+).

4.3.2 Overexpression of Shep in the Germline Disrupts Eggshell Patterning.

To investigate the reason behind the sterility observed in flies overexpressing Shep, we examined the eggshell of mature eggs laid by these females. Development of the chorionic dorsal appendages, which are the most prominent eggshell structures synthesized by two populations of dorsal-anterior follicle cells during late oogenesis, is often used as a readout of proper body axis patterning that is established throughout oogenesis (Stein and Stevens 2014; Osterfield *et al.* 2017). For this analysis, F1 females were prepared as previously described for the fertility assay (Section 4.3.1). However, after three consecutive days of mating, we transferred the mated females into a fly cage apparatus to collect mature eggs. Following an acclimatization period, we examined the eggshells, specifically the dorsal appendages, of the collected mature eggs under a microscope. For controls, the germline-specific driver was crossed with w^{1118} control flies, which do not overexpress any specific transgene, or with a UASp construct overexpressing plain eGFP.

Similar to wild-type flies, both control groups produced normal dorsal appendages, where a pair of chorionic structures were well separated and symmetrically positioned anteriorly on either side of the dorsal midline (**Figure 4.13A**). This observation indicates that these structures underwent proper tube formation, elongation, and maturation, ultimately forming three-dimensional, paddle-shaped appendages. In contrast to this, regardless of the specific Shep isoform overexpressed in the germline, the development of the dorsal appendages was compromised in 100% of the investigated mature eggs (**Figure 4.13B–D**). Strikingly, overexpression of the various Shep isoforms resulted in different phenotypes, unique to the specific overexpressed isoform. Germline overexpression of Shep-E exhibited the mildest phenotype, followed by Shep-A with a slightly more severe phenotype, while Shep-C displayed the most extreme defects. Overexpression of Shep-A specifically resulted in a distinct phenotype by the formation

of a single, centrally positioned, and shortened appendage (**Figure 4.13B**). This unique phenotype, which to our knowledge has not been previously described in the literature, bears a notable resemblance to a professional boxing glove, as demonstrated in panels B–Biii of **Figure 4.13**. Rather than undergoing elongation and maturation, cells appeared to accumulate at the base of the appendage, giving rise to the phenotype observed.

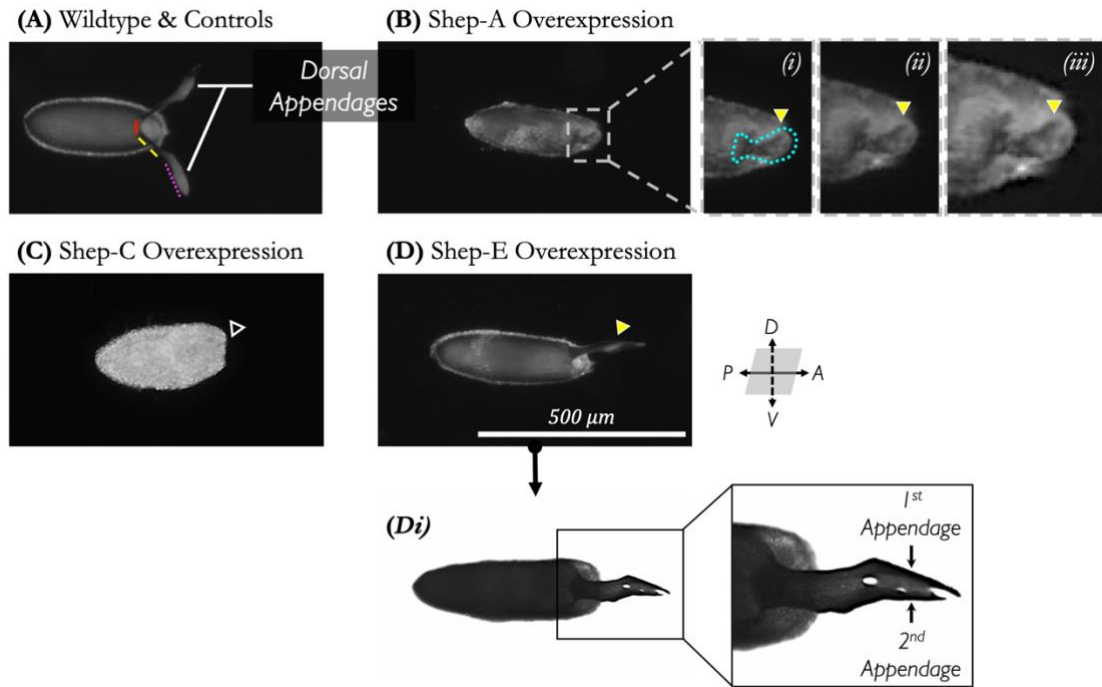


Figure 4.13: Germline overexpression of Shep disrupts the formation of mature eggs' dorsal appendages. (A) A mature wild-type or control egg shows two dorsal appendages, each characterized by a stalk (indicated by the yellow dashed-line) and a paddle (shown by the magenta dotted-line). These chorionic structures are distinctly separated (as indicated by the red solid line), located on the dorsal-anterior side of the eggshell, and positioned laterally on either side of the midline. (B–D) Germline overexpression of three Shep isoforms: A, C, and E. Three distinct phenotypes arise from Shep overexpression, each unique to the specific isoform overexpressed. Solid yellow triangles point towards the existing dorsal appendages, and white outlined triangles indicate the presumptive locations of the dorsal appendages. (B) Overexpression of Shep-A in the germline results in a single, short dorsal appendage (outlined in cyan dotted-line). Panels labelled with Roman numerals (i, ii, and iii) display higher magnification views of the dorsal appendages from the corresponding highlighted region. (C) Dorsal appendages are absent in mature eggs laid by females overexpressing Shep C in the germline. (D) Germline overexpression of Shep-E resulted in fused dorsal appendages. Panel (Di) shows bright-field illumination of a mature egg exhibiting fused dorsal appendages, which was captured from a slightly different angle to clearly demonstrate this fusion. Scale bar, 500 μm . Fly genotype(s) – Panel A (*w⁻; osk-Gal4:VP16/+; UASp-eGFP/+*), B (*w⁻; UASp-mCh:ShepA/+; osk-Gal4:VP16/+*), C (*w⁻; UASp-mCh:ShepC/+; osk-Gal4:VP16/+*), and D (*w⁻; UASp-mCh:ShepE/+; osk-Gal4:VP16/+*).

In contrast, germline overexpression of Shep-C resulted in a more severe phenotype, whereby no distinguishable appendages were developed. Additionally, the eggshell appearance notably differed from that of both the other experimental and control groups (**Figure 4.13C**). Furthermore, the germline overexpression of Shep-E gave rise to a fused dorsal appendage, as depicted in panels D–Di of **Figure 4.13**. Contrary to the effects of overexpressing Shep-A, these dorsal appendages appear to have undergone elongation without reaching proper maturation, as evidenced by the length of the stalk and the shape of the paddles (i.e., the extremities of each appendage).

Overall, our results demonstrate that overexpression of various Shep isoforms specifically in the female germline disrupts axis patterning, which is reflected in the morphogenesis of the dorsal appendage. Here, these phenotypes provide additional evidence for Shep’s cell nonautonomous effects, where its overexpression in the germline influences the adjacent somatic follicle cells that give rise to the dorsal appendages. Given the disruption observed in the formation of dorsal appendages, together with Shep’s subcellular localization at the oocyte’s dorsal-anterior cortex (**Figure 3.5B**), it is possible that Shep is involved in the regulation of the Gurken signalling pathway.

4.3.3 Germline Overexpression of Shep Disrupts the Expression of Key Maternal mRNAs

4.3.3.1 Overexpression of Shep Disrupts Gurken Protein Levels in the Oocyte

Disruption of the dorsal appendages (see a detailed description of this process in Section 5.2.2) suggests that Shep overexpression in the germline affects the Gurken signalling pathway. Based on the defects we observed, the Shep protein subcellular localization at the dorsal-anterior corner of the oocyte, and the corresponding literature, we hypothesized that *gurken* is downregulated when Shep is overexpressed. To test this hypothesis, we performed an immunostaining experiment to visualize Gurken protein, complemented by qPCR for the quantification of *gurken* mRNA. Briefly, flies carrying the *oskar*-Gal4:VP16 driver were mated with both wild-type control flies (w^{1118}) and those harbouring mCherry-tagged Shep constructs, with the aim of overexpressing Shep in

the germline. F1 adult females from both the control groups and those overexpressing Shep in the germline were isolated and prepared for dissection and staining, as described in Sections 2.5.4 and 2.5.10 of the ‘Materials & Methods’ Chapter.

Based on our previous characterization of Shep protein subcellular distribution pattern (see Section 3.4.3 of Chapter 3), we anticipated that Shep would colocalize with Gurken protein. Indeed, our immunostaining experiment demonstrated that this is the case: Shep colocalizes with Gurken protein at the dorsal-anterior corner of late-stage oocytes (Figure 4.14), where Gurken signalling helps specifying the populations of cells needed to give rise to the dorsal appendages (Berg 2005).

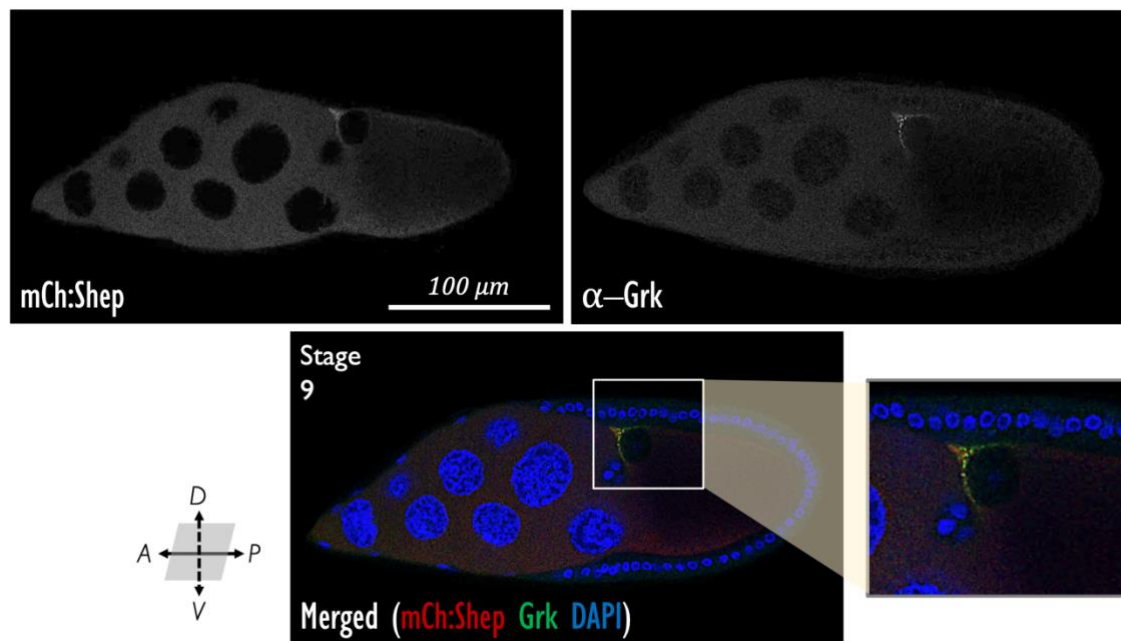


Figure 4.14: Shep colocalizes with Gurken protein at the dorso-anterior corner of late-stage oocytes. Fluorescent immunohistochemistry was used to visualize the Gurken protein (green) of stage-9 oocytes expressing mCherry-tagged Shep isoforms (red). Nuclei are marked by DAPI staining (blue). After the establishment of the anterior-posterior axis during mid-oogenesis, Gurken localizes to the dorsal-anterior corner of the oocyte. Different Shep isoforms (A, C, and E) colocalize with Gurken protein. The staining shown here is representative of all three isoforms. Scale bar, 100 μ m. Fly genotype(s) – The displayed egg chamber is representative of Shep-A/-C/-E germline overexpression (w^- ; UASp-mCh:ShepA/+; *osk*-Gal4:VP16/+), (w^- ; UASp-mCh:ShepC/+; *osk*-Gal4:VP16/+), and (w^- ; UASp-mCh:ShepE/+; *osk*-Gal4:VP16/+).

Next, we systematically examined the effects of overexpressing Shep in the female germline on the expression pattern of Gurken protein during oogenesis. In the wild-type control, Gurken protein is expressed predominantly in the germline syncytium

from the early stages of oogenesis, with enrichment observed in the oocyte (Figure 4.15A) until the anterior-posterior axis is established during mid-oogenesis. In *D. melanogaster* oogenesis, the Gurken signalling pathway plays a key role in the establishment of both the anterior-posterior and dorsal-ventral posterior axes during mid- and late oogenesis, respectively (Milas and Telley 2022). After the establishment of the anterior-posterior axis during mid-oogenesis, Gurken protein expression becomes restricted to the dorsal-anterior corner of oocytes (see Figure 4.15C). This localization pattern is key because it specifies the dorsal fate of the eggshell and body axis of the future embryo (Roth and Lynch 2009).

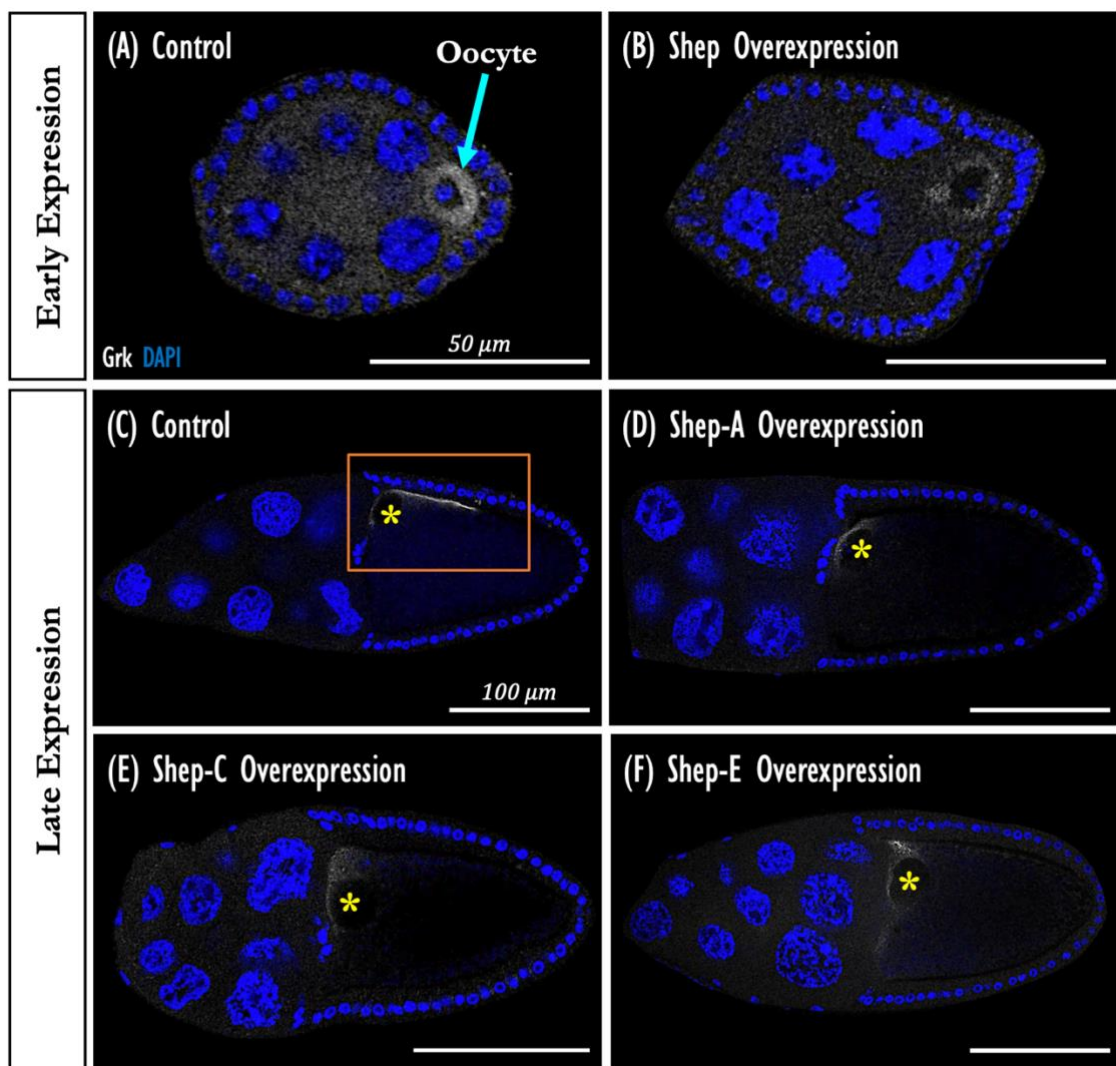


Figure 4.15: Germline overexpression of Shep disrupts Gurken expression. Visualization of the Gurken protein in stage-10 oocytes using fluorescent immunohistochemistry. **(A & B)** During the early stages of oogenesis, within the germline syncytium, Gurken protein is enriched in the oocyte towards the posterior in control flies or in those overexpressing Shep. Here, the the first round of Gurken signalling induces the establishment of the anterior-posterior axis. **(C)** After this

first body axis is established during mid-oogenesis, Gurken localizes to the dorsal-anterior corner of stage 10 oocytes, as indicated by the orange box. Here, the second round of Gurken signalling occurs, establishing the dorsal-anterior axis. However, the overexpression of Shep disrupts Gurken protein expression and distribution patterns. Asterisks indicate the location of the oocyte nucleus. Scale bar represents 50 μm for early-stage oocytes and 100 μm for late-stage oocytes. Fly genotype(s) – Panels A & C (*w⁻/w¹¹⁸*; *osk-Gal4:VP16/+*), D (*w⁻*; UASp-mCh:ShepA/+; *osk-Gal4:VP16/+*), E (*w⁻*; UASp-mCh:ShepC/+; *osk-Gal4:VP16/+*), and F (*w⁻*; UASp-mCh:ShepE/+; *osk-Gal4:VP16/+*). Panel B represents Shep-A/-C/-E germline overexpression.

When the various isoforms of Shep were overexpressed in the germline using the *osk-Gal4:VP16* driver, the localization of Gurken to the oocyte was not affected (**Figure 4.15B**). Even though the protein levels at these early stages appear to be slightly reduced compared with the controls, this reduction does not seem to disrupt the first round of Gurken signalling necessary for establishing the anterior-posterior axis. This is indicated by the proper positioning of the oocyte's nucleus at the dorsal-anterior corner, as shown in the late-stage egg chambers in **Figure 4.15** (compare panels C-F).

At later stages of oogenesis, the expression pattern of Gurken protein in oocytes overexpressing Shep deviates from the pattern observed in the control ovaries (compare panel C to D-F of **Figure 4.15**). Gurken protein expression becomes constricted along the dorsal midline, primarily accumulating in close proximity to the oocyte nucleus, rather than extending towards the posterior end (**Figure 4.15D-F**). Occasionally, it also extends deeper towards the ventral side, a pattern not observed in any of the control or wild-type ovaries. The disruption in the distribution of Gurken was relatively similar across the different overexpression isoforms of Shep. Interestingly, the disruption we observed here in the Gurken localization pattern does not seem to correlate with the severity observed in the dorsal appendages analysis.

Altogether, our results demonstrate that overexpression of Shep in the female germline disrupts Gurken protein expression at the dorsal-anterior corner of late-stage oocytes. This disruption ultimately leads to the defects observed in the formation of the dorsal appendages, as detailed in Section 5.2.2.

Given the apparent reduction in Gurken protein in ovaries overexpressing Shep (**Figure 4.15D-F**), we subsequently proceeded to quantify *gurken* mRNA levels. To assess

whether *gurken* mRNA levels decreased with germline overexpression of Shep, we performed reverse-transcription qPCR, as outlined in Section 2.1.3.4 of the ‘Materials & Methods’ Chapter. Briefly, we extracted RNA from both control and Shep overexpression ovaries. This RNA was then subjected to DNase I treatment to remove gDNA. Subsequently, the treated RNA samples were converted into cDNA templates for qPCR reactions. Interestingly, our results indicate that there is no significant difference between the overexpression groups and the control (**Figure 4.16**), as determined by the one-way ANOVA ($p= 0.7095$) with Tukey’s HSD test for multiple comparisons. These findings suggest that the observed reduction in Gurken protein levels, as revealed by immunostaining, may be a consequence of defects in the translation of *gurken* mRNA, or an improper anchoring of either *gurken* mRNA or Gurken protein at the antero-dorsal corner of the oocyte. Additional tests are required to ascertain which of these scenarios is occurring.

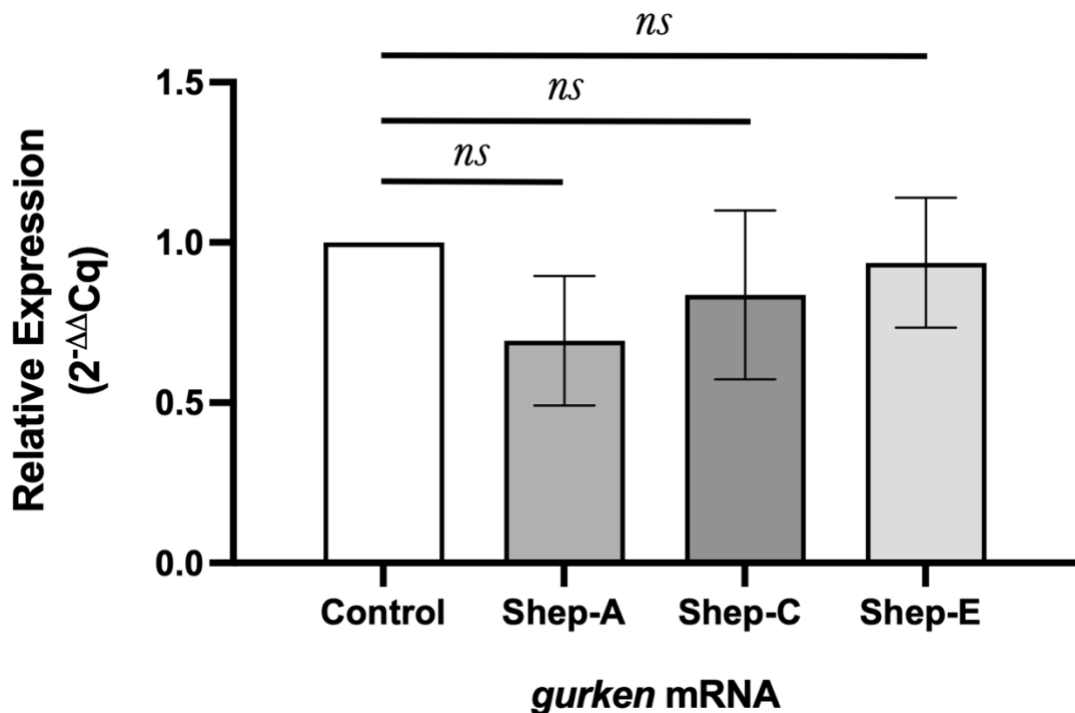


Figure 4.16: *gurken* mRNA levels in ovaries overexpressing Shep protein in the germline. Comparison of *gurken* mRNA expression values obtained by RT-qPCR suggests that overexpression of different Shep protein isoforms (namely, -A, -C, and -E) in the female germline has no significant effect on the relative expression of *gurken* mRNA levels between experimental groups and control. The mRNA of the *ribosomal protein L32* (also referred to as *rp49*) was used as a reference gene to normalize the relative expression of *oskar* mRNA within each sample. Three biological replicas were obtained, each with technical triplicates. Error bars represent mean \pm SEM. One-way ANOVA with Tukey’s multiple comparison test was performed between all group

combinations (ns: not significant, $p > 0.05$) using GraphPad. Fly genotype(s) – cDNA library for RT-qPCR obtained from the ovaries of control females ($w^-/w^{1118}; ; osk-Gal4:VP16/+$) and those overexpressing Shep-A/-C/-E in the germline ($w^-; UASp-mCh:ShepA/+; osk-Gal4:VP16/+$), ($w^-; UASp-mCh:ShepC/+; osk-Gal4:VP16/+$), and ($w^-; UASp-mCh:ShepE/+; osk-Gal4:VP16/+$).

4.3.3.2 Germline Overexpression of Shep Leads to a Reduction in Oskar Protein Levels

Given the ectopic premature translation of *oskar* observed in the *shep* loss-of-function mutants (see Section 4.2), we aimed to determine if the overexpression of Shep would have the opposite effect, i.e., a re-enforcement of *oskar* mRNA translational repression.

To investigate the effects of Shep overexpression in the germline on *oskar* mRNA and its protein expression during *D. melanogaster* oogenesis, we performed fluorescence *in situ* hybridization to visualize *oskar* mRNA, followed by immunostaining to detect the Oskar protein. Briefly, flies carrying the *oskar-Gal4:VP16* driver were mated with both wild-type control flies and those harbouring mCherry-tagged Shep constructs, aiming to overexpress Shep in the germline. Adult F1 females from both the control groups and those overexpressing Shep in the germline were isolated and prepared for dissection and staining, as described in Sections 2.5.4 and 2.5.10 of the ‘Materials & Methods’ Chapter.

Following the dissection and staining of these ovaries, we examined their egg chambers using confocal microscopy to visualize the expression patterns of both the *oskar mRNA* and Oskar protein. Although both the accumulation of *oskar* mRNA in these oocytes during the early stages of oogenesis (data not shown) and its transport to the posterior pole at mid-oogenesis seem unaffected, its expression and anchoring at the posterior pole were impacted. Our results revealed that the overexpression of any Shep isoform disrupted *oskar* expression at the posterior pole of stage 10 oocytes (**Figure 4.17**). The disruptions observed because of Shep overexpression appear to affect *oskar* expression at both the mRNA and protein levels, as judged by the intensity of their immunofluorescent signals (**Figure 4.17**).

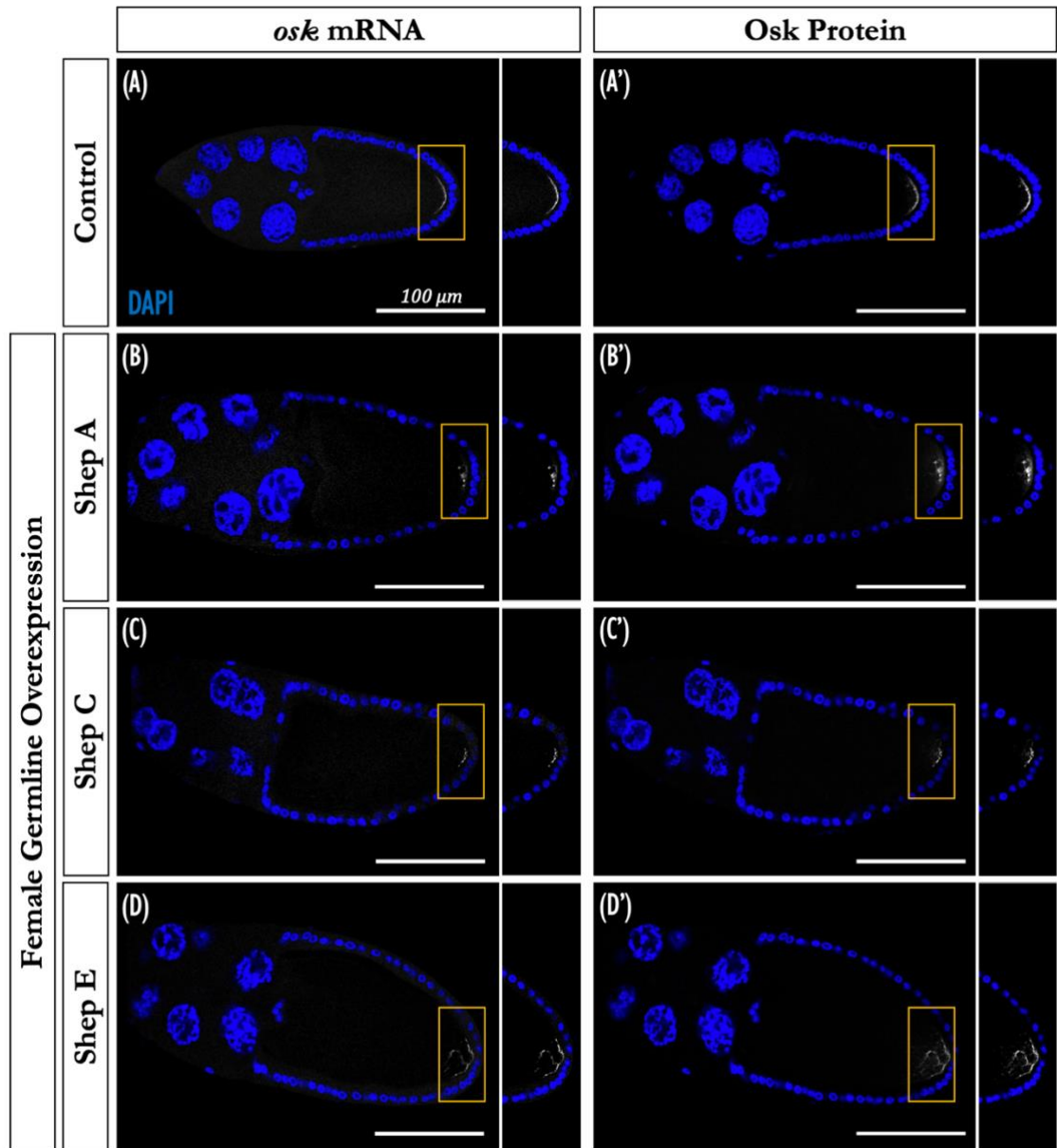


Figure 4.17: Germline overexpression of Shep disrupts *oskar* expression. Visualization of *oskar* mRNA and Oskar protein in stage-10 oocytes using fluorescence *in situ* hybridization coupled to immunohistochemistry. **(A)** In the control, both *oskar* mRNA and its protein localise to the oocyte's posterior pole by stage-10 of oogenesis. **(B–D)** Shep overexpression appears to interfere with the expression and anchoring of both *oskar* mRNA and Oskar protein at the posterior pole. In all panels, the orange box indicates the magnified posterior portion of the oocyte. Scale bar, 100 μ m. Fly genotype(s) – Panel A (*w⁻/w¹¹¹⁸*; *osk*-Gal4:VP16/+), B (*w⁻*; UASp-mCh:ShepA/+; *osk*-Gal4:VP16/+), C (*w⁻*; UASp-mCh:ShepC/+; *osk*-Gal4:VP16/+), and D (*w⁻*; UASp-mCh:ShepE/+; *osk*-Gal4:VP16/+).

Similar to the variation in severity observed in the dorsal appendages analysis (see Section 4.3.2), the phenotypic disruption to *oskar* expression resulting from germline overexpression of Shep displayed a similar trend: Shep-E showed the mildest phenotype, followed by Shep-A, with Shep-C exhibiting the most drastic defects. First,

oskar mRNA and protein levels were notably reduced due to the overexpression of various Shep isoforms, as shown in **Figure 4.17** (compare panels A and B–D). Second, the overexpression of Shep appears to potentially interfere with the anchoring of *oskar* mRNA and Oskar protein at the posterior pole, as they both appear dissociated from the oocyte’s posterior cortex. In the examples given, this is particularly evident in the overexpression of Shep-E (**Figure 4.17D**). However, this phenotype is also evident in the Shep-C and -A overexpression samples. Moreover, the characteristic crescent pattern of the *oskar* at the posterior of stage-10 oocytes also appears narrower sometimes when compared with the control groups, as illustrated here with Shep C (**Figure 4.17C**). Altogether, these results indicate that Shep is indeed involved in the regulation of *oskar* expression, possibly as a regulator of mRNA stability and/or translation, which is consistent with our previous analyses and hypotheses.

Next, we decided to validate the observations obtained in our immunostaining analysis to confirm the reduction of *oskar* at both the mRNA and protein levels. This was achieved by conducting both Western blot and RT-qPCR analyses.

Remarkably, our immunoblotting data revealed that the overexpression of Shep in the germline led to a reduction in the Oskar protein levels (**Figure 4.18**). The amount of Oskar protein in these ovaries mirrored the phenotypic severity previously described in the immunostaining and the dorsal appendage analyses (see Section 5.2.2). In fact, a reduction is observed in both the long and short isoforms of the Oskar protein, with the short isoform being more severely affected (**Figure 4.18**). The spatial localization of these maternal determinants to the posterior pole of oocytes is critical for posterior patterning and pole plasm assembly in *D. melanogaster*. However, these isoforms of Oskar, synthesized at the oocyte’s posterior by alternative translation initiation from two in-frame start codons within *oskar* mRNA, have been demonstrated to play distinct biological roles (Markussen *et al.* 1995; Jeske *et al.* 2015). Given that the long Oskar isoform is essential for the proper anchoring of both *oskar* mRNA and Short Oskar to the posterior cortex of the oocyte (Vanzo and Ephrussi 2002), the observed reduction in long Oskar protein could justify the *oskar* anchoring phenotype observed in our immunostaining results (see **Figure 4.17D**). Our findings suggest that overexpression of

Shep downregulates both isoforms of the Oskar protein, indicating a potential regulatory role for Shep within the *oskar* mRNP complex. In addition, the reduction differences in Oskar protein levels observed with various Shep overexpressions may suggest that the regulation of *oskar* mRNA by Shep could be isoform-specific.

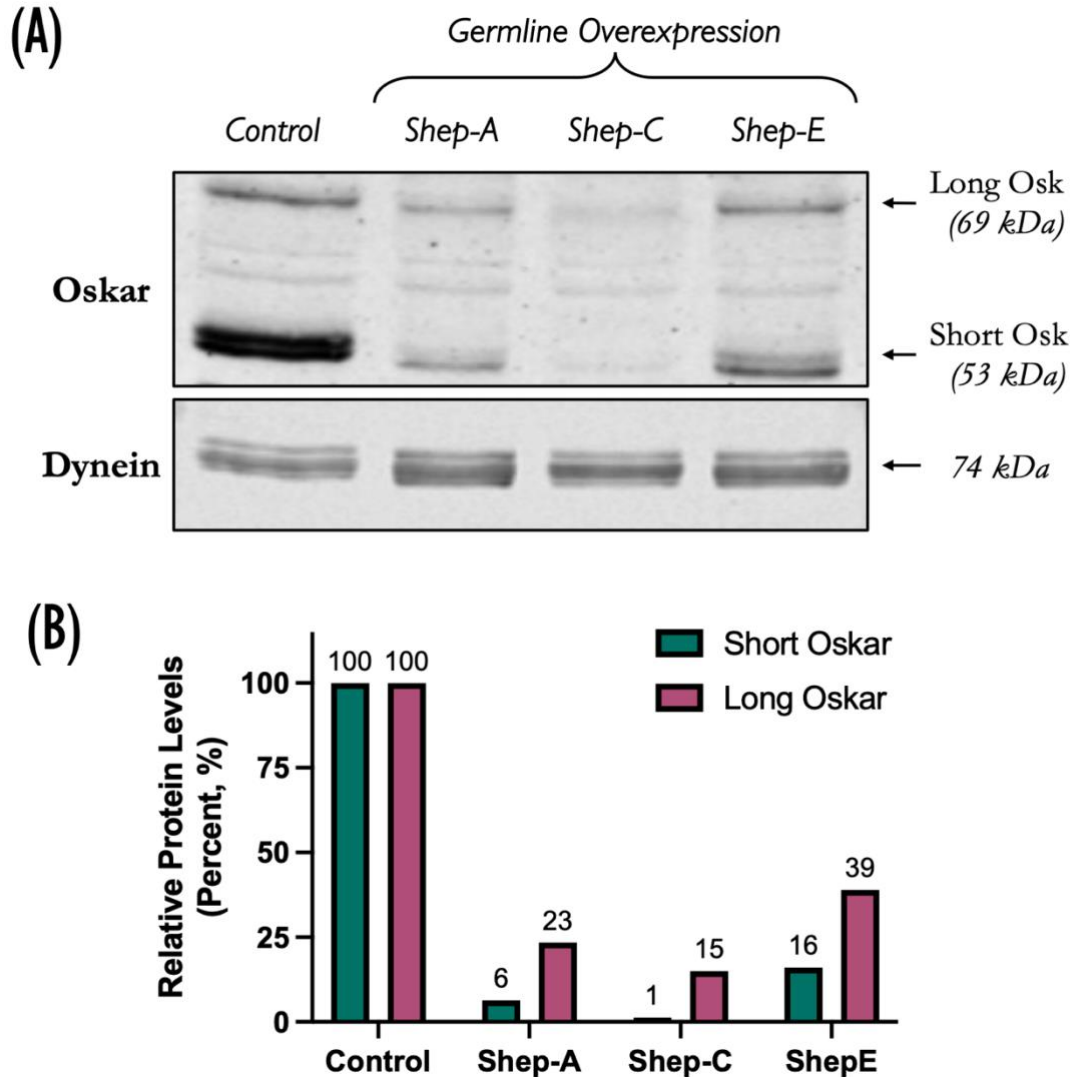


Figure 4.18: Germline overexpression of Shep reduces Oskar protein levels. (A) Western blot analysis of Oskar protein from crude extracts of control ovaries or those overexpressing Shep isoforms (A, C, or E). Two isoforms of Oskar exist: the Short and Long Oskar proteins (53 kDa and 69 kDa, respectively). In wild-type ovaries, Short Oskar is the most abundant isoform of the two polypeptides and appears as a duplet because of post-translational modifications (Riechmann *et al.* 2002). Overexpression of Shep in the germline reduces Oskar protein expression levels. The motor protein dynein (74 kDa) was used as a loading control. The ‘Precision Plus Protein Dual Color Standards’ (BioRad) was used as a protein ladder. (B) Quantification of Oskar protein expression. The relative expression levels are determined from the band intensity on membranes, normalized against Dynein. Fly genotype(s) – Protein extracts obtained from the ovaries of control females (w^{1118}/w^- ; $osk-Gal4:VP16/+$) and those overexpressing Shep-A/-C/-E in the germline (w^- ; UASp-mCh:ShepA/+; $osk-Gal4:VP16/+$), (w^- ; UASp-mCh:ShepC/+; $osk-Gal4:VP16/+$), and (w^- ; UASp-mCh:ShepE/+; $osk-Gal4:VP16/+$).

Having addressed the fact that Shep overexpression reduces Oskar protein levels, we next turned our attention to what is happening at the mRNA level. As previously described and detected by *in situ* hybridisation (**Figure 4.17**), *oskar* mRNA levels also appeared to be reduced upon Shep overexpression. Therefore, to determine whether Shep could regulate the stability or translation of *oskar* mRNA, we opted to quantify the levels of *oskar* mRNA using RT-qPCR. Here, we followed a similar experimental procedure to that previously described for *gurken* mRNA (as outlined in Section 2.1.3.4 of the ‘Materials & Methods’ Chapter).

From three biological replicates, each with technical triplicates, our qPCR results showed that there is no significant difference between the overexpression groups and the control (see **Figure 4.19**), as determined by the one-way ANOVA ($p= 0.9343$) with Tukey’s HSD test for multiple comparisons. Despite the variability in the data, these findings suggest that the reduction in Oskar protein levels upon Shep overexpression, which we previously detected by Western blotting (see **Figure 4.18**), may be a consequence of having less protein translated rather than less mRNA. Furthermore, a possible explanation for the discrepancy observed between the confocal analysis and the RT-qPCR results in terms of *oskar* mRNA levels might be the improper anchoring at the posterior pole. Thereby, causing the mRNA to drift away and subsequently diffuse into the oocyte by cytoplasmic streaming. Overall, our findings suggest that Shep overexpression downregulates Oskar protein levels through translational control.

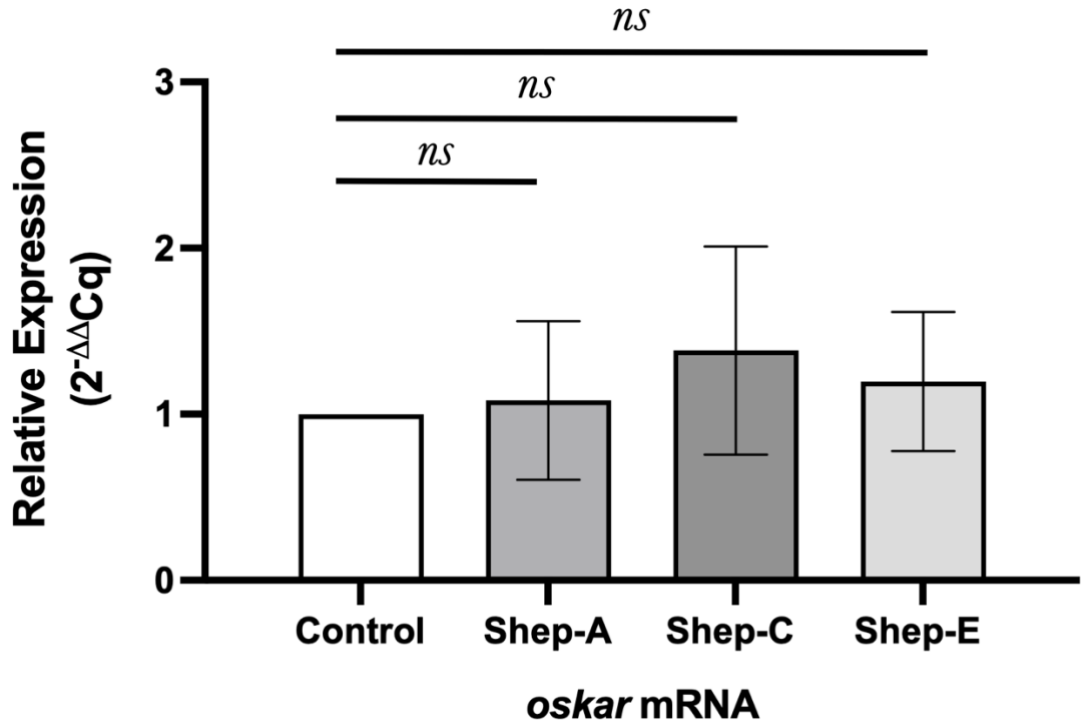


Figure 4.19: *oskar* mRNA levels in ovaries overexpressing Shep in the germline. Comparison of *oskar* mRNA relative expression values obtained by RT-qPCR suggests that overexpression of different Shep protein isoforms (namely, -A, -C, and -E) in the female germline has no significant effect on the relative expression of *oskar* mRNA levels between experimental groups and control. The mRNA of the *ribosomal protein L32* (also referred to as *rp49*) was used as a reference gene to normalize the relative expression of *oskar* mRNA within each sample. Three biological replicates were obtained, each in technical triplicates. Error bars represent mean \pm SEM. One-way ANOVA with Tukey's multiple comparison test was performed between all group combinations (ns: not significant, $p > 0.05$) using GraphPad. Fly genotype(s) – cDNA library for RT-qPCR obtained from the ovaries of control females (*w*⁻/*w*¹¹¹⁸; ; *osk*-Gal4:VP16/+) and those overexpressing Shep-A/-C/-E in the germline (*w*⁻; UASp-mCh:ShepA/+; *osk*-Gal4:VP16/+), (*w*⁻; UASp-mCh:ShepC/+; *osk*-Gal4:VP16/+), and (*w*⁻; UASp-mCh:ShepE/+; *osk*-Gal4:VP16/+).

4.4 Conclusions

In conclusion, the results presented here offer novel insights into the role of Shep during *D. melanogaster* oogenesis, especially its involvement in the post-transcriptional regulation of key maternally loaded mRNAs: *oskar* and *gurken*. This is particularly interesting, given that the role of Shep in oogenesis had not been elucidated before.

In our *shep* loss-of-function analysis, we revealed dysregulation in the *oskar* expression pattern, where Oskar protein was ectopically and prematurely translated in the oocyte's cytoplasm before reaching the posterior cortex. Additionally, when we downregulated *shep* by inducing the expression of RNAi hairpins in the female germline, egg chambers

displayed an aberrant organization. Specifically, the 16-cell syncytium was no longer fully encapsulated by a monolayer of follicular epithelium. Interestingly, this phenotype resembled the phenotype observed upon downregulating the Delta-Notch signalling pathway (detailed in Section 3.5.1 of Chapter 3), pointing to a possible link between Shep and this signalling pathway. Furthermore, RNAi-mediated depletion of *shep* levels in the germline affected neighbouring somatic cells in a cell nonautonomous manner: it partially disrupted the migration of border cells and downregulated Lamin C expression in stalk cells.

Overexpression of Shep in the germline leads to partial arrest of oogenesis and sterility in female flies. Among other factors, sterility can be attributed to the disruption in the formation of mature eggs' dorsal appendages, which serve as breathing tubes for the embryo before hatching (Berg 2005). This provides further evidence for a cell nonautonomous role of Shep during oogenesis, whereby altering its levels in the germline affects neighbouring somatic cells. Upon further investigation of *gurken* expression, it became apparent that overexpressing Shep in the germline dysregulates Gurken protein expression or localization at the dorsal-anterior corner of stage-10 oocytes, ultimately leading to disruption of the dorsal appendages. In conjunction with overexpressing Shep in the germline and by employing a combination of techniques such as *in situ* hybridization, immunostaining, western blotting, and RT-qPCR, our results alluded to a post-transcriptional regulatory role for Shep in controlling the translation of *oskar* mRNA at the posterior pole of the oocyte.

❧ Chapter V ❧

Characterizing the Composition of
the Ovarian Shep mRNP
Complexes

5 Chapter V – Characterizing the Composition of the Ovarian Shep mRNA Complexes:

The findings presented in the previous chapters indicate that Shep is a novel regulatory component of both the *oskar* and *gurken* mRNA complexes. In addition, throughout our characterization of Shep, we observed that some of Shep subcellular distribution patterns and phenotypes bear notable resemblance to other genes either presented in our work or reported in the existing literature. For example, the amorphic egg chamber phenotype observed when downregulating *shep* levels in the germline resembles the phenotype observed following the disruption of Delta-Notch signalling pathway (Compare **Figures 3.10-11 & 4.4**). Therefore, our analysis thus far hinted at potential associations between the cytoplasmic Shep protein and several candidates, including *Delta*, *Notch*, and *shotgun* (*Drosophila* DE-Cadherin), *bicoid*, and *nanos*. Moreover, Shep was identified as a potential binding partner of the RNA-binding protein PTB in a yeast two-hybrid screen (López de Quinto, unpublished). This finding becomes particularly interesting given that both Shep and PTB post-transcriptionally regulate the expression of *oskar* and *gurken* mRNAs in the female germline (see Chapter 4; Besse *et al.* 2009; McDermott and Davis 2013). In light of our previous findings and the aforementioned parallelisms between Shep and PTB, we decided to molecularly characterize the mRNA complexes with which Shep associates during oogenesis and to further investigate whether the RNA-binding proteins Shep and PTB physically interact with each other.

Aims
7) Analyse the sequence of target and candidate mRNA molecules to identify potential Shep-binding motifs.
8) Map <i>in vitro</i> the binding region of Shep within key maternal mRNAs.
9) Confirm the <i>in vivo</i> interaction of Shep with ovarian mRNA complexes.
10) Investigate the physical interactions between Shep and PTB <i>in vivo</i> .

5.1 Characterization of Interaction Between Shep and its mRNA Targets

In this study, we employed a sequence-based computational analysis to identify potential Shep mRNA targets and their Shep-binding motif. We also used an *in vitro* RNA-centric assay to map the Shep-binding interaction within these mRNA targets.

5.1.1 Computational Prediction of Shep-Binding Sites within Candidate mRNAs

We began our analysis by searching for potential Shep-binding motifs in the primary mRNA sequence of transcripts suspected to interact with Shep. Although the consensus sequence recognized by Shep within its RNA targets has yet to be experimentally validated, Ray *et al.* (2013) reported three 7-mer RNA-binding motifs for *Drosophila* Shep using RNAcompete: ‘WAUWUWD’, ‘WUAUWWA’, and ‘AUAUUWD’ (see **Figure 5.1A**). Based solely on these three motifs, Shep appears to preferentially interact with AU-rich sequences. RNAcompete is conceptually similar to the traditional motif identification approach SELEX (Gerstberger *et al.* 2013; Cook *et al.* 2015; Ray *et al.* 2017). While both methods involve incubating RNA-binding proteins with an *in vitro* transcribed pool of RNA molecules, to then conduct RBP pulldowns followed by RNA recovery, RNAcompete and SELEX mainly differ in the final RNA identification step: RNAcompete uses microarray hybridization whereas SELEX relies on sequencing.

To analyze the distribution of the three motifs within each mRNA of interest, we retrieved sequences corresponding to the coding sequence and untranslated regions from FlyBase for each potential Shep mRNA target and then searched for the three Shep-binding motifs using SnapGene. Subsequently, for each mRNA target, we summed up the total number of potential Shep RNA-binding sites to their respective regions (i.e., 5'UTR, CDS, and 3'UTR) and then normalized these counts to 1 kilobase (Kb). Our analysis focused on nine mRNAs critical for ovarian development: *oskar*, *bicoid*, *nanos*, *gurken*, *Notch*, *Delta*, *shotgun*, *female sterile (1) K10*, and *tsunagi* (also known as *Y14*). These mRNAs were either previously demonstrated or hypothesized by us to potentially interact with Shep. However, in the case of *Y14*, it was included in our analysis as a negative control for the *in vitro* RNA-pulldown assay (detailed in Section 5.1.2).

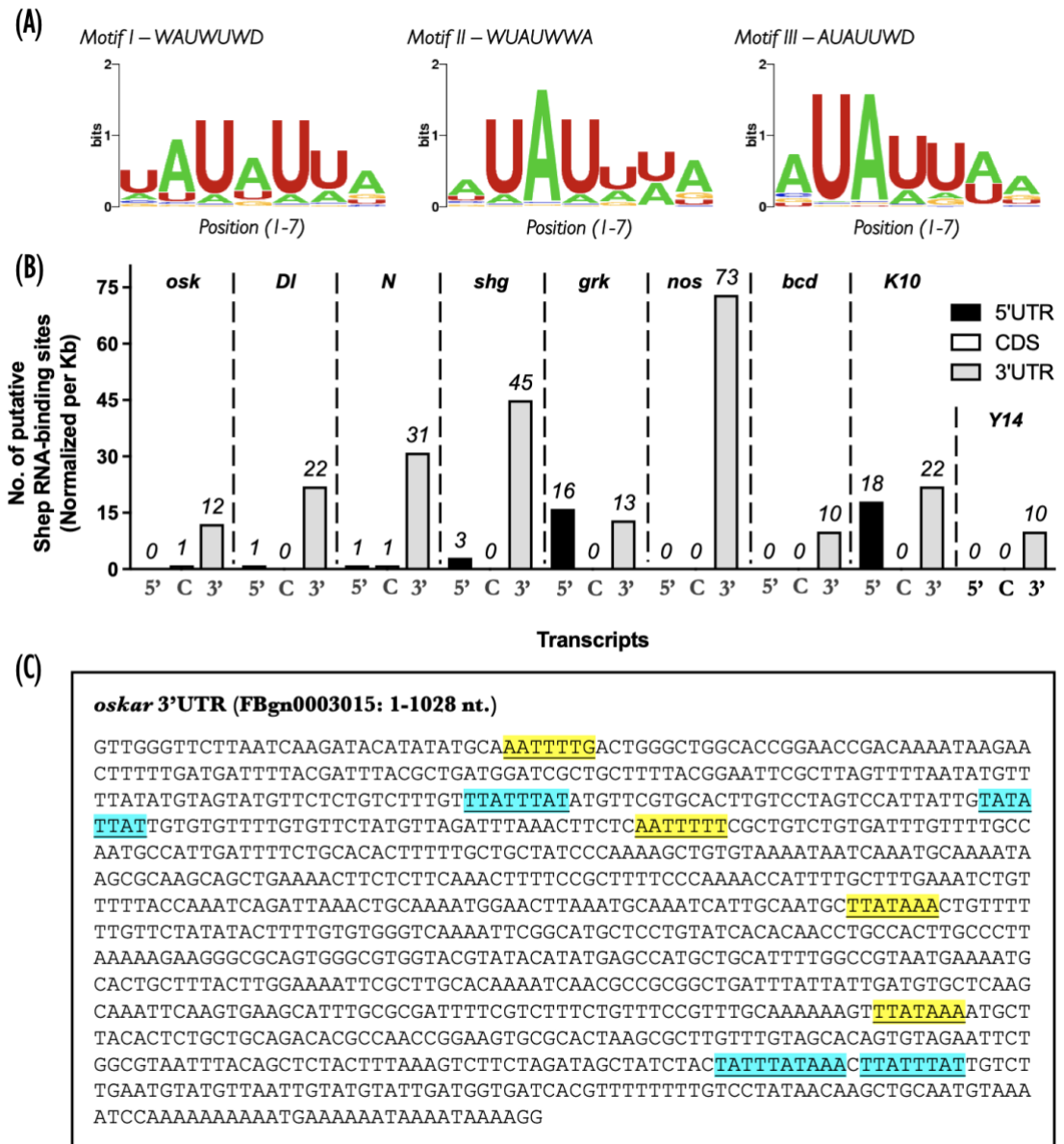


Figure 5.1: Computational characterization of Shep-binding sites within various mRNA targets. (A) Three 7-nucleotide consensus RNA-binding motifs of the *Drosophila* Shep protein: WAUWUWD, WUAUWWA, and AUAUUWD, as identified by the RNAcompete study by Ray *et al.* (2013). In these consensus sequences, the IUPAC nucleotide code is used: the letter ‘W’ represents either A or T bases, while ‘D’ stands for A, G, or T. The sequence logo, a graphical representation of the consensus, is depicted in 2-bits to convey information about the conservation and frequency of motif sequences at each position. (B) The graph shows the combined number of putative Shep RNA-binding sites identified in each transcript, based on the three RNAcompete motifs, normalized per kilobase (Kb). For each transcript, the mRNA is split into three distinct regions: the 5' untranslated region (5'UTR, 5'), coding sequence (CDS, C), and 3' untranslated region (3'UTR, 3'). For the exact values of the binding sites for each motif (i.e., raw data), refer to Appendix 5. Abbreviations – *osk*: *oskar*, *DI*: *Delta*, *N*: *Notch*, *shg*: *shotgun*, *grk*: *gurken*, *nos*: *nanos*, *bcd*: *bicoid*, *K10*: *female sterile (1) K10*, *Y14*: *tsunagi*. (C) The Shep RNA-binding motifs found within the *oskar* 3'UTR. Sequences underlined and highlighted in yellow represent individual motifs, while those in cyan indicate overlapping regions containing two motifs.

Irrespective of the mRNA investigated, our analysis of the normalized data indicated that the Shep-binding motifs were preferentially located within the 3'UTR (**Figure 5.1B**). Although we observed a pronounced enrichment of putative Shep RNA-binding sites within the 3'UTR sequence, a handful of potential sites were also identified outside this region, specifically within the 5'UTR of *Delta*, *Notch*, *shotgun*, *gurken*, and *K10*, as well as in the coding sequence of *oskar* and *Notch*. Notably, these potential Shep-binding sites were found to span the entire length of *Notch* mRNA only (**Figure 5.1B**). These observations indicate that while Shep interactions are more likely to occur within the 3'UTR, they could also extend beyond it, suggesting that Shep might influence various aspects of mRNA processing and regulation depending on the location of its interaction site.

As a final part of the same analysis, we also scrutinized the nucleotide base composition of each Shep putative binding site identified within the investigated mRNAs using the RNAcompete motifs (Ray *et al.* 2013). Upon examination, it became apparent from the sequences that these binding sites were indeed highly enriched in both adenine and uridine bases (see example in panel C of **Figure 5.1**). These observations strongly suggest that Shep likely functions as a *trans*-acting RNA-binding protein, binding to AU-rich elements within the 3'UTR, potentially regulating mRNA stability and translation. This is in fact consistent with the translational repression role we previously revealed for Shep in both our loss- and gain-of-function analyses (refer to Sections 4.2 and 4.3 of Chapter 4).

5.1.2 Biochemical Mapping of Shep Binding to mRNA Targets

Based on the computational analysis results showing numerous putative Shep-binding sites predominantly along 3'UTR regions, we hypothesized that Shep would preferentially bind the 3'UTR sequences of the potential mRNA targets. Therefore, to experimentally validate these interactions in the ovary and to identify any potential isoform-specific preferences for each mRNA, we performed an RNA-affinity pulldown assay. This *in vitro* RNA-centric biochemical assay involved the synthesis of biotinylated mRNA transcribed *in vitro* to selectively capture RNA-binding proteins from an ovarian cell lysate under native conditions.

To analyze the entire set of mRNAs studied in this project, we began by individually cloning the coding sequence and untranslated regions of both *gurken* and *shotgun* mRNAs into a pBlueScript II vector. Biotinylated sense mRNA probes were then synthesized *in vitro* for the coding sequence and untranslated regions of *oskar*, *Delta*, *Notch*, *shotgun*, and *gurken*, as well as the 3'UTR of *nanos*, *bicoid*, and *K10*. The coding sequence of *Y14* was used as a negative control, as detailed in Besse *et al.* (2009). Biotinylated probes in equimolecular amounts were then coupled to streptavidin particles conjugated to magnetic beads. Concurrently, a soluble ovarian protein extract was prepared from wild-type control females (i.e., *w¹¹⁸* stock), generating a pool of soluble proteins that were further incubated with the preformed biotinylated-streptavidin complexes. The mRNA-protein complexes assembled *in vitro* were precipitated using a magnet and then washed to remove unbound proteins. Proteins bound to these mRNA probes, which were present in the immunoprecipitants, were then analyzed via western blotting. It is noteworthy that by adopting the buffer compositions and conditions described in Besse *et al.* (2009), we expect our analysis to therefore reveal direct interactions between RBPs and biotinylated probes. In their study, they demonstrated that Kinesin-1, a component of the *oskar* mRNP transport machinery that lacks RNA-binding capacity, was not detected in the precipitated bound fraction of *oskar* mRNA. Our group has further confirmed these observations for other components of the *oskar* mRNP (data not shown).

Consistent with our computational analysis predictions, we found that Shep preferentially binds to the 3'UTR of the investigated mRNAs, including *oskar*, *Delta*, *Notch*, *shotgun*, *gurken*, *nanos*, *bicoid*, and *K10* (**Figure 5.2A-D**). Interestingly, we showed that both both the large and small Shep isoforms bind exclusively to the 3'UTR of *oskar* mRNA (**Figure 5.2A**). No interactions were observed between Shep and the coding sequence of the negative control *Y14* mRNA. For both *Delta* and *Notch* mRNAs, however, Shep primarily associates with their 3'UTRs and, to a lesser extent, with their 5' untranslated regulatory regions (see panels A & B of **Figure 5.2**). Notably, the coding sequence of *Notch* was the only coding sequence region among the investigated mRNA candidates to show any interaction with Shep (**Figure 5.2B**). This interesting observation aligns with

our computational motif prediction analysis, which indicated that *Notch* had the only and highest motif count among all the analysed transcripts specifically within the coding sequence region (Figure 5.1B & Appendix 5.1-5.4).

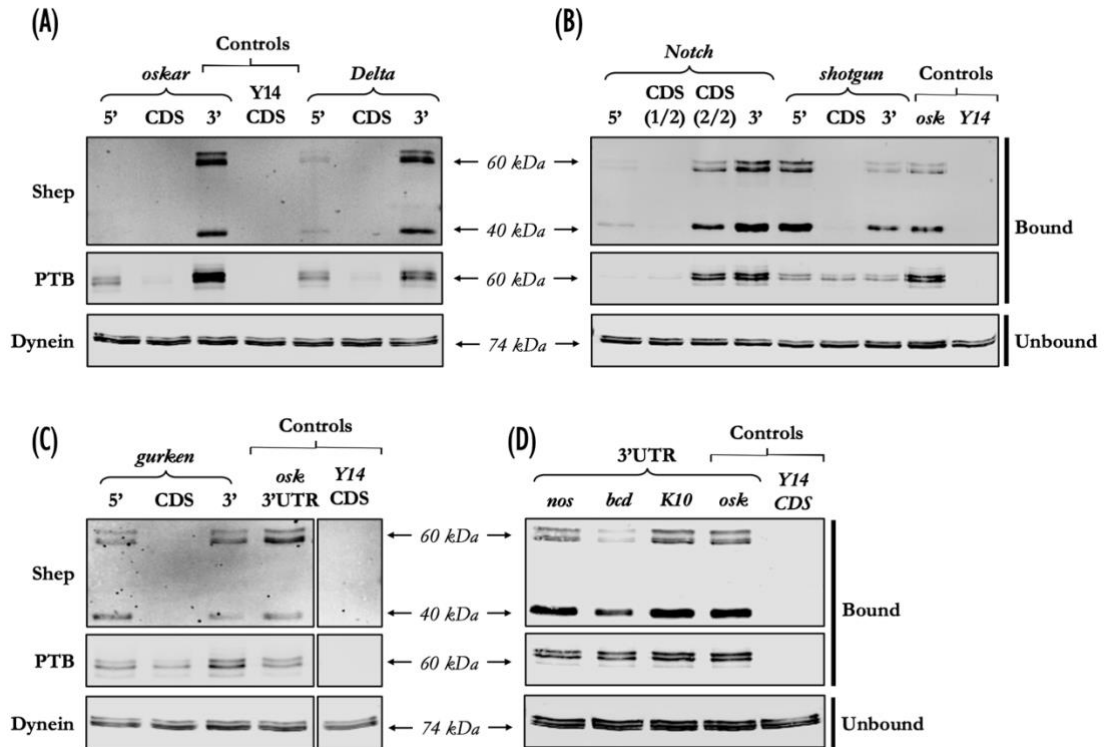


Figure 5.2: Mapping the interactions of Shep with its various mRNA targets. (A–D) RNA-affinity pulldown assay using *n¹¹⁸* ovarian extracts with biotinylated mRNA either covering the entire length of *oskar*, *Delta*, *Notch*, *shotgun*, and *gurken* mRNA or the 3'UTR region of *nanos*, *bicoid*, and *K10* mRNA. In each experiment, the coding sequence of *Y14* was used as a negative control, while the *oskar* 3'UTR was used as a positive control and as a reference for comparison. Proteins in both bound and unbound fractions were visualized by Western blot analysis. The bound fractions were probed with anti-Shep and anti-PTB antibodies. The dynein heavy chain was used as a loading control in the unbound fractions. The ‘Precision Plus Protein Dual Color Standards’ (BioRad) was used as a protein ladder with each run. Abbreviations – 5': 5'UTR. 3': 3'UTR

Remarkably, for *shotgun* mRNA, Shep displays a pronounced preference for the 5'UTR instead and, to a lesser degree, for its 3'UTR (see panel B of Figure 5.2). This observation was unexpected because it did not correlate with the computational analysis predictions, where only 3 motif sequences were found within the 5'UTR compared with the 45 motifs in the 3'UTR (Figure 5.1B). Furthermore, the interactions of the Shep protein with both the 5'UTR and 3'UTR of *gurken* mRNA are arguably similar, as shown in Figure 5.2C. This pattern appears to align positively with our motif analysis. However, the 40 kDa isoform of Shep showed preferential binding to the 5'UTR of

gurken mRNA. Finally, we also confirmed the interaction of Shep with the 3'UTR of other maternally loaded and developmentally significant mRNAs, such as *nos*, *bcd*, and *K10* (Figure 5.2.D).

Shep protein often displays an isoform-specific preference within the same region of a given mRNA. This means that although various Shep isoforms bind to these RNA regions, not all isoforms bind with equal affinities. Notably, the 40 kDa isoform tends to exhibit more pronounced binding than the 60 kDa isoforms. This was particularly evident in the 3'UTR of *bcd* (Figure 5.2D). Interestingly, this trend is also observed with other mRNA targets, such as the 3'UTR of *shotgun*, the coding sequence of *Notch*, and the 5'UTR of *gurken* (see panels B & C of Figure 5.2).

To determine whether the RNA-binding proteins Shep and PTB could be part of similar mRNP complexes, we extended our analysis to include PTB and tested whether PTB binds the same set of mRNA targets that were previously characterized for the interaction with Shep. We observed strong binding of PTB to the *oskar* 3'UTR, and weaker binding to the 5'UTR, demonstrating that our experimental conditions recapitulate previously published results (Besse *et al.* 2009). Interestingly, our results show that the binding pattern of PTB to *Delta*, *Notch*, *shotgun*, *gurken*, *nanos*, *bicoid*, and *K10* mRNAs is comparable to that of Shep (Figure 5.2A-D). However, in contrast to Shep, which showed no binding to the coding sequences of *shotgun* and *gurken* mRNAs, we observed that PTB binds to both these mRNA regions (Figure 5.2B-C).

Overall, our findings collectively demonstrate that Shep not only has distinct binding preferences towards different regions of its mRNA targets but also does so in an isoform-specific manner. These RNA-affinity pulldown assay results also provide a novel characterization of the RBP-mRNA interactions for both Shep and PTB. Additionally, these results indicate that Shep and PTB associate with similar mRNAs, potentially collaborating within the same mRNP complexes to influence different aspects of the cytoplasmic post-transcriptional regulation of these target mRNAs. Finally, these findings also suggest that an RNA molecule typically interact with multiple RBPs at any given time, likely within the mRNP complex. This is evidenced by the

distinct binding patterns exhibited by both Shep and PTB towards the analyzed mRNA molecules.

5.2 Nature of the Interaction Between the RNA-Binding Proteins Shep and PTB

Based on the evidence from our lab's yeast two-hybrid screen data, which suggests a potential direct interaction between Shep and PTB, we have further demonstrated that these RNA-binding proteins localize to the same region within the posterior cortex of the oocyte. Additionally, our findings from the RNA-affinity pulldown assay demonstrated that Shep and PTB bind a similar set of ovarian mRNA targets. Therefore, we next decided to determine whether Shep and PTB may physically interact with each other at the protein level within the mRNP complexes in which they are found.

5.2.1 Does Shep Directly Interact with PTB *in vivo*?

We initially conducted a co-immunoprecipitation assay to pull-down Shep-GFP associated complexes and assess whether PTB co-precipitates, which would indicate an *in vivo* direct interaction between these two proteins. To achieve this, soluble ovarian protein extract from female flies of the Shep GFP protein-trap line (CC00236) was incubated with anti-GFP nanobodies coupled to magnetic agarose beads. GFP-tagged Shep-associating complexes were immunoprecipitated as described in the 'Material & Methods' Chapter, along with its interacting RNA molecules and protein partners, potentially including endogenous PTB. Ovarian extract from flies expressing untagged eGFP under the control of the *osk-Gal4:VP16* driver were used as a negative control. After immunoprecipitation of Shep-GFP and untagged eGFP, western blot analysis of the immunoprecipitants was performed to examine the presence of PTB and Shep in the pulled-down fractions.

Our results showed that the commercial anti-GFP antibodies captured and immunoprecipitated both untagged eGFP and Shep-GFP from the ovarian extracts, as seen in **Figure 5.3A**. Notably, these results revealed that the endogenous PTB protein associated with Shep-GFP but not with the untagged eGFP control, as demonstrated in

panel A of **Figure 5.3**. These findings indicate that Shep and PTB can interact with each other as part of common mRNP complexes. However, given the distribution pattern of these proteins within the ovary, especially the cytoplasmic nature of Shep, it is plausible to question whether the interactions observed here are RNA-dependent. In other words, PTB might have co-immunoprecipitated with Shep-GFP through their interactions with common mRNA molecules, rather than through a direct protein-protein interaction. Importantly, the experimental conditions used here, which are also suitable for RNA immunoprecipitation (detailed later in this Chapter), preserve mRNP complexes.

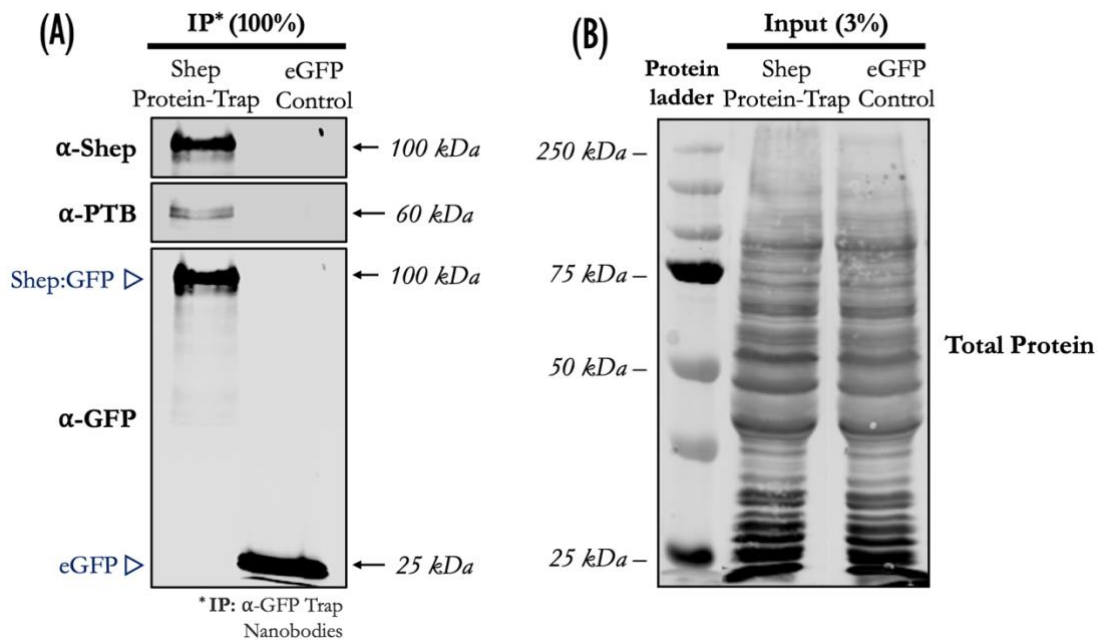


Figure 5.3: Endogenous PTB protein co-immunoprecipitates with Shep protein. Co-immunoprecipitation assay of GFP from untagged-eGFP control (see fly genotype below) and GFP-tagged Shep protein-trap (CC00236) from soluble ovarian protein extracts using nanoantibodies against GFP. Proteins in the bound **(A)** and input **(B)** fractions were visualized using Western blotting. The bound fractions were probed with anti-Shep, anti-PTB, and anti-GFP antibodies. Total protein stain of the input fractions with Revert™ 700 (Licor) was used as a loading control. In the membrane probed with the anti-GFP antibody, GFP-tagged Shep (approximately 100 kDa) and untagged eGFP (around 25 kDa) were detected in the different experimental samples. The ‘Precision Plus Protein Dual Color Standards’ (BioRad) was used as a protein ladder with each run. A minimum of three independent experiments were conducted. Abbreviations – IP: Immunoprecipitation. Fly genotype(s) – Ovaries from untagged eGFP control females (*w⁻; ask-Gal4:VP16/Cyo; UASp-eGFP/TM3,Sb*) and the Shep GFP protein-trap line (*w⁻; Shep:GFP^{CC00236}/ Shep:GFP^{CC00236}*).

5.2.2 Is the Shep-PTB Interaction RNA-Dependent?

To determine whether the interaction between Shep and PTB proteins is nucleic acid dependent, we performed a co-immunoprecipitation procedure similar to the one described above. However, this time, the ovarian cell lysates were subjected to a nuclease treatment. Given the cytoplasmic distribution pattern of Shep in the ovaries, it is reasonable to assume that its interaction with PTB occurs in the cytoplasm and is independent of DNA-mediated interactions. For this reason, ovarian cell lysates from Shep-GFP protein-trap (CC00236) and untagged eGFP control female flies were treated with ribonucleases to digest any potential RNA-mediated interactions. A combination of both RNase A and RNase T1 was used to ensure complete digestion of the RNA molecules. Before using ribonucleases in the co-immunoprecipitation assay, we first tested various combinations of temperatures and incubation periods to determine the optimal conditions for complete RNA digestion within an ovarian lysate (data not shown). The exact optimal conditions used for RNA digestion are described in the ‘Materials & Methods’ Chapter.

To examine whether the interaction between Shep and PTB proteins is RNA-dependent, we performed western blot analysis on the immunoprecipitants of Shep-GFP and untagged eGFP, which were obtained from the RNase-treated ovarian lysate. Remarkably, in the absence of intact RNA molecules, the endogenous PTB protein no longer co-immunoprecipitated with Shep-GFP, as demonstrated in panel A of **Figure 5.4**. These results, in conjunction with our RNA-affinity pulldown assays, demonstrate that both Shep and PTB interact *in vivo* through their association with common mRNA targets. Interestingly, the lack of direct physical interaction between Shep and PTB contradicts the findings obtained from the lab’s yeast two-hybrid assay. A possible explanation for this discrepancy could be that Shep was identified either as a false positive hit in the screen, or through an unexpected RNA-mediated interaction within the yeast. This hypothesis is supported by the fact that only one positive clone for Shep was identified among the 73 independent clones in total. Alternatively, it is possible that a direct Shep-PTB interaction occurs primarily through the small Shep isoform, which are not tagged in the protein-trap line used in our co-immunoprecipitation analysis.

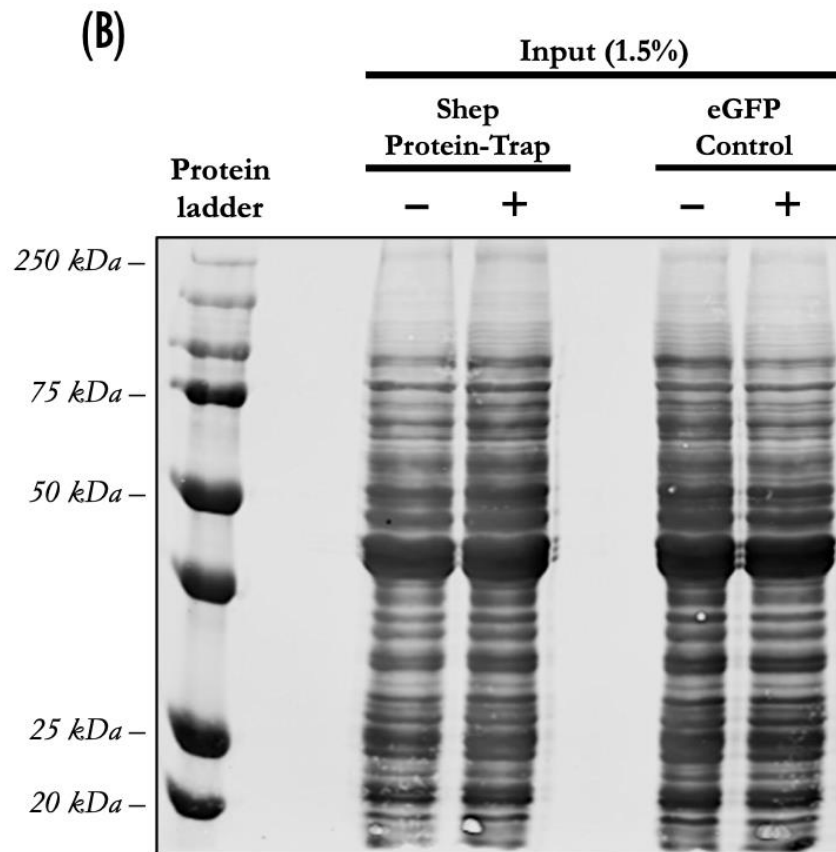
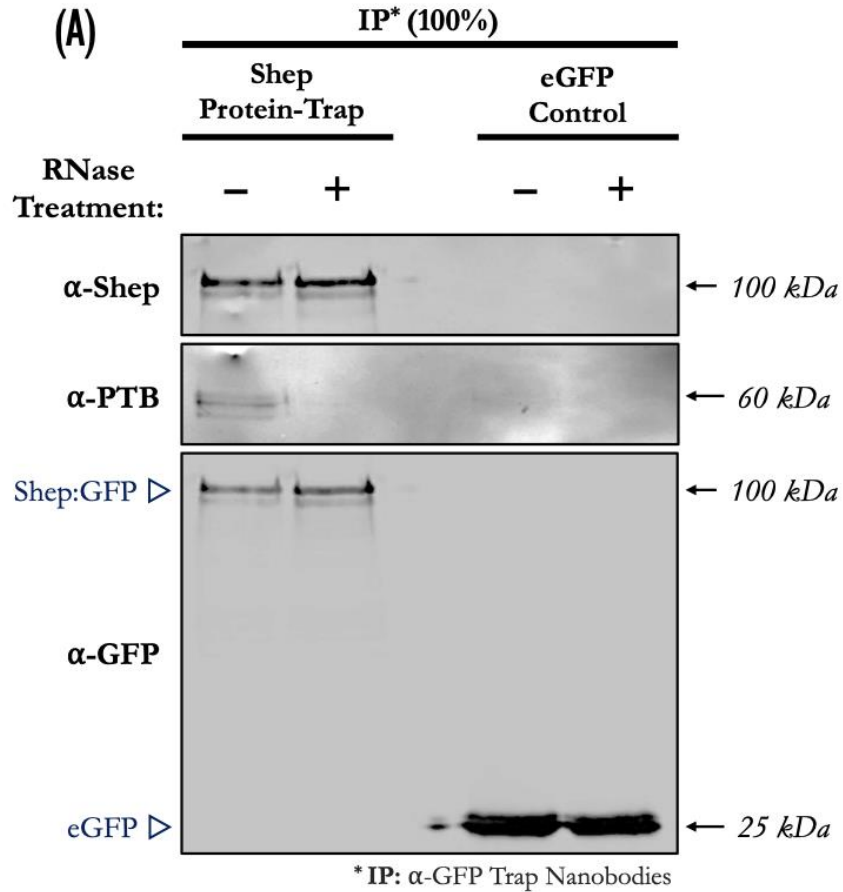


Figure 5.4: PTB protein interacts with Shep in an RNA-dependent manner. Co-immunoprecipitation assay of GFP was conducted using soluble ovarian protein extracts from untagged eGFP control (see fly genotype below) and GFP-tagged Shep protein-trap (CC00236). The extracts were either untreated (denoted as –) or treated (denoted as +) with ribonucleases. GFP in both experimental samples was immunoprecipitated using nanoantibodies against GFP. Proteins in the bound **(A)** and input **(B)** fractions were visualized using Western blotting. The bound fractions were probed with anti-Shep, anti-PTB, and anti-GFP antibodies. A total protein stain with Revert™ 700 (Licor) was used as a loading control in the input fractions. In the membrane probed with the anti-GFP antibody, GFP-tagged Shep (approximately 100 kDa) and untagged eGFP (around 25 kDa) were detected in the different experimental samples. The ‘Precision Plus Protein Dual Color Standards’ (BioRad) was used as a protein ladder with each run. Two independent experiments were conducted. Abbreviations – IP: Immunoprecipitation. Fly genotype(s) – Ovaries from untagged eGFP control females (*w⁻; osk-Gal4:VP16/Cyo; UASp-eGFP/TM3,Sb*) and the Shep GFP protein-trap line (*w⁻; Shep:GFP^{CC00236}/ Shep:GFP^{CC00236}*).

5.3 Identification of Ovarian mRNAs Bound to Shep *in vivo*

Based on our findings thus far, which encompass the subcellular distribution pattern of Shep in the ovary, its expression regulation in the soma, functional analyses, and RNA-affinity pulldown assay results, we aimed to obtain *in vivo* evidence for the interaction between Shep and different ovarian mRNAs. These mRNAs include *oskar*, *gurken*, *Delta*, *Notch*, and other candidate mRNAs. Investigating these interactions would elucidate the potential biological role that Shep may play in *D. melanogaster* oogenesis. To identify the mRNA targets with which Shep associates *in vivo*, we performed RNA immunoprecipitation (RIP) followed by semi-quantitative RT-PCR analysis. The relative abundance of mRNA in experimental samples was evaluated based on early detection of RT-PCR amplification and the degree of enrichment in Shep-GFP immunoprecipitants compared with the control. Therefore, detecting enriched RT-PCR amplification in Shep-GFP – and not in the control – suggests an *in vivo* interaction between Shep and the investigated mRNA.

Because of the limited availability of antibodies validated for immunoprecipitation experiments using the endogenous Shep protein, we decided to characterise the mRNA targets of the GFP-trapped Shep in the protein-trap line using nanobody-based reagents specifically designed for efficient GFP immunoprecipitation (ChromoTek GmbH, USA). Magnetic beads coupled to anti-GFP nanobodies were used to immunoprecipitate the GFP-tagged Shep protein from ovarian extract, prepared from

roughly 250 pairs of Shep protein-trap (CC00236) female flies. As part of this process, mRNA molecules associated with Shep were also co-immunoprecipitated. To account for potential non-specific binding to the GFP tag and beads, control flies expressing untagged eGFP under the control of the *osk-Gal4:VP16* driver were used for comparison. Western blot analysis was carried out to ensure that the GFP-labelled Shep or plain eGFP in the samples were successfully immunoprecipitated (see **Figure 5.5C**). mRNA molecules from the immunoprecipitant fractions were purified and treated with DNase I before being converted into cDNA, as described in the ‘Material & Methods’ Chapter. Despite the genomic DNA being digested, the primers used for this analysis were designed to distinguish spliced mRNAs from their unspliced counterparts.

In the semi-quantitative approach carried out here, we compared the differences in amplicon intensities between Shep-GFP and the control. The number of PCR cycles used for each target mRNA varied, ranging between 25 and 39 cycles, depending on their relative abundance in the RIP samples. Remarkably, we found that the immunoprecipitation of Shep-GFP specifically co-precipitated *oskar*, *gurken*, *nanos*, *bicoid*, *Delta*, *Notch*, and *shotgun* mRNAs. These interactions are evident from the differential amplification observed for each target mRNA between the Shep-GFP and control immunoprecipitated samples (**Figure 5.5A**). Interestingly, these *in vivo* interactions are consistent with our *in vitro* findings obtained from the RNA-affinity pulldown assay, as detailed previously in Section 5.1.2 of this Chapter. It is noteworthy that these results were obtained from two distinct and independent biological replicates of RIP experiments.

Overall, our data provide *in vivo* evidence suggesting that Shep interacts with specific ovarian mRNAs, namely *oskar*, *gurken*, *nanos*, *bicoid*, *Delta*, *Notch*, and *shotgun* mRNAs. Within the ovary, it is highly plausible that Shep, as part of their cytoplasmic mRNP complexes, influences the post-transcriptional regulation of these mRNAs. Such regulation may encompass aspects such as mRNA stability, translation efficiency, subcellular localization, or a combination of these.

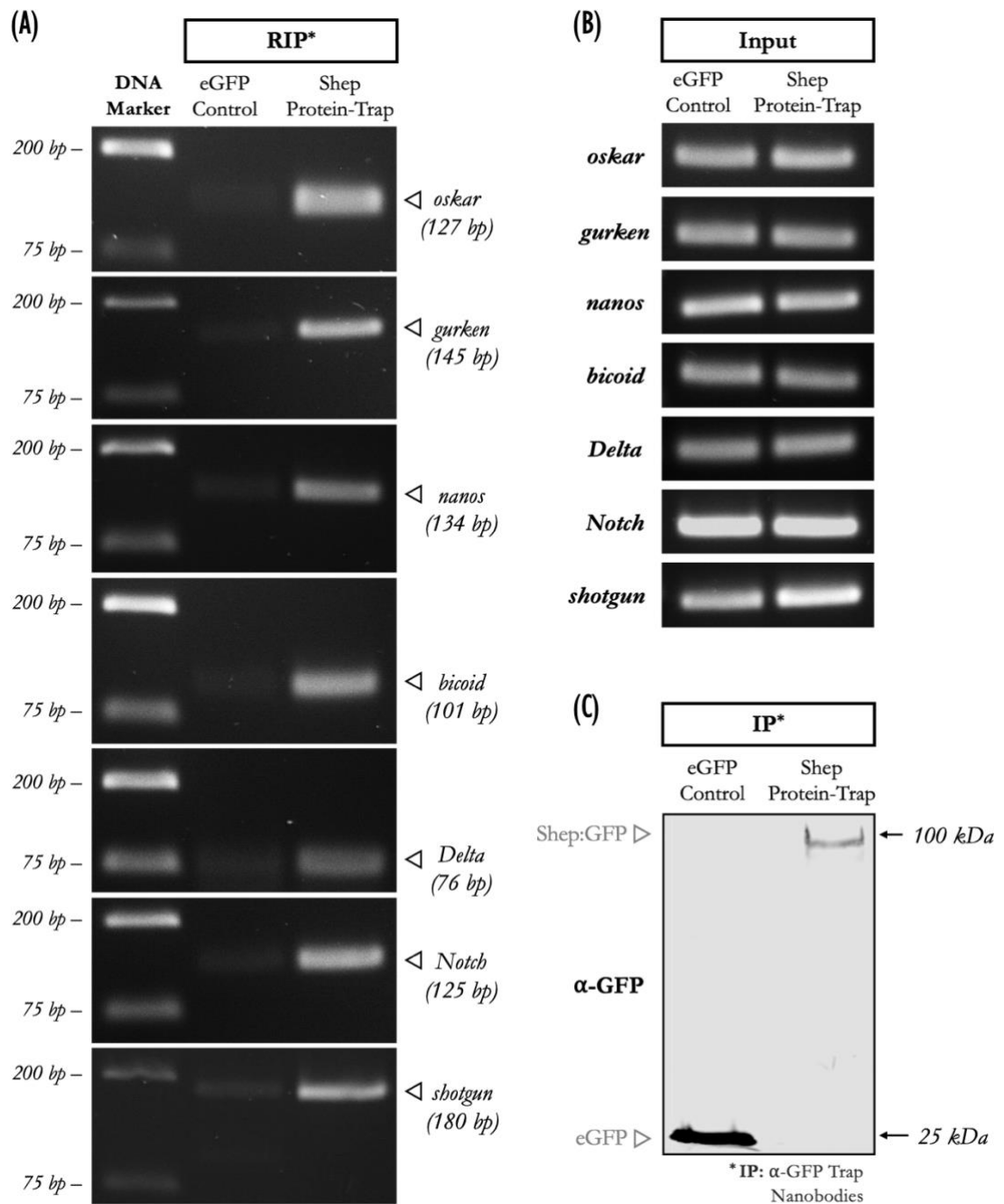


Figure 5.5: Differential enrichment of mRNAs bound to Shep in the ovary. Identification of *in vivo* interactions between Shep and seven putative interacting mRNA targets: *oskar*, *gurken*, *nanos*, *bicoid*, *Delta*, *Notch*, and *shotgun* mRNA. **(A)** Lysates derived from the ovaries of both untagged-eGFP control flies and GFP-tagged Shep protein-trap (CC00236) flies were subjected to immunoprecipitation using nanoantibodies against GFP. RNA was extracted from these immunoprecipitated fractions. The precipitates were examined for the presence of seven putative interacting mRNAs using RT-PCR: *oskar* (27 cycles), *gurken* (39), *nanos* (33), *bicoid* (39), *Delta* (35), *Notch* (35), and *shotgun* (37) mRNAs. Analysed on a 2% agarose gel, all seven mRNAs were all specifically enriched in the Shep-GFP precipitate. A GeneRuler 1 kb plus DNA ladder (Thermo Fisher Scientific) was used with each run. **(B)** RT-PCR amplifications of mRNAs recovered from the input fractions were similarly analyzed and used as controls. **(C)** Validation of immunoprecipitations via western blot analysis. The immunoprecipitants, which were eluted in the SDS-PAGE loading buffer, were electrophoresed on a 10% SDS-PAGE gel. The nitrocellulose

membrane was then probed with anti-GFP for the detection of GFP-tagged Shep (around 100 kDa) and untagged eGFP (around 25 kDa) in the different experimental samples. The ‘Precision Plus Protein Dual Color Standards’ (BioRad) was used as a protein ladder. Abbreviations – RIP: RNA immunoprecipitation, IP: Immunoprecipitation. Fly genotype(s) – Ovaries from untagged eGFP control females (*w¹¹¹⁸; osk-Gal4:VP16/Cyo; UASp-eGFP/TM3,Sb*) and the Shep GFP protein-trap line (*w¹¹¹⁸; Shep:GFP^{CC00236}/ Shep:GFP^{CC00236}*).

5.3.1 Optimization of the RIP Protocol

Multiple optimization attempts were performed to refine the conditions for the RNA immunoprecipitation protocol. The aim was to find robust experimental conditions to purify RNA molecules for the RNA-seq analysis, which will allow the identification of novel Shep mRNA targets. During these attempts, we tested different buffer compositions and changed both the duration and number of washes. Immunoprecipitated RNAs were evaluated based on four parameters: their quantity, purity, quality, and detection of differential expression between experimental groups.

In our initial attempt at the RIP procedure, we adhered to the general procedure outlined in Clare Alexandra Pritchard’s thesis (Cardiff University, 2016), but used buffers with compositions similar to those described by Gonsalvez *et al.* (2010). Unfortunately, this initial attempt failed to meet any of the four evaluation parameters. First, very little RNA was immunoprecipitated, as indicated by Qubit readings of around 5ng/μL. Second, the absorbance ratio at 260nm and 280nm was used to assess the purity of RNA in the different experimental RIP samples. A low A₂₆₀/A₂₈₀ ratio of less than 1.2 was obtained using a Nanodrop spectrophotometer, indicating the presence of proteins, phenols, or other contaminants in the RNA samples, which could interfere with downstream experiments such as RNA-seq. Third, despite the use of ribonuclease disinfectants and careful handling of the samples throughout the procedure, we observed compromised integrity of the immunoprecipitated RNAs when analyzed using high-sensitivity RNA ScreenTape. RNA degradation was observed, as indicated by the presence of smears and low molecular weight RNA molecules of 25 bp (**Figure 5.6B**). Despite digesting the gDNA with DNase I provided by the RNAqueous Micro Total RNA Isolation kit, ScreenTape analysis showed the presence of high molecular weight nucleic acid fragments of above 6Kb, which could indicate the presence of gDNA contamination (**Figure 5.6B**). Finally, RT-PCR failed to detect any differential enrichment between the Shep-GFP and control RIP samples (data not shown). For

instance, it was not until the 35th PCR cycle that *oskar* amplicons were equally detected in both experimental groups, indicating no differential enrichment between them (data not shown). In our more recent successful attempts, using a similar amount of starting cDNA template, *oskar* mRNA was differentially enriched in the Shep-GFP RIP sample. In this case, the *oskar* amplicon was detected around the 25th PCR cycle in the Shep-GFP sample and was absent in the control sample, (see **Figure 5.5A**). This is ten cycles earlier than our initial unsuccessful attempt.

In the most recent successful attempt and its replica, we used Tris-HCl based buffers with lower ionic strength, as detailed in the ‘Materials & Methods’ Chapter. As additional precautionary measures against both protein and RNA degradation, we supplemented the buffers with a protease inhibitor cocktail, PMSF protease inhibitor, and RiboShield RNase inhibitor. Although the RNAqueous Micro Total RNA Isolation kit was used for RNA purification, gDNA was not digested with the DNase I supplied by the kit. Instead, RNase-free DNase I from Thermo Fisher Scientific was used. Importantly, we found that the modifications made to the buffer compositions resulted in a significant improvement in the quality and yield of immunoprecipitated RNA.

5.3.2 Quality Control Check of RIP Samples for Transcriptome Sequencing

To ascertain whether the RNA molecules immunoprecipitated with Shep-GFP and untagged eGFP proteins using the newly optimized RIP buffer composition and protocol were suitable for the RNA-seq workflow, we evaluated the immunoprecipitated RNAs from both experimental groups based on four criteria: quantity, purity, quality, and size distribution. These criteria were established in line with the specifications recommended by the manufacturers of RNA Library Prep Kits and commercial providers of the RNA-seq service. For example, Novogene recommends the following specifications: an RNA sample of at least 100 ng concentration (at least 3 ng/ μ L) and an ideal A260/A280 absorbance ratio measurement of 2 or higher. Quantification of the RNA molecules immunoprecipitated with either Shep protein-trap or untagged eGFP as bait using Qubit revealed a concentration of around 30 ng/ μ L, as shown in **Table 5.1**, yielding a total amount of 600 ng of

precipitated RNA. RNA purity was evaluated using a NanoDrop spectrophotometer. The A260/280 absorbance ratio of approximately 2.0 for both experimental groups indicated pure RNA samples (Table 5.1).

Table 5.1: Quality control of RIP samples to check RNA compatibility for transcriptome sequencing. *: RNA eluate of 20µL. Abbreviations – HS: high sensitivity, RIP: RNA Immunoprecipitation

Novogene Requirements					
Sample Type		Concentration	Total Amount	Purity	
RNA		≥ 3 ng/µL	≥ 100 ng	A260/280 > 2.0	
Method:		HS RNA Qubit Assay		NanoDrop	
#	RNA sample	Concentration	Total Amount*	A260/280	
1	eGFP RIP	33 ng/µL	660 ng	2.11	
2	Shep-GFP RIP	31 ng/µL	620 ng	2.00	

Analytical assessment of RNA quality was conducted using the High Sensitivity RNA ScreenTape assay and analyzed using the TapeStation system. The immunoprecipitated RNA samples from both experimental groups were intact. This was demonstrated by the intensity and integrity of the RNA bands, as shown in **Figure 5.6A**. Additionally, the absence of smearing and the lack of a shift towards shorter fragment sizes indicate the intactness of the RNA within both samples. Furthermore, the RNA size distribution in both samples ranged from 100 to 6,000 bp, with an average length of approximately 2Kb (**Figure 5.6C**). This implies the presence of a diverse, representative, and high-quality pool of mRNA molecules.

Overall, the results presented here indicate that the optimization of the buffers and conditions of our RIP experiments – as detailed in the ‘Materials & Methods’ Chapter – effectively immunoprecipitated RNAs associated to a GFP-tagged RNA binding protein such as Shep, yielding the total RNA quantity and quality required for RNA sequencing.

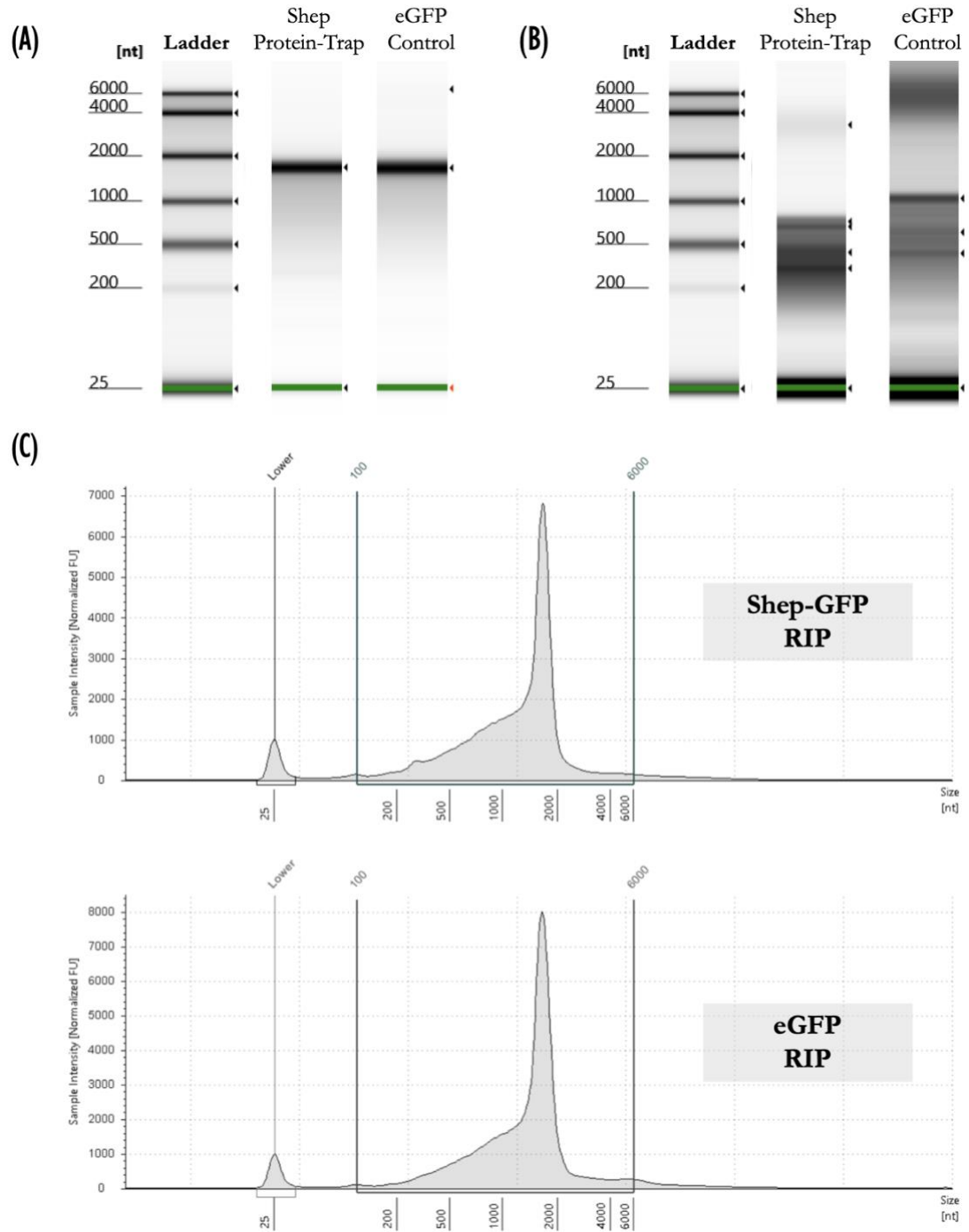


Figure 5.6: Analysis of RNA from the RNA immunoprecipitation experiments. RNA samples were analyzed for quality and intactness on the TapeStation system (Agilent Technologies) using the High Sensitivity RNA ScreenTape assay, followed by RNA immunoprecipitation, RNA purification, and DNase I treatment. RNAs were immunoprecipitated from the ovaries of homozygous Shep protein-trap (CC00236) and eGFP control (*asf-Gal4:VP16 > UASp-eGFP*) adult females using GFP nanoantibodies for immunoprecipitation. **(A – B)** Microfluidic-based electrophoretic separation. The ScreenTape assay employs precast multilane gels and microfluidics in an automated system to produce gel-like images of RNA immunoprecipitants. Images reflect the quality of RNA derived from two separate RIP attempts: the optimized **(A)** and initial pre-optimization **(B)** RIP protocols. The 25nt position is highlighted in green. **(C)** Electropherogram

of immunoprecipitated RNA derived from the optimized RIP protocol. The panels illustrate the size distribution of RNA immunoprecipitants obtained from Shep-GFP (Top) and eGFP (Bottom), showing average RNA lengths of 1,799 and 1,945 nt, respectively. Abbreviations – RIP: RNA Immunoprecipitation, nt: nucleotide.

5.4 Conclusions

In conclusion, our findings provide novel characterizations of the interactions of Shep with key mRNAs and the RNA-binding protein PTB in the *D. melanogaster* ovary. It also sheds light on the potential cytoplasmic associations of Shep with multiple mRNPs during oogenesis, thereby enhancing our current understanding of the complex and intricate regulation of RNA during this critical stage of development. A combination of approaches was employed to characterize the interactions of Shep with mRNAs and PTB in the context of oogenesis. These included sequence-based computational analyses and several *in vivo* and *in vitro* biochemical assays.

Using an RNA-affinity pulldown assay, we have shown that Shep preferentially binds to the 3'UTR regions of multiple mRNAs, a finding that correlates with the predictions we scored using Shep RNA-binding motifs reported by Ray *et al.* (2013). Additionally, we observed that Shep exhibited isoform-specific binding to distinct regions within the investigated mRNA targets. Additionally, from the motifs themselves and our sequence analysis of the putative Shep-binding sites, it appears that Shep preferentially associates with AU-rich sequences predominantly found within the 3'UTR of the investigated mRNAs. If this proves to be true, then it increases the likelihood that Shep operates as a *trans*-acting RNA-binding protein that interacts with AU-rich elements within 3'UTR sequences, thereby regulating the stability and translation of mRNAs. These observations not only reinforce our hypothesis regarding the cytoplasmic role of Shep in the regulation of mRNA stability and translation but also complement our results obtained from the functional analyses detailed in Chapter 4.

Remarkably, co-immunoprecipitation of Shep-GFP using anti-GFP nanobodies revealed that Shep and PTB interact primarily by associating with common mRNAs in the ovaries. Moreover, given our RNA-affinity pulldown assay results, which showed both proteins as components of numerous common mRNP complexes, and taking into

account their posterior crescent subcellular localization pattern in stage 10 oocytes and the observed translational repression of *oskar* mRNA, it would be interesting to explore how Shep and PTB collaboratively regulate *oskar* mRNA.

Finally, through the optimization of the RNA immunoprecipitation protocol and subsequent RT-PCR enrichment analysis, we have provided *in vivo* evidence of the interaction between Shep and the mRNP complexes of *oskar*, *gurken*, *nanos*, *bicoid*, *delta*, *notch*, and *shotgun* mRNAs in fly ovaries. This has allowed us to hypothesize the potential biological roles of Shep during *D. melanogaster* oogenesis based on the phenotypes we observed in the previous chapters or presented in the literature. Additionally, we demonstrated that our RIP protocol can yield RNA of a concentration and quality suitable for RNA-seq, which will allow us to identify new RNA targets of Shep, and other RNA-binding proteins, in the future.

⌘ Chapter VI ⌘

Discussion

6 Chapter VI – Discussion:

6.1 Overview of the State of Knowledge Prior to Our Study

The *Drosophila* gene *alan shepard* (*shep* for short) encodes a highly conserved RNA-binding protein equipped with two RNA-recognition motifs (Matzat *et al.* 2012). Ever since the *shep* gene was identified by Armstrong *et al.* (2006) in a forward genetic screen for its role in sensory-motor responsiveness to gravity, research has primarily focused on its function within the nervous system. The *Drosophila* Shep protein functions in the cytoplasm and nucleus of both neurons and glia. During metamorphosis, Shep has been shown to regulate neuronal remodelling either through direct regulation of transcriptional programs (Chen *et al.* 2017; Chen *et al.* 2018) or by acting as a negative regulator of the gypsy insulator complex (Matzat *et al.* 2012; Chen *et al.* 2019; Chen *et al.* 2021).

In addition to its nuclear functions, studies have shown, using immunofluorescence, that Shep is also localized to the cytoplasm (Chen *et al.* 2014; Schachtner *et al.* 2015). Cytoplasmic Shep functions in the dendrite branching of nociceptive neurons, regulates the organization of neuronal clusters in the peripheral nervous system, and orchestrates the organization of axons within the ventral nerve cord (Schachtner *et al.* 2015). Although a role for the Shep protein has been demonstrated in the nervous system, its cytoplasmic function remains poorly understood, particularly the mechanisms through which it mediates post-transcriptional gene regulation.

Before our analysis, the expression and role of Shep in *Drosophila* oogenesis remained largely unexplored. Our lab's yeast two-hybrid screen pointed to an interaction between Shep and PTB, an RNA-binding protein known for its role in the translational repression of *oskar* mRNA in the ovary (Besse *et al.* 2009). Additionally, considering the results of my initial GFP protein-trap screening of candidate proteins, which revealed Shep localization at the posterior pole of the oocyte, we set out to comprehensively characterize its expression pattern, subcellular distribution, and elucidate its post-transcriptional regulatory role within the female germline.

6.2 *shep* Displays a Complex Gene Expression Pattern in *Drosophila* Ovaries

6.2.1 Transcription of Multiple *shep* Transcripts Gives Rise to Several Protein Isoforms

The *Drosophila shep* gene is predicted to produce eight distinct transcripts, which give rise to six unique polypeptides. In our initial analysis of the modENCODE RNA-seq dataset from both virgin and mated females (Consortium *et al.* 2011), we concluded that several *shep* transcripts were expressed in the ovaries (**Figure 6.1**). However, we could not determine which specific *shep* transcripts they were because the reads mapped to exons that are shared among several transcripts. Therefore, before this study, out of the eight transcripts listed on FlyBase, *shep-F* was the only transcript we could confidently ascribe as being expressed in the ovaries. Identifying transcript F from these datasets was straightforward because the enriched reads uniquely mapped to a transcript-F-specific exon encoding a region of the CDS, as depicted in **Figure 6.1** (see region labelled #2 highlighted in blue). Furthermore, transcripts A and F both have reads aligning to an exon that is unique to them, making it impossible to ascertain from the modENCODE dataset if only one or both are expressed, as illustrated in the region labelled #6 in **Figure 6.1** (highlighted in orange).

A similar issue also arose with transcripts E, H, and I, as shown in the region labelled #1 in **Figure 6.1** (highlighted in yellow). For transcripts B and D, however, the sparse read counts aligned to their unique transcript-specific exon suggest potential expression, albeit at very low levels (see regions #4 in green and #5 in cyan of **Figure 6.1**, respectively). In fact, this low abundance of transcript B correlates with the band intensity observed in our semi-quantitative RT-PCR, as illustrated in panel B of **Figure 3.4** (see Chapter 3). Furthermore, transcript G appears not to be expressed in the ovaries, as evidenced by the absence of reads aligning to its transcript-specific exon, as shown in the purple highlighted region of **Figure 6.1** (see region labelled #3).

Given the previously discussed challenges posed by the shared exons across multiple transcripts, determining the expression of individual transcripts from the modENCODE dataset becomes difficult. Nonetheless, by using various combinations

of primer pairs that utilize those transcript-specific exons detailed in Section 3.4.2, our semi-quantitative RT-PCR analysis has defined the profile of *shep* transcripts in ovarian tissues, offering insights at the individual transcript level. Specifically, we showed that transcripts A, B, E, F, H, and I are expressed in the fly ovary, whereas transcripts D and G are not. Our findings from the *D. melanogaster* ovary align with the modENCODE RNA-seq data, which suggests that multiple *shep* mRNA molecules are transcribed in the ovary.

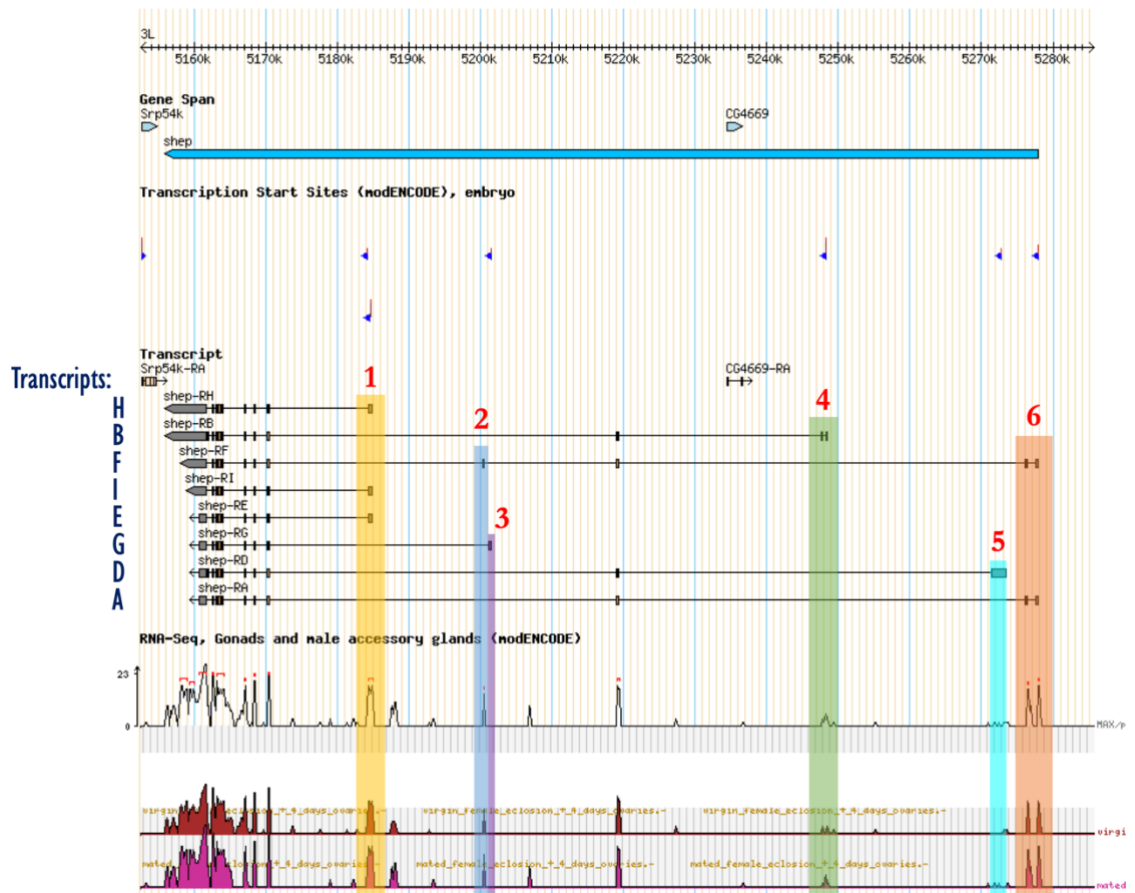


Figure 6.1: modENCODE transcriptome data of the *shep* gene in the *Drosophila* ovary.

Analysis of the modENCODE transcriptome profile for the *shep* gene in *Drosophila* ovaries was performed using FlyBase. The *D. melanogaster* genome viewer, GBrowse, displays four selected tracks: gene span, transcription start sites from the modENCODE embryonic dataset, transcript, and RNA-seq profiles from the modENCODE Gonads dataset. The *shep* gene locus (shown in blue at the very top) along with the eight possible transcripts produced. The exon-intron layout of the transcripts (A, B, D, E, F, G, H, and I) is represented by wide bars and black lines. Exons are colour-coded to denote different mRNA regions: light orange for the coding sequence and gray for untranslated regions. The orientation of the gene and transcripts is indicated by the pointed ends, reflecting the direction of transcription. The RNA-seq profile of the *shep* gene in adult female gonads 4 days post eclosion is displayed using a log₂ signal scale view, with an increased vertical spacing to improve visibility of the tracks. Each coloured track represents adult females from which ovarian samples were taken: maroon for virgin females (mE_mRNA_A_VirF_4d_ovary)

and pink for mated females (mE_mRNA_A_MateF_4d_ovary). The transparent track offers a consensus view of the transcriptome, with RNA-seq signals showing exceptional read enrichment, as indicated by the peaks marked by the red horizontal lines above. Signals mapped to the genomic minus strand are shown. Unique exons common to specific subsets of transcripts are highlighted: orange (#6) for transcripts A and F, and yellow (#1) for transcripts E, H, and I. Meanwhile, regions highlighted in blue (#2), purple (#3), and cyan (#5) correspond to a transcript-specific exon unique to transcripts F, G, and D, respectively. The green-highlighted area (#4) indicates low transcription activity in transcript B. The modENCODE transcriptome data of the *shep* gene was accessed and retrieved from FlyBase (Gramates *et al.* 2022).

As previously mentioned, the *shep* gene is predicted to transcribe eight distinct transcripts that arise from different transcription start sites and alternative splicing. At the protein level, however, only 6 unique isoforms are translated from these transcripts because the coding sequences of transcripts B/D and E/G are identical (**Figure 1.14**). While many of these transcripts may differ in their coding sequences, the predominant differences are found in their untranslated regions, highlighting the complexity of post-transcriptional regulation that these transcripts undergo. Additionally, the use of different transcription start sites by the *shep* gene to produce these various transcripts further underscores the intricate control to which the gene is subjected (**Figure 1.13**).

Our western blot analysis confirmed the various Shep isoforms that were expressed in the ovaries (**Figure 3.3**). Specifically, three bands of Shep were detected and initially categorized based on their relative size: the ‘large’ isoforms could include Shep A, F, B, and D with predicted molecular weights ranging from 53 to 62 kDa; while the ‘small’ isoforms may correspond to Shep G, E, H, and I, each around 40 kDa in weight. Determining the identities of the isoforms expressed in the ovary was challenging because each band can represent multiple Shep isoforms. Therefore, through our RT-PCR analysis, which revealed the specific transcripts expressed in the ovary, we were able to pinpoint and identify the isoforms detected by Western blotting. As shown in **Figure 6.2**, the three distinct bands of Shep detected in the ovaries represent the expression of all six isoforms: isoforms A and F at around 70 kDa, isoform B at approximately 65 kDa, and isoforms E, H, and I each at around 40 kDa. Interestingly, although transcript B was relatively less abundant than the other transcripts (refer to panel B of **Figure 3.4** in Chapter 3), its protein expression appears relatively comparable to the other isoforms. This implies that the translational efficiency of *shep* transcripts is dynamically regulated at the post-transcriptional level. The observed discrepancy

between mRNA levels and protein abundance is a phenomenon frequently observed in various biological systems (Gygi *et al.* 1999; Nie *et al.* 2006; Pascal *et al.* 2008; Vogel and Marcotte 2012). Further investigation is needed to address how Shep translation is regulated by mechanisms that either enhance the translation of transcript **B** or hamper the translation of the other isoforms.

Altogether, cross-referencing findings from both RT-PCR and Western blot analyses allowed us to determine that all Shep isoforms are expressed in the ovary, despite the fact that only six of the eight *shep* transcripts annotated in FlyBase are expressed (**A**, **B**, **E**, **F**, **H**, and **I**), while **D** and **G** are not. This was possible because the coding sequence of transcript **D** is identical to that of **B**, and similarly, the coding sequence of transcript **G** is also indistinguishable from that of **E**.

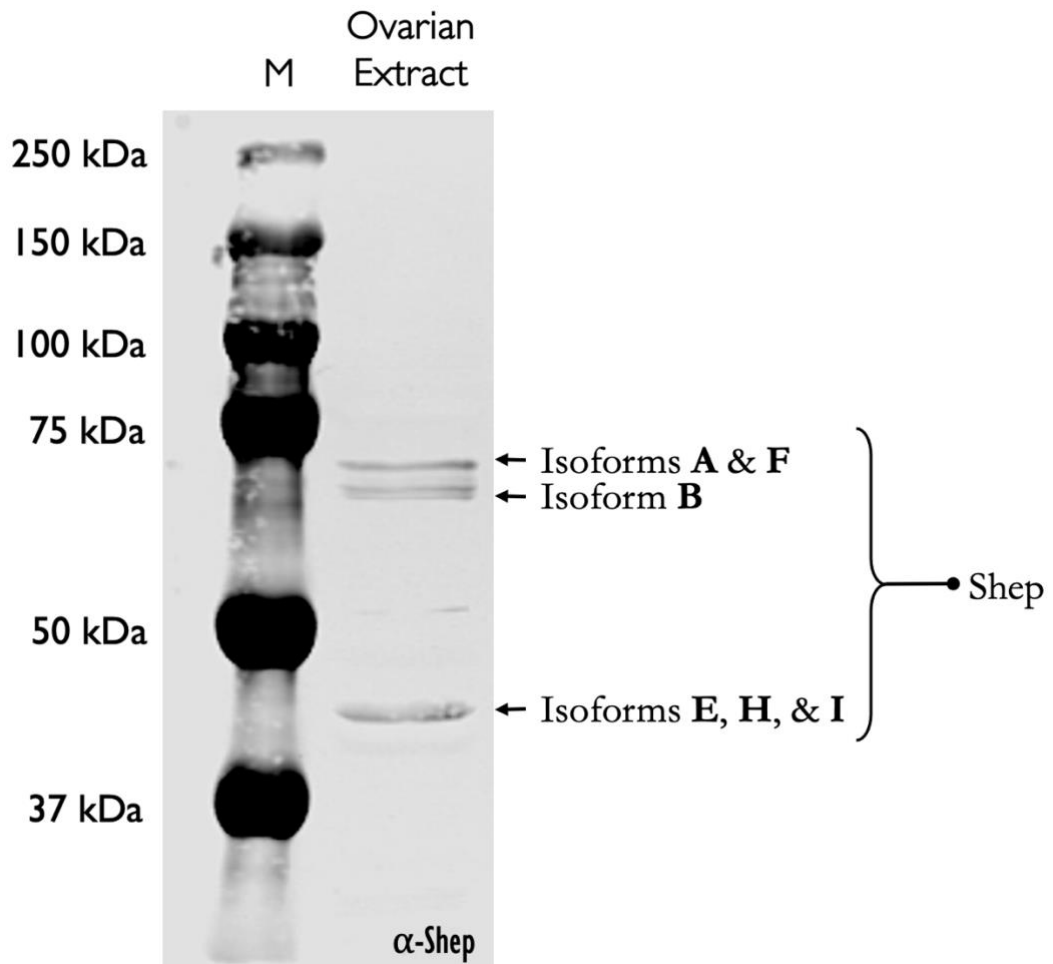


Figure 6.2: Protein expression of Shep isoforms in the *Drosophila* ovary. Ovarian crude extracts, equivalent to one pair of ovaries, from *w¹¹¹⁸* females were electrophoresed on a 9% SDS-PAGE gel. Shep was detected using Western blotting by probing with rabbit anti-Shep antibodies. A protein marker is in lane labelled 'M' (precision plus protein dual color standard, BioRad). The

Shep proteins ran slightly higher than their predicted molecular weights, with apparent molecular weights of approximately 40, 65, and 70 kDa.

6.2.2 Shep Expression Pattern is Altered in the CC00236 Protein-Trap Line

Given the precise location of the GFP-cassette insertion in the *shep* gene locus, it could theoretically trap the large isoforms (A, B, D, and F) and potentially also the smaller isoforms (E, G, H, and I). Moreover, using ovarian extracts from the Shep GFP protein-trap line, our Western blot analysis showed that the large Shep isoforms were shifted to 100 kDa, reflecting the added molecular weight of GFP (roughly 25 kDa; Prasher *et al.* 1992). Surprisingly, the small isoforms of Shep (i.e., E, H, and I), normally observed at 40 kDa in wild-type ovaries, were neither detected nor shifted to around 65 kDa in the protein-trap line ovaries (as shown in **Figure 3.3**).

The absence of the small Shep isoforms indicates that the Shep CC00236 protein-trap flies are null mutants for these isoforms. Based on these observations, we hypothesized that the chromosomal location of the GFP-trap cassette insertion into the 5'UTR region of the small isoforms might interfere with either their transcription or translation. Upon further investigation of the precise location of the CC00236 P-element insertion, we found that the GFP-trap cassette is situated between the 6th and 7th nucleotides of the transcription start site (TSS_mE1_005337), which is responsible for the transcription of transcripts E, H, and I (**Figure 6.3B**). Given that the GFP-cassette insertion is located within this transcription start site window, the most plausible explanation for the absence of the small Shep isoforms in the protein-trap line is a disruption of their transcription rather than their translation. This also implies potential interference with the promoter's *trans*-acting regulatory elements necessary for transcribing these specific transcripts. Further experimental investigation is needed to validate our hypothesis.

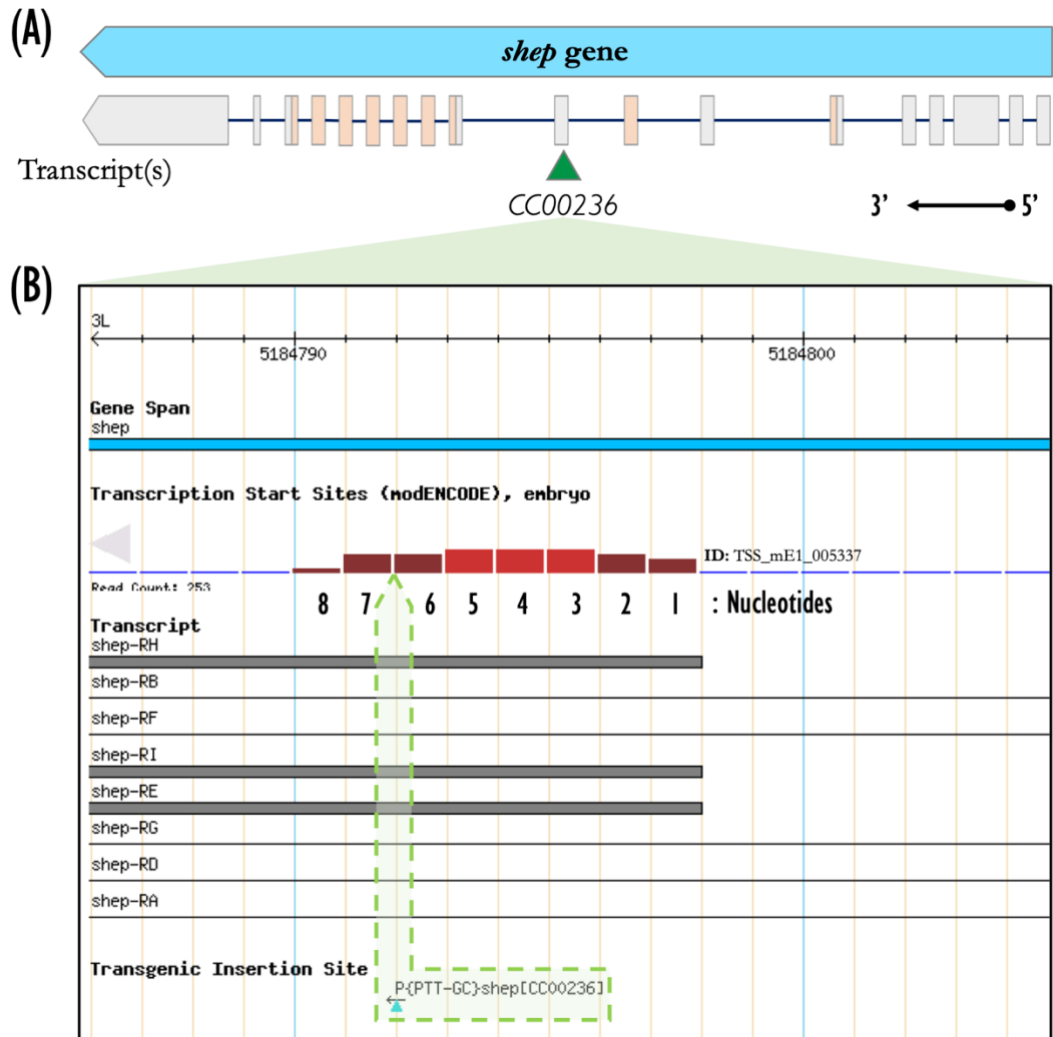


Figure 6.3: Protein-trap insertion hints at transcriptional disruption of Shep small isoforms. (A) The gene of *shep* locus (shown in blue) and the exon-intron structure model of the eight transcripts produced are depicted based on FlyBase. The transcript model shows the exon-intron layout, representing the combined structure of individual transcripts (namely, A, B, D, E, F, G, H, and I). In this representation, exons are colour-coded to differentiate the various regions of an mRNA: the coding sequence is denoted in orange and the untranslated regions in gray. The green triangle represents the insertion site of the P-element GFP cassette within the CC00236 protein-trap line. The diagram is oriented to match the direction of transcription, as indicated by the pointed ends of the gene and transcript model. **(B)** An expanded view of the insertion site at a single-nucleotide base resolution on the *D. melanogaster* genome viewer. Four selected tracks are displayed on the GBrowse viewer: the gene span, transcription start sites (TSS) from the modENCODE embryonic dataset, transcript, and transgenic insertion site. The red bars on the TSS track represent a single nucleotide base of the TSS window (TSS_mE1_005337). In the transcript track, gray bars correspond to the exons encoding the beginning of the 5'UTR region of the *shep* transcripts -RH, -RI, and -RE. Introns associated with the other transcripts are depicted as black lines. The CC00236 protein-trap cassette is inserted into the narrow TSS window responsible for the transcription of *shep* transcripts E, H, and I, specifically between the 6th and 7th nucleotides of the transcription start site. For more details on the exact genomic location of the transposable element insertion, refer to FBti0099746 on FlyBase. Abbreviation – UTR: untranslated region.

Regardless of the mechanisms underlying the loss of the small Shep isoforms observed in the GFP protein-trap line, these findings suggest that the absence of these isoforms may account for the phenotypes detailed in Section 3.3 of Chapter 3, wherein we demonstrated that flies of the protein-trap line had smaller ovaries, fewer ovarioles, and laid fewer eggs compared with the wild-type. In line with our findings, ovary size has been shown to correlate with both the number of ovarioles and egg production (Boulétreau-Merle *et al.* 1982). Moreover, in light of the small ovary size we observed, which appears to be a direct result of fewer ovarioles, it is worth noting that a body of evidence supports the notion that the number of ovarioles is established during larval development in the early gonads (Sarikaya *et al.* 2012; Sarikaya and Extavour 2015; Zhao *et al.* 2022). These studies have shown that intercellular signalling pathways, such as BMP and Hippo, intricately regulate the sorting of terminal filament cells through morphogenesis. This process plays a pivotal role in guiding the formation of distinct niches, each housing germline stem cells, which in adult ovaries become ovarioles. These findings highlight the importance of early developmental events in determining the final number of ovarioles in the adult female reproductive system. Given our preliminary results and findings from the literature, Shep makes a compelling candidate for further investigation as a potential player in early ovarian development, which ultimately influences the adult ovary.

6.2.3 Shep Isoforms Display Different Subcellular Distribution Patterns Reflecting Their Diverse mRNP Complex Associations

In the germline, the localization of GFP-trapped Shep to the oocyte cortex, particularly its enriched accumulation at the posterior pole and the dorsal-anterior corner adjacent to the oocyte's nucleus, strongly suggested an association of Shep with both the *oskar* and *gurken* mRNP complexes (Figure 3.1 & Figure 3.5). This subcellular distribution pattern is consistent with our *in vitro* and *in vivo* biochemical evidence (Figures 5.2 & 5.5). Moreover, when examining the subcellular distribution pattern of several Shep mCherry-tagged isoforms (A, C, and E), we observed that all three isoforms were enriched at the dorsal-anterior cortex of the oocyte (Figure 4.14). Interestingly, while Shep-E and -C displayed a crescent-shaped enrichment at the posterior cortex, isoform

A did not exhibit this posterior crescent pattern (**Figure 3.6**). The differences in the subcellular localization at the posterior cortex may arise from structural variations among these isoforms, as discussed in the following section.

The localization of Shep to the entire cortex of the oocyte suggests that it could mediate interactions with other mRNP complexes, beyond its interactions with *oskar* and *gurken*. In agreement with this hypothesis, using the *in vitro* RNA-affinity pulldown assay and *in vivo* RNA immunoprecipitation followed by RT-PCR, we demonstrated that Shep associates with *Notch*, *Delta*, *DE-Cadherin*, *nanos*, *bicoid*, and *K10* mRNAs. Interestingly, when *shep* levels were downregulated in the female germline using RNAi hairpins, we observed amorphic egg chambers (**Figure 4.4**). These phenotypes resembled those seen in both *Notch* and *Delta* RNAi-mediated downregulation (**Figures 3.10 & 3.11**), as well as germline clone mutants reported in the literature (López-Schier and St. Johnston 2001; Torres *et al.* 2003). Additionally, we observed a cell non-autonomous phenotype whereby RNAi-mediated downregulation of *shep* in the germline impacts border cell migration, as discussed in Section 4.1.2. Altogether, the phenotypes mentioned above, along with others discussed in this section, further point to Shep's association with these mRNAs, indicating a potential role for Shep in their cytoplasmic post-transcriptional regulation. Taking into account the exclusive cytoplasmic localization of Shep within various ovarian cell types, it is highly unlikely that Shep directly regulates transcription or is involved in post-transcriptional regulation within the nucleus.

In the ovarian soma, Shep is expressed in nearly all subpopulations of follicle cells, yet its role within these cell types remains unclear. However, given the early and pronounced expression of Shep in border cells, Shep may modulate the epithelial-to-mesenchymal transition (EMT) of these cells in manner similar to that of its mammalian ortholog, RBMS3 (Block *et al.* 2021). In breast cancer cell lines, overexpression of RBMS3 was sufficient to induce EMT, while its downregulation significantly impaired metastasis. In *Drosophila*, Shep has not yet been demonstrated to facilitate cell migration. Further exploration is required to determine whether Shep may be involved in mediating EMT within border cells.

6.2.4 Shep Isoforms Displaying a Posterior Crescent Pattern

When analyzing the subcellular distribution pattern of Shep using the GFP protein-trap line, the posterior crescent was particularly interesting (**Figure 3.1C**). This is not only because Shep was identified by our lab's yeast two hybrid screen as a potential binding partner of PTB, a negative translational regulator of *oskar* mRNA, but also because it suggests that similar to PTB, Shep may be a part of the *oskar* mRNP complex (Besse *et al.* 2009). Consistent with this localization pattern, we also observed the posterior crescent with the UAS-driven mCherry-tagged Shep-E and -C, but not with isoform A (**Figure 3.6**), demonstrating that the localization of Shep to the posterior cortex of the oocyte is isoform-specific. Considering that these mCherry-tagged transgenes only consist of the coding sequences without the untranslated regions, the differences observed in the subcellular distribution patterns of Shep are likely mediated at the protein level. In other words, the directed localization of certain mCherry-tagged Shep recombinant isoforms to the posterior pole of the oocyte is determined by the protein sequence, rather than the asymmetric localization or local translation of the mRNA. Nonetheless, this does not rule out the possibility that mechanisms involving the asymmetric subcellular localization and local translation contribute to the posterior enrichment of endogenous *shep* mRNA transcripts, thereby leading to the expression of specific Shep protein isoforms.

The N-terminal extension found in isoform A and the 11-amino acid linker located between the two RRM domains of Shep-E are the main structural differences among the three Shep isoforms studied here (**Figure 4.11A**). A possible mechanism explaining Shep's posterior accumulation might involve an amino acid motif in the N-terminal extension sequence of isoform A that prevents its posterior localization. Alternatively, the 11-amino acid linker of isoform E may contain a localization signal essential for posterior accumulation of Shep. However, given that Shep-C lacks both the N-terminal extension and the 11-amino acid linker regions, its ability to localize to the posterior cortex suggests an alternative mechanism, which remains to be elucidated.

Using both RT-PCR and Western blot analyses, we showed that the Shep GFP protein-trap line is null for the small isoforms, while isoforms A, F, and B are expressed and tagged in these ovaries (**Figure 3.3**). Therefore, the enrichment of GFP-trapped Shep proteins at the posterior pole of the oocyte suggests that the posterior crescent observed in the Shep GFP protein-trap line could be attributed to either isoform B or F, especially because our previous findings indicated that mCherry-tagged Shep-A does not localize to the posterior cortex. Since the coding sequences of both isoforms A and F are almost identical (**Figure 1.14**), the posterior crescent observed in the Shep GFP protein-trap line is most likely due to the accumulation of isoform B.

Similar to Shep-A, isoform B also has an N-terminal extension region, which is absent in Shep-E (**Figure 1.14**). Given the aforementioned reasoning, Shep-B is likely the isoform that localizes to the oocyte's posterior pole, similar to isoform E (**Figure 3.1** and **3.6D**). Notably, they both contain the 11-amino acid linker region, which is absent in Shep-A, the isoform that fails to enrich at the posterior. Therefore, by comparing the structural differences and subcellular distribution patterns of isoforms A, B, and E, it appears that the 11-amino acid linker situated between the RRM domains of Shep protein may serve as a 'posterior' localization signal, mediating the isoform-specific localization to the oocyte posterior cortex. One possible way to test this intriguing hypothesis is to artificially insert the 11-amino acid linker sequence located between the RRM domains of the Shep protein into isoform A and test whether this chimeric Shep-A protein can now localize to the posterior pole of the oocyte. Furthermore, generating a fluorescently tagged Shep-B to track its subcellular localization pattern would be beneficial, as it would help us test the hypothesis that the 11-amino acid linker, rather than the N-terminal extension region, may contribute to the posterior localization of Shep.

6.3 The Notch Signalling Pathway Regulates *shep* Expression

Among the various signalling pathways that regulate oogenesis, such as Notch, JAK-STAT, EGF, BMP, Hh, and Wg (Poulton and Deng 2007; Hinnant *et al.* 2020), we focused on how the Delta-Notch signalling pathway could influence Shep expression. We did this based on two key observations: (1) a notable temporal correlation between

Notch signalling activity during oogenesis and Shep expression in ovarian somatic cells (Section 1.4.3 and **Figure 3.5**); and (2) the amorphic egg chamber phenotypes, resulting from *shep* RNAi-mediated downregulation in the female germline, closely resembled those observed in both *Delta* and *Notch* RNAi experiments (**Figures 3.10, 3.11, and 4.4**).

Interestingly, we noticed that the Shep expression pattern seems to coincide with the two distinct rounds of follicle cell differentiation governed by Notch signalling (Xu and Gridley 2012). As a result of the Delta-Notch signalling occurring in the germarium, polar and stalk follicle cells are among the first follicle cell types to differentiate (Torres *et al.* 2003). It is also within these cell types that the first apparent signal of the Shep protein in the ovarian somatic cells was detected (see **Figure 3.5 Bii-iv** of Chapter 3). Furthermore, from mid-oogenesis onwards, Shep expression was detected throughout the entire follicle epithelium (see **Figure 3.5B**), coinciding with the second round of the Notch-Delta signalling pathway (Deng *et al.* 2001; López-Schier and St. Johnston 2001; Xu and Gridley 2012).

Given the temporal correlation between Shep expression and the activity of the Notch signalling pathway, we investigated whether the Delta-Notch signalling pathway is involved in regulating Shep expression in the follicular epithelium. Indeed, our *in vivo* induced RNAi-mediated knockdown analysis, which involved depleting the levels of either *Delta* in the female germline or *Notch* in the ovarian soma, revealed that the expression of the Shep protein in the somatic follicular epithelium is regulated by the Delta-Notch signalling pathway (as detailed in Section 3.5 of Chapter 3).

Our analysis, however, did not determine whether the Delta-Notch signalling pathway regulates *shep* expression either directly by affecting its transcription – as it is the case with Notch target genes – or indirectly by altering the levels of regulators of *shep* mRNA translation or stability (**Figure 6.4**). This indirect regulation hypothesis is based on the assumption that, while the *shep* mRNA is transcribed, it remains either untranslated or repressed in the cytoplasm. Nonetheless, to further shed light on the most probable scenario, we examined *shep* mRNA expression using publicly available single-cell

transcriptome data and analyzed the *shep* gene region for the presence of Notch *cis*-regulatory elements.

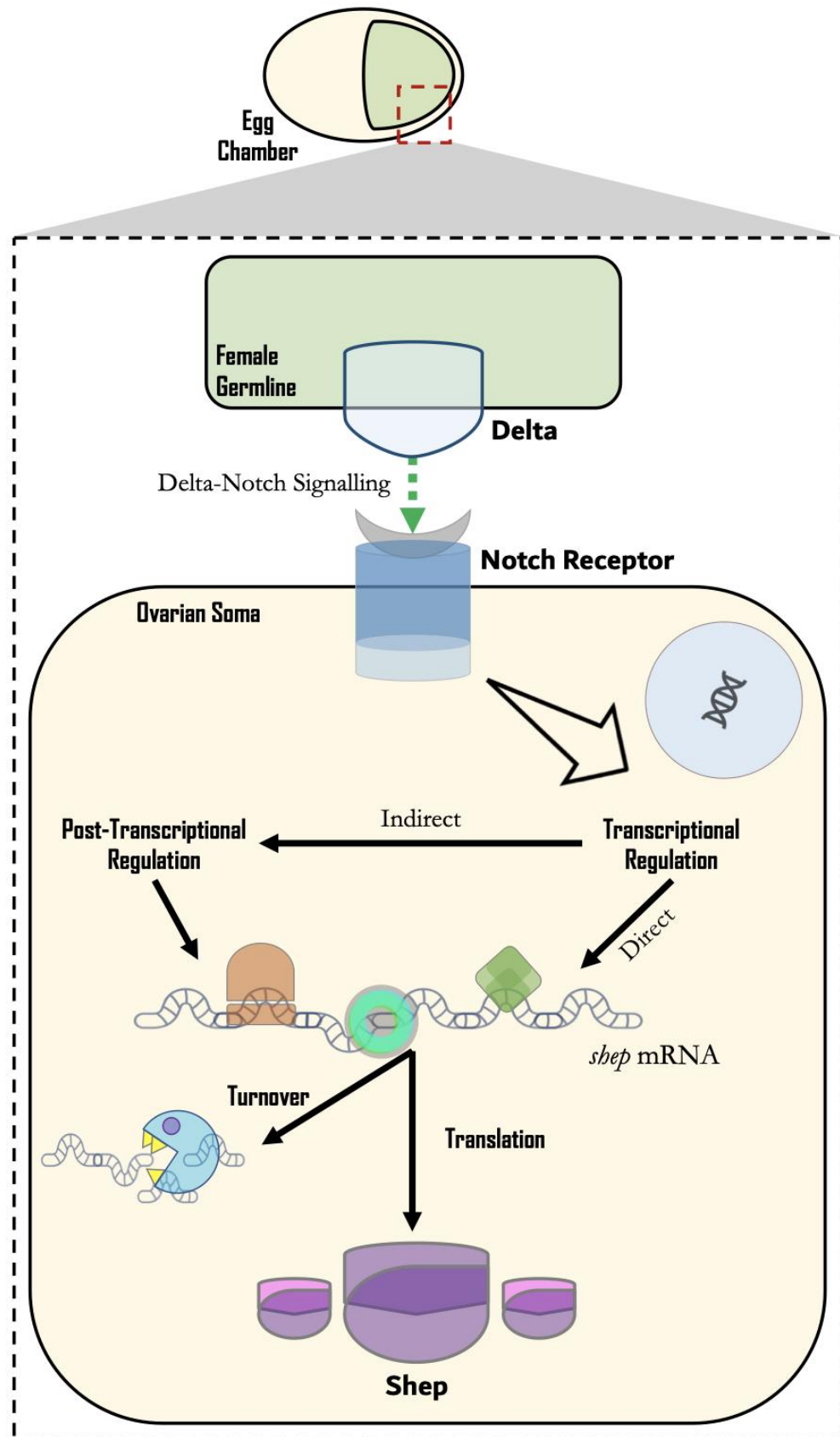


Figure 6.4: Model of the Delta-Notch signalling pathway regulating *shep* gene expression in the ovarian soma. During *D. melanogaster* oogenesis, Delta presented on the cell surface of the 16-cell germline syncytium activates Notch signalling in the overlaying somatic cells, inducing their differentiation. RNAi-mediated depletion of Delta in the germline or Notch receptor in the ovarian soma revealed that the Delta-Notch signalling pathway affect the Shep protein expression in cells of the follicular epithelium. This regulation could result from either a direct effect on *shep* gene transcription or an indirect effect by altering the expression levels of regulators influencing *shep* mRNA translation or stability.

6.3.1 *shep* Represents a Novel Target Gene of the Notch Signalling Pathway

In an effort to provide experimental evidence lending support to either of our previously proposed hypotheses that the Delta-Notch signalling pathway regulates *shep* expression in somatic ovarian cells, we analyzed the adult *Drosophila* ovary transcriptome using single-cell RNA-seq data from the Fly Cell Atlas project (Li *et al.* 2022). Interestingly, while *shep* transcripts are sparsely detected in some ovarian somatic cells during the early stages of oogenesis, their expression within the follicular epithelium predominantly peaks from mid-oogenesis onwards, specifically at stages 5/6 (see regions outlined in purple and yellow in **Figure 6.5**). This expression pattern indicates that *shep* transcription in ovarian somatic cells becomes increasingly active at these stages. Notably, the transcriptional activity of *shep* during oogenesis described here aligns with the expression pattern of the Shep protein that we characterized using the GFP protein-trap line (as detailed in Section 3.4.3 of Chapter 3). Given the somatic expression profile of the *shep* mRNA, it seems likely that the Delta-Notch signalling pathway regulates *shep* expression at the transcriptional level, rather than through an indirect mechanism affecting the cytoplasmic fate of the *shep* mRNAs. Further experimental validation is required to confirm the proposed model.

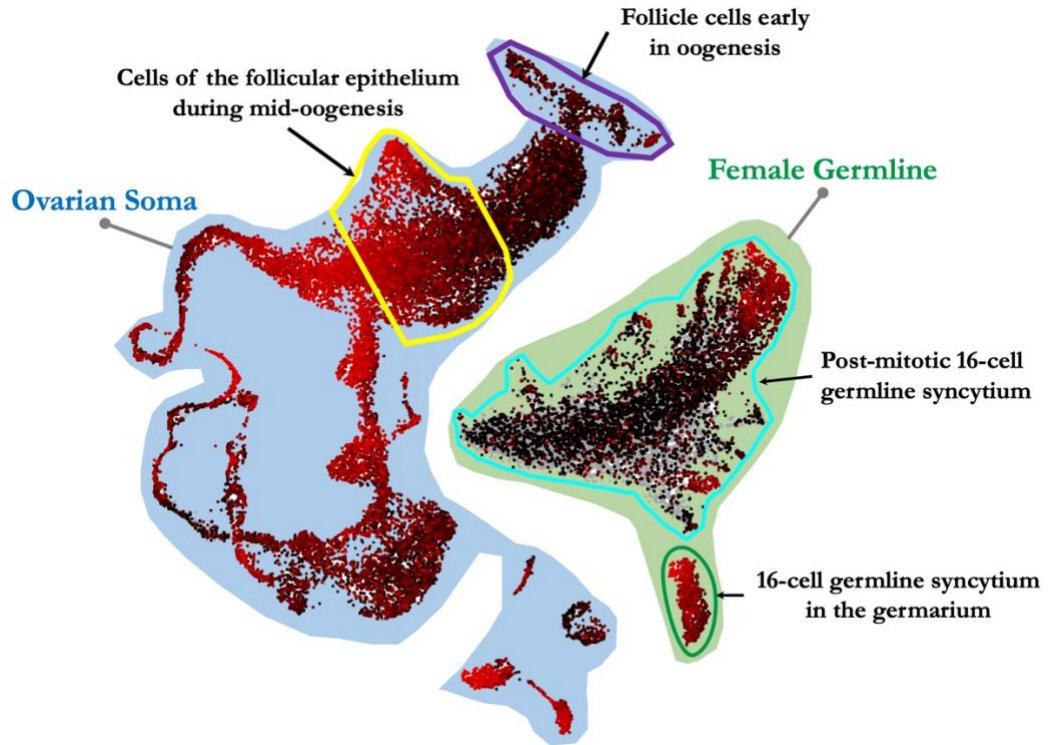


Figure 6.5: Single-cell transcriptome profile of *shep* in adult fly ovaries. The UMAP plot visualizes the *shep* mRNA expression profile within the ovary, derived from the Fly Cell Atlas single-cell RNA-seq dataset. Each data point (dot) on the plot represents the transcriptome profile from a single nucleus: red dots indicate cells that express *shep* at high levels, while black dots indicate those that express low levels of *shep*, and gray dots indicate the absence of *shep* expression. Single-cell RNA-seq profiles obtained from female germline cells are highlighted in green and those from ovarian somatic cells are highlighted in blue. Nuclei from somatic cells during early and mid-oogenesis are outlined in purple and yellow, respectively. According to the current annotations, *shep* expression in the female germline is divided into two clusters: the early 16-cell germline syncytium situated in regions 2a and 2b of the germarium (outlined in dark green), and the late post-mitotic 16-cell germline syncytium within developing egg chambers (outlined in cyan). The ‘annotated ovary loom stringent 10x’ dataset from the Fly Cell Atlas project (Li *et al.* 2022) was visualized using Scope, accessible at https://scope.aertslab.org/#/FlyCellAtlas/*/welcome. Abbreviation – UMAP stands for Uniform Manifold Approximation and Projection.

To further strengthen our proposition that the Delta-Notch signalling pathway may regulate the transcription of *shep* in ovarian somatic cells, we examined the *shep* gene locus for the presence of *cis*-regulatory elements that mediate the transcriptional activation of Notch target genes.

In *Drosophila*, the nuclear effector Suppressor of Hairless (Su(H)) is a DNA-binding transcriptional regulator that function as a key downstream component of the Notch intercellular signalling cascade (Bray and Furriols 2001). As the primary effector of the

Notch pathway, Su(H) targets specific genes by binding to the ‘GIGRGAR’ motif (Morel and Schweisguth 2000; Stroebel and Erives 2016). In the absence of Notch signalling, Su(H) recruits various co-repressors to repress its target genes. Upon activation of the Notch signalling pathway, however, the intracellular domain of the Notch transmembrane receptor undergoes proteolytic cleavage and nuclear translocation, transducing the intercellular signal to the nucleus (Xu and Gridley 2012). Once inside the nucleus, the Notch intracellular domain interacts with Su(H), causing a switch in the transcriptional activity of Su(H), thereby activating the transcription of Notch-responsive genes by recruiting distinct co-activator complexes (Bray and Furriols 2001; Lai 2002).

Therefore, we analyzed the *shep* gene locus for the presence of the ‘GIGRGAR’ motif, which is recognized by the conserved transcriptional co-repressor, Su(H). Interestingly, our examination revealed that the extended gene region of *shep* contains a total of 26 potential sites for Su(H) binding (Figure 6.6). Notably, a minimum of two, if not more, Su(H)-binding sites are situated upstream of each of the six transcription start sites identified by FlyBase within this genomic region. The location of these Su(H)-binding sites suggests their potential role in regulating the transcription of the *shep* gene, further supporting the idea that the Delta-Notch signalling pathway regulates the transcription of the *shep* gene.

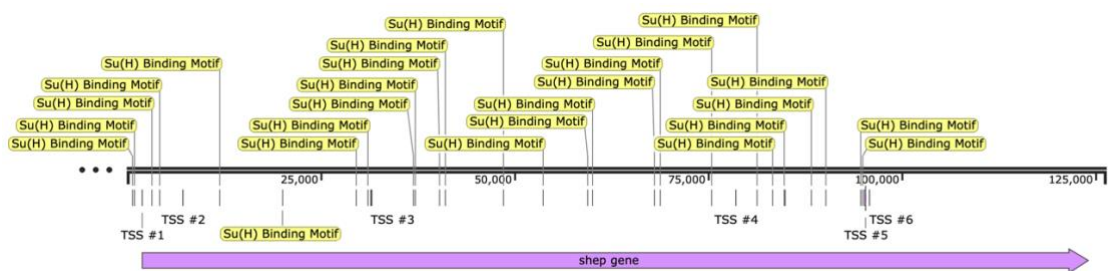


Figure 6.6: Potential binding sites for Su(H) in the *Drosophila shep* gene. Schematic representation showing the six transcription start sites (TSS #1-6) within the extended gene region of *shep*, which encompasses the *shep* gene region (depicted in purple) and an additional 2 Kb both upstream and downstream of the gene. A total of 26 Su(H) DNA-binding sequence motifs (represented as GTGRGAR and depicted in yellow) were identified within the *shep* gene region, with each transcription start site being preceded by at least two Su(H)-binding sites, if not more. The chromosomal location of the transcription start sites was obtained from FlyBase, according to the modENCODE embryonic dataset. The direction of transcription of the *shep* gene is indicated by the direction of the arrowhead. The extended gene region of *shep* was retrieved from

FlyBase (FBgn0052423; 126,124 bp) Abbreviations – Su(H): Suppressor of Hairless, TSS: transcription start site.

Interestingly, *shep* may not only represent a novel Notch target gene, but it is possible that Shep protein represents a new regulator of the Notch-Delta signalling pathway acting at the mRNA level. This idea is supported by both our findings from RNA-affinity pulldown and RNA-immunoprecipitation assays, which demonstrated an interaction between the Shep protein and both *Delta* and *Notch* mRNAs (**Figures 5.2 & 5.5**). Additionally, the amorphic egg chamber phenotype observed from RNAi-mediated *shep* downregulation closely resembles those of defective Notch signalling (**Figures 3.10, 3.11, & 4.4**). Given these observations, it is compelling to consider whether Shep protein levels reciprocally modulate the activity of the Notch signalling pathway by a feedback loop mechanism in which Shep post-transcriptionally regulates *Delta* and *Notch* mRNAs in the cytoplasm, affecting their translation or turnover.

6.4 Shep Levels in the Germline Affect Border Cell Migration

The border cell migration defects observed when Shep levels were downregulated by the induction of RNAi hairpins suggest that Shep functions in the female germline to regulate the migratory process of border cells. In our *shep* RNAi experiments the border cell cluster completed migration, but one or several cells detached and lagged behind during migration, suggesting a problem with the adhesion and/or traction required for migration to occur. There are two possible scenarios by which the depletion of Shep levels in the germline might lead to this cell non-autonomous phenotype: either through an intercellular signalling pathway or by differential expression of adhesion molecules between the 16-cell germline cyst and the migrating border cell cluster.

Given that we have demonstrated a direct interaction between Shep protein and both *Notch* and *Delta* mRNAs using RNA-affinity pulldown assays, and have also provided *in vivo* evidence of Shep's association with their mRNP complexes, it is plausible that Shep post-transcriptionally regulates these mRNAs to modulate the activity of the Notch signalling pathway. This hypothesis is particularly attractive considering that the Delta-

Notch signalling pathway is essential for normal border cell migration and is activated in border cells during migration (Wang *et al.* 2007).

Another interesting possibility is the differential expression of DE-Cadherin molecules between the nurse cells and the migrating border cell cluster. The *Drosophila* DE-Cadherin, encoded by the gene *shotgun* (*shg*), is a transmembrane protein crucial for the proper movement, cohesion, and adhesion of the border cell cluster over the nurse cells. Altering the levels of DE-Cadherin in either the border cells or the germline cells disrupts the migration of the border cell cluster across the surface of germline cells toward the anterior face of the oocyte (Niewiadomska *et al.* 1999). Using an RNA-affinity pulldown assay, we have demonstrated that Shep directly interacts with both untranslated regions of the *Drosophila* *DE-Cadherin* mRNA. In fact, Shep binds more strongly to the 5'UTR of *DE-Cadherin* than to its 3'UTR. Additionally, the two predicted Shep-binding sites we found in the 5'UTR of *DE-Cad* mRNA (Appendix 5) are both located upstream of a Hrp38 binding site (CAGGGCGCGCACUGUACGAG), which is essential for IRES-dependent translation of *DE-Cadherin* (Ji and Tulin 2012). Therefore, it is possible that Shep post-transcriptionally regulates *DE-Cadherin* mRNA expression in an IRES-dependent manner, thereby disrupting border cell migration. Interestingly, Chen *et al.* (2018) demonstrated that *shg* mRNA levels were upregulated in cultured *Drosophila* BG3 cells following *shep* depletion using dsRNA directed against *shep*. In light of our data and the existing literature, it is possible that the border cell phenotype observed upon RNAi-mediated downregulation of *shep* in the female germline alters the adhesion properties on the nurse cell surface by changing the levels of DE-Cadherin, thereby impacting the migration of the border cell cluster.

Another phenotype unrelated to border cells, which lends support to the idea that Shep levels in the germline might affect adhesion molecules, comes from the amorphic egg chambers observed upon downregulating *shep* in the germline using RNAi hairpins (Figure 4.4). During the early stages of oogenesis, in the germarium, DE-Cadherin levels are upregulated, especially within the oocyte, which is crucial for its posterior positioning within the egg chamber (Godt and Tepass 1998b). The mis-positioning of

the oocytes within our amorphic egg chambers (**Figure 4.4C**) hints at a disruption of the DE-Cadherin-mediated posterior positioning of the oocyte, potentially caused by the depletion of Shep levels in the germline. Interestingly, Ji and Tulin (2012) observed a similar oocyte mislocalization phenotype when the IRES-dependent translation of DE-Cadherin was disrupted, leading to decreased DE-Cadherin levels.

Taken together, our findings in conjunction with those from the literature point to a role for Shep in the post-transcriptional regulation of *DE-Cadherin*. Specifically, Shep may function as a translational regulator, controlling DE-Cadherin levels in the ovary, especially within the germline, leading to the observed cell non-autonomous defects.

6.5 Shep & PTB are Part of Similar mRNP Complexes in the Ovary

PTB is a multifunctional nucleocytoplasmic protein whose post-transcriptional regulatory roles have been extensively studied (Romanelli *et al.* 2013). In the *Drosophila* ovary, PTB localizes to the cytoplasm of the oocyte and to the nucleus of both nurse cells and follicle cells (Besse *et al.* 2009). Based on our *in vivo* characterization, Shep primarily localizes to the cytoplasm of both the oocyte and follicle cells. Interestingly, both Shep and PTB exhibit a distinct subcellular localization pattern within the oocyte cytoplasm at the posterior cortex (**Figures 3.1**), which coincides with the localization pattern of *oskar* mRNA. This posterior crescent, a hallmark of proteins involved in regulating the *oskar* mRNP complex, suggests either a protein-protein interaction between Shep and PTB or their association with common mRNAs. In fact, our RNA-affinity pulldown assays have shown that both Shep and PTB interact with not only *oskar* mRNA, but also with other key maternal mRNAs, including *Delta*, *Notch*, *shotgun* (*Drosophila* DE-Cad), *gurken*, *bicoid*, *nanos*, and *K10*.

In addition to associating with similar mRNP complexes, it is worth noting that both Shep and PTB produce phenotypes in the ovary that are reminiscent of each other. Specifically, downregulation of either protein leads to defects in border cell migration (**Figure 4.5** and López de Quinto, unpublished). Moreover, both PTB and Shep interact with and post-transcriptionally regulate *gurken* mRNA expression at the dorsal anterior corner (**Figures 4.15** and **5.2**; McDermott and Davis 2013). This local

translation of *gurken* within the oocyte is crucial for the proper specification of follicle cells and the correct positioning of dorsal-anterior chorion structures, including the dorsal appendages. Consequently, aberrant levels of either PTB or Shep have been shown to impact the patterning and morphogenesis of the dorsal appendages (**Figure 4.13**; McDermott and Davis 2013). This biological readout and the phenotypic resemblances between Shep and PTB, in conjunction with their associations with common mRNP complexes, suggest that these proteins may regulate similar post-transcriptional processes in the cytoplasm and potentially work redundantly with one another.

Through both loss-of-function and overexpression functional analyses, we showed that Shep post-transcriptionally regulates the expression of both *gurken* and *oskar* mRNAs in the female germline (**Figures 4.15 & 4.17**). Interestingly, these findings correlate with those reported for PTB in the oocyte, where it regulates *gurken* mRNA localization at the dorsal-anterior cortex and suppresses *oskar* translation before the mRNA reaches the posterior pole (Besse *et al.* 2009; McDermott and Davis 2013). These results collectively provide evidence suggesting a potential direct interaction between Shep and PTB. However, when we investigated the protein-protein interaction between Shep and PTB using an *in vivo* co-immunoprecipitation assay, we found that PTB immunoprecipitated with the GFP-trapped Shep protein in an RNA-dependent manner (**Figures 5.3 & 5.4**). The co-immunoprecipitation assay findings indicate that Shep and PTB interact indirectly by associating with the same mRNAs, which highlights the importance of RNA molecules as potential docking platforms for mediating protein-protein interactions and emphasizes the role of mRNAs beyond just encoding proteins. It also highlights the significance of RNA molecules in coordinating the assembly of sophisticated mRNP complexes.

The Shep GFP protein-trap line is a versatile tool that is powerful for the *in vivo* characterization of gene expression patterns and protein subcellular distributions. It is also compatible with various biochemical assays, including RNA-affinity pulldown, RNA immunoprecipitation, and co-immunoprecipitation. In fact, we used GFP-trapped Shep to co-immunoprecipitate PTB and to test the direct interaction between

them. However, a key limitation of our approach is that, in the CC00236 protein-trap line, the GFP-cassette insertion within the *shep* gene leads to the tagging of only the large isoforms of Shep. Additionally, because of the specific location of the insertion, these flies are null for the small isoforms, as demonstrated by our western blot analysis. Given this limitation and considering our isoform-specific distribution analysis – which shows mCherry-tagged Shep-E consistently exhibiting an enrichment at the oocyte's posterior cortex – it is plausible that the small isoforms of Shep could be the ones physically interacting with PTB at the protein level. Further tests are needed to determine whether the direct interaction between Shep and PTB is isoform-specific, especially if the small isoforms of Shep could physically interact with PTB.

A potential experimental strategy to confirm the protein-protein interaction between Shep and PTB would be to conduct an *in vitro* co-immunoprecipitation assay using exogenously expressed and purified recombinant proteins. Because PTB fusion proteins individually tagged with either GST or MBP epitopes in their purified forms are already available in the lab (Besse *et al.* 2009), it would be prudent to prioritize expressing different Shep isoforms in *E. coli* for their subsequent purification. This is something that we have attempted but could not complete due to time constraints (as outlined in Section 2.2.3 of the 'Materials & Methods' Chapter). Here, we have chosen to use the Spot tag, a novel tagging system from ChromoTek, to generate fusion proteins for Shep-E and other isoforms (A and C). This novel 12-amino acid peptide tag was specifically designed for high-affinity purification using commercial single-domain nanoantibodies, making it suitable for *in vitro* co-immunoprecipitation assays.

An alternative strategy to investigate the dynamic relationship between Shep and PTB would have been to immunoprecipitate the endogenous Shep protein from wild-type control ovaries, rather than using Shep GFP protein-trap ovaries and anti-GFP antibodies. Alternatively, we could repeat our co-immunoprecipitation experiments with ovarian extracts expressing the individual mCherry-tagged Shep isoforms, using the mCherry protein as the biochemical tag.

6.6 Regulatory Role of Shep as a Component of *oskar* & *gurken* mRNP Complexes

The CC00236 GFP protein-trap line showed cortical expression of the Shep protein in oocytes during the mid- to late stages of oogenesis (**Figure 3.5**). This observation aligns with what was briefly mentioned by Buszczak *et al.* (2007), who generated the protein-trap we used throughout our study. Interestingly, in their Carnegie Protein Trap Library collection, they observed that a protein trap in the gene CG32423 (i.e., Shep) displayed a posterior localization pattern, a pattern they also noted in other genes, such as EIF-4E and Tral. Remarkably, Olesnický *et al.* (2018) demonstrated that in the nervous system, Shep interacts not only with the RNA-binding proteins Tral and Bel (regulator of EIF-4E) but also with Yps. During oogenesis, all of these proteins exhibit a crescent pattern at the oocyte's posterior pole, similar to *oskar* (**Figure 3.1**; Buszczak *et al.* 2007; Yarunin *et al.* 2011). Additionally, Bel, EIF-4E, and Yps exert either direct or indirect translational control over the *oskar* mRNP complex (Mansfield *et al.* 2002; Yarunin *et al.* 2011). However, while Tral accumulates at the posterior pole during stages 9–10, it appears to have no effect on *oskar* expression (Wilhelm *et al.* 2005).

The *oskar* mRNA is normally translationally repressed until it reaches the posterior pole by late stage 8 or early stage 9 (Gunkel *et al.* 1998; Yano *et al.* 2004; Besse *et al.* 2009). However, in our Shep loss-of-function analysis using different trans-heterozygous combinations of deficiencies and *shep* P-element insertion mutants, we observed a visible dot of Oskar protein aggregation within the oocyte's cytoplasm far from the posterior cortex, especially from mid-oogenesis onwards (stages 6–10). These phenotypes shown in **Figures 4.9** and **4.10** are indicative of an ectopic, premature translation of *oskar* mRNA. Interestingly, a similar *oskar* phenotype has been demonstrated using loss-of-function mutants for PTB, a translational repressor of *oskar* (Besse *et al.* 2009). This is particularly intriguing, given that Shep was identified as a potential binding partner of PTB in a yeast two-hybrid screen (López de Quinto, unpublished). In light of our observations that the translational repression of *oskar* mRNA may be prematurely relieved in various *shep* loss-of-function mutant backgrounds, our results suggest that Shep may function as a repressor of *oskar* translation. This repression could be achieved through directly controlling the *oskar*

mRNA fate, or by modulating either the action of PTB itself or other post-transcriptional regulatory factors part of the *oskar* mRNP complex.

Similarly, both Krauss *et al.* (2009) and Zimyanin *et al.* (2007) observed the ectopic translation and aggregation of Oskar protein, which appeared as a dot in the oocyte's cytoplasm, even before the mRNA localized to the posterior cortex. They also observed that in some instances, the Oskar protein no longer forms a tight crescent at the posterior pole of stage 9 oocytes. Instead, it appears either as a cloud or a dot near the posterior cortex (**Figure 4.10** in Chapter 4). Together, these findings imply an interplay between Myosin-V and Kinesin-1 that prevents ectopic accumulation of *oskar* mRNA in the cytoplasm, outside the oocyte's posterior cortex. Additionally, they found that the Oskar protein plays an active role in the reinforcement of the oocyte's polarity by recruiting both microtubule plus ends and Par-1 to the posterior cortex. Given these findings, it is therefore possible that the *oskar*-related phenotypes observed in the Shep trans-heterozygous mutants may be a consequence of defective transport machinery, improper localization of *oskar* mRNA, or both. Despite this, the exact mechanism underlying the ectopic expression of Oskar protein outside the posterior cortex from stage 9 oocytes onwards remains to be fully characterised.

The novel role of Shep as a translational repressor of the *oskar* mRNP complex we proposed is based on the complementary results obtained in the loss-of-function and overexpression experiments. When Shep was overexpressed in the female germline, we observed a reduction in both isoforms of the Oskar protein at the oocyte's posterior pole (**Figures 4.17 & 4.18**), with no significant change in *oskar* mRNA levels (**Figure 4.19**). In tandem with these findings, we also showed a direct interaction of the various Shep isoforms with the 3'UTR of *oskar* mRNA (**Figure 5.2A** in Chapter 5). Interestingly, Shep binds exclusively to the 3'UTR of *oskar*, as opposed to other tested mRNAs where binding was detected in regions beyond the 3'UTR, such as the 5'UTR and coding sequence (**Figure 5.2**). This, combined with the observation that the predicted Shep-binding sites on *oskar* mRNA are AU-rich sequences, altogether lend further support to our proposed novel role for Shep as a post-transcriptional regulatory component of the *oskar* mRNP complex during oogenesis, potentially controlling its

translation. We propose that Shep acts as a translational suppressor by binding to AU-rich sequence motifs situated in the 3'UTR of *oskar* mRNA and potentially other mRNA targets (Figure 6.7). However, the underlying mechanism of its role as a translational suppressor has yet to be elucidated.

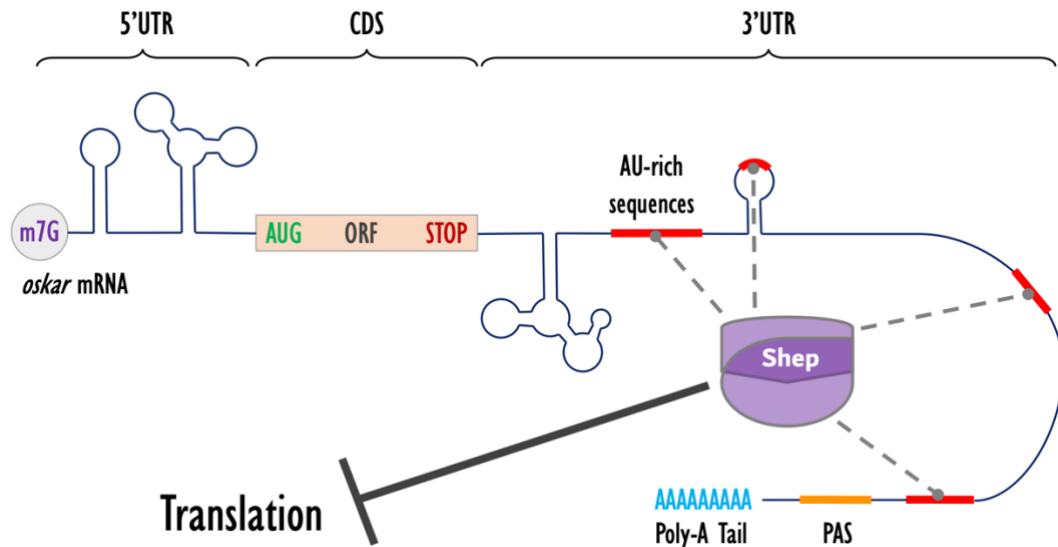


Figure 6.7: Translational repression of the *oskar* mRNP complex by the cytoplasmic Shep.

A schematic presentation illustrating a proposed model for *Drosophila* Shep mechanism of action as a post-transcriptional regulatory component within the *oskar* mRNP complex. This model highlights Shep protein interactions with AU-rich sequences at the 3'UTR, which facilitates the repression of *oskar* mRNA translation during oogenesis. Abbreviations – 5' and 3' untranslated region (5'- and 3'-UTR). m7G: 5'-7-methylguanosine cap. AUG (written in green): start codon of the open reading frame (ORF, written in dark gray). STOP (written in red): stop codon of the Oskar protein ORF. PSA (shown in orange): polyadenylation signal. Poly-A Tail: polyadenylic acid tail.

Parallelisms from non-*Drosophila* orthologs can be drawn to support our proposed model. For example, the human Shep homolog RBMS3 inhibits the translation of *twist1* mRNA in breast cancer by directly binding to the 3' untranslated region of the mRNA (Zhu *et al.* 2019). Similarly, the ortholog of Shep in *C. elegans*, Sup-26, negatively regulates the translation of the sex-determining gene *tra-2* by directly binding to *cis*-regulatory elements found in the 3'UTR of its mRNA (Mapes *et al.* 2010). These conserved mechanisms may provide insight into the potential function of Shep in regulating *oskar* mRNA.

With regards to *gurken*, we showed that the overexpression of various Shep isoforms disrupts the development of the egg's dorsal appendages (Figure 4.13). Upon further

examination, we observed a reduction in the Gurken protein at the dorsal-anterior corner of the oocyte by immunostaining (**Figure 4.15**). This specific corner corresponds to the region where follicle cells receive Gurken signalling from the underlying oocyte, instructing them to give rise to the dorsal appendages. Moreover, our RT-qPCR analysis showed no significant change in *gurken* mRNA levels upon the overexpression of Shep (**Figure 4.16**). Given these results, the apparent reduction in Gurken protein can be attributed to one of two scenarios in which Shep post-transcriptionally regulates the *gurken* mRNP complex: either through translational repression or due to improper localization of the mRNA. Further experimental investigations are required to address how Shep modulates *gurken* expression within the oocyte.

A loss-of-function mutant affecting the Sqd-associated protein, K10, causes a dorsalization phenotype of the eggshell due to ectopic Gurken protein expression from the mislocalized *gurken* mRNA (Serano *et al.* 1995; Saunders and Cohen 1999; Johnstone and Lasko 2001). Importantly, since we have demonstrated that Shep binds to the 3'UTR of *K10* mRNA (**Figure 5.2D**), it is possible that the Shep protein may indirectly influence *gurken* expression by modulating the levels of K10 within the oocyte's cytoplasm.

Collectively, our preliminary data, obtained from combining various *shep* P-element insertions with different deficiencies on the third chromosome, indicate that while *shep* loss-of-function leads to an ectopic premature translation of *oskar* and disrupts its tight localization at the oocyte's posterior pole, while the expression and localization of *gurken* mRNA at the dorsal-anterior cortex appear unaffected (**Figure 6.8**). These observations support the notion that Shep's regulation of mRNA cytoplasmic fate varies based on its associated mRNP complex. Within the *oskar* mRNP complex, Shep may function as a translational repressor, while in the *gurken* mRNP complex, it may influence a different aspect of post-transcriptional regulation. In addition to analyzing Shep RRM mutants that disrupt its RNA-binding capacity (see Section 6.9), it would be insightful to investigate whether the Shep overexpression phenotypes can be rescued using RNAi hairpins targeting *shep* in the female germline.

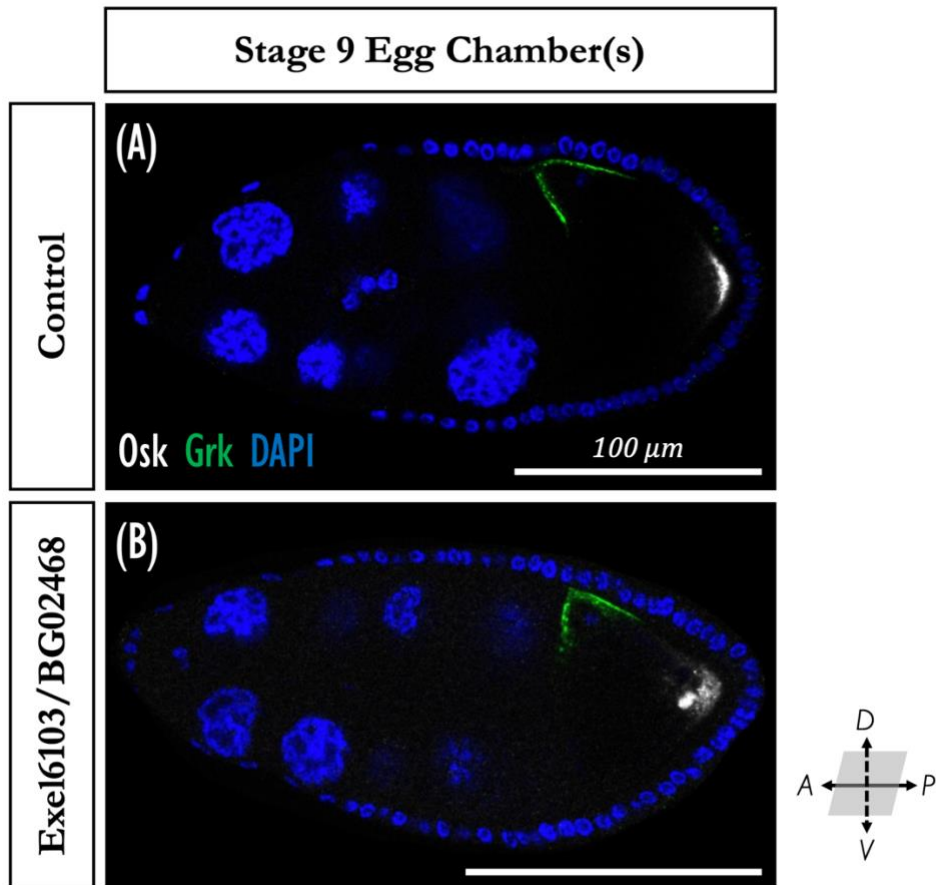


Figure 6.8: *shep* loss-of-function backgrounds dysregulated *oskar* expression but not *gurken*. (A) In control ovaries, the Oskar protein (shown in white) accumulates exclusively at the posterior cortex of stage 9 oocytes, while Gurken protein (shown in green) is localized to the dorsal-anterior corner. (B) In ovaries with a *shep* P-element insertion in the Exel6103 deficiency background, the Oskar protein appears as a dot near the oocyte's posterior cortex and no longer retains its normal tight crescent shape. However, the expression of the Gurken protein appears unaffected. Nuclei are marked by DAPI staining (blue). Scale bar, 100 μm. Fly genotype(s) – Panels A (*w*; *Df(3L)^{Exel6103}/+*) and B (*w*; *Df(3L)^{Exel6103}/shep^{BG02468}*).

Finally, despite our extensive efforts, all attempts to obtain an *oskar*-related phenotype by RNA-mediated downregulation of *shep* in the female germline proved unsuccessful (Table 4.1). The lack of any *oskar* dysregulation can be attributed to two key possible factors: (1) the presence of an arsenal of functionally redundant proteins within the *oskar*mRNP complex, which ensures its proper localization and local translation at the posterior cortex of the oocyte (Kugler and Lasko 2009); (2) given that Shep protein is expressed in germline stem cells (Figure 3.5Bi), its half-life should be taken into account, as these early and pre-assembled mRNP complexes may already be sufficient to ensure proper post-transcriptional regulation of *oskar* mRNA during oogenesis. Therefore, regardless of the Gal4 driver we use to express *shep* RNAi hairpins in the

germline, it is not possible to downregulate Shep levels in those early germline stem cells. In general, tight regulation of key maternal mRNAs within the oocyte is essential to ensure precise spatial-temporal translation of proteins required for embryonic patterning and development (Kuersten and Goodwin 2003). To the best of our knowledge, post-transcriptional regulation defects related to key maternal mRNAs in the *Drosophila* ovary, as reported in the literature, resulted from either overexpression or loss-of-function mutants, not from RNAi-mediated downregulation (Van Buskirk *et al.* 2000; Bullock and Ish-Horowicz 2002; Yano *et al.* 2004; Zimyanin *et al.* 2007; Besse *et al.* 2009; Krauss *et al.* 2009; McDermott and Davis 2013; Nieuwburg *et al.* 2017).

6.7 Potential Regulatory Role of Shep as a Component of the *nanos* mRNP Complex

During early oogenesis, Nanos facilitates the establishment and maintenance of germline stem cells (GSC) within the germarium (Parisi and Lin 2000; Wang and Lin 2004). Additionally, the rapid downregulation of Nanos within the GSC's committed daughter cells aids in their differentiation and entry into oogenesis. A collection of translational regulators, including the Bam and Sxl proteins, repress the expression of *nanos* by interacting with various *cis*-acting elements located within its 3'UTR (Li *et al.* 2009; Chau *et al.* 2012; Malik *et al.* 2020). Given the observed expression of Shep in the GSC (**Figure 3.5Ai**) and its interaction with the *nanos* 3'UTR (**Figure 5.2D**), it would be interesting to further explore whether Shep could be a novel component of the *nanos* mRNP complex, potentially recruited to post-transcriptionally regulate its expression. If proven true, there could be two possible mechanisms at play. First, Shep may maintain GSC identity by promoting *nanos* translation in a manner similar to its mouse ortholog. Notably, RBMS3 facilitates the translation of the key pancreatic transcription factor *ptf1a* mRNA by binding to its 3'UTR (Lu *et al.* 2012). Second, although Shep could function as a negative translational regulator of *nanos*, it is functionally repressed in GSCs and awaits derepression in germline-committed daughter cells to promote differentiation.

The research group that previously showed Shep as a direct interactor and negative regulator of the gypsy chromatin insulator complex in the *Drosophila* central nervous

system (CNS; Matzat *et al.* 2012) has also identified Rumpelstiltskin (Rump), a hnRNP M homolog, as an antagonist of gypsy chromatin insulator enhancer-blocking and barrier activities in non-CNS tissues (King *et al.* 2014). Interestingly, Rump associates with the *cis*-acting localization signal located in the 3'UTR of *nanos* mRNA during oogenesis, and this association is essential for the localization of *nanos* mRNA at the oocyte's posterior cortex (Jain and Gavis 2008). Given this functional similarity across diverse *Drosophila* tissue types and our RNA-affinity pulldown assay results, which revealed a direct interaction between Shep and the *nanos* 3'UTR *in vitro*, it would be interesting to explore whether Shep operates redundantly to Rump in modulating *nanos* mRNA posterior localization.

6.8 Shep is Not Nuclear in any of the Ovarian Cell Types

The GFP protein-trap line showed that Shep is primarily cytoplasmic across various ovarian cell types. However, given the caveat that the CC00236 GFP protein-trap flies are null for the small Shep isoforms, we expanded our approach to better characterize Shep's subcellular distribution by including both *in vivo* tagging of different isoforms (Shep-A, -C, and -E) and immunostaining of the endogenous protein. Both methods confirmed our findings from the protein-trap line analysis, suggesting that Shep is predominantly cytoplasmic, rather than nuclear, within the *D. melanogaster* ovary. Interestingly, our observations of Shep's cytoplasmic pattern align with the findings of Chen *et al.* (2014). In their study, they demonstrated that both *shep* mRNA and protein are localized to the cytoplasm of stage-8 oocytes using *in situ* hybridization and immunostaining, respectively. This was primarily done to test the specificity of their *in situ* probes and the polyclonal anti-Shep antiserum for their potential application in the nervous system.

Beyond Shep's documented nuclear localization (Chen *et al.* 2018), isoform E in particular has also been characterized for its interaction with core components of the gypsy chromatin insulator complex, which plays a role in regulating gene expression in the nervous system (Matzat *et al.* 2012; Chen *et al.* 2019). The same research group demonstrated that when Shep was overexpressed in *Drosophila* muscle tissue using the Mef2-Gal4 driver, insulator activity was successfully repressed (Matzat *et al.* 2012).

These findings suggest not only that Shep's role as a negative regulator of insulator activity is specific to cells of the central nervous system, but also that there is a concentration-dependent threshold of Shep protein required for its nuclear role. Interestingly, however, when we overexpressed Shep-E in various ovarian cell types, Shep did not localize to the nucleus. Importantly, the mechanisms underlying the shuttling of Shep between the cytoplasm and nucleus remain unclear, especially its import to the nucleus. Additionally, its nuclear localization signal sequence has yet to be identified.

Nucleocytoplasmic trafficking of Shep: The nuclear localization signal (NLS) is typically a stretch of amino acid residues within proteins that acts as a molecular cue, mediating the transport of proteins from the cytoplasm to the nucleus (Cokol *et al.* 2000; Lange *et al.* 2007). Members of the importin superfamily are crucial components of the nuclear transport mechanism, as they enable the recognition and transportation of cargo proteins across the nuclear envelope through the nuclear pore complex (NPC; (Kosugi *et al.* 2009; McLane and Corbett 2009). In general, nuclear import of cargo proteins is initiated by the formation of a ternary complex with members of importin superfamily (Kosugi *et al.* 2009; Kimura and Imamoto 2014). In the classical importin pathway, importin- α and importin- β 1 form a heterodimer. Importin- α recognizes the NLS on a cargo protein and functions as an adaptor protein between the cargo and importin- β to promote transportation into the nucleus (Goldfarb *et al.* 2004; Miyamoto *et al.* 2016; Oka and Yoneda 2018). Importin- β , on the other hand, mediates interactions with the nuclear pore complex to facilitate nuclear transportation across the nuclear envelope (Weis 2003; Freitas and Cunha 2013). Once importin- β 1 docks the cargo-importin complex to the NPC, the cargo is released through the binding of Ran-GTP to importin- β 1 in the nucleus (Kosugi *et al.* 2009; Oka and Yoneda 2018).

Since the NLS signal residues and mechanism of nuclear import for Shep have neither been described nor proposed in the literature, we aimed to identify a putative NLS signal sequence within the Shep protein, thereby providing novel insights into the potential mechanism behind its nuclear localization in non-ovarian cell types. We utilized a combination of sequence-based computational analyses (detailed in Section

2.7.3). Briefly, it involved searching for both classical and non-canonical NLS consensus sequences that have been described in the literature to mediate interactions with the nuclear import machinery (Romanelli and Morandi 2002; Kosugi *et al.* 2009; Freitas and Cunha 2013; J. Lu *et al.* 2021). In addition, we used NLS prediction tools employing various algorithms and patterns to identify potential NLS sequences based on characteristics such as amino acid composition, charge distribution, secondary structure prediction, and previously characterized sequences.

In our analysis of the six annotated Shep primary amino acid sequences using NLS prediction tools, as outlined in Section 2.7.3 of the ‘Materials & Methods’ Chapter, we identified a novel putative nuclear localization signal: ‘PKKKN’. This short amino acid signal is located directly downstream of the RRM2 domain in all Shep isoforms (**Figure 6.9A-B**). This predicted monopartite NLS follows the classical 4-residue pattern that necessitates the presence of 3 consecutive basic amino acids and a special case amino acid, such as proline (see underlined sequences in text).

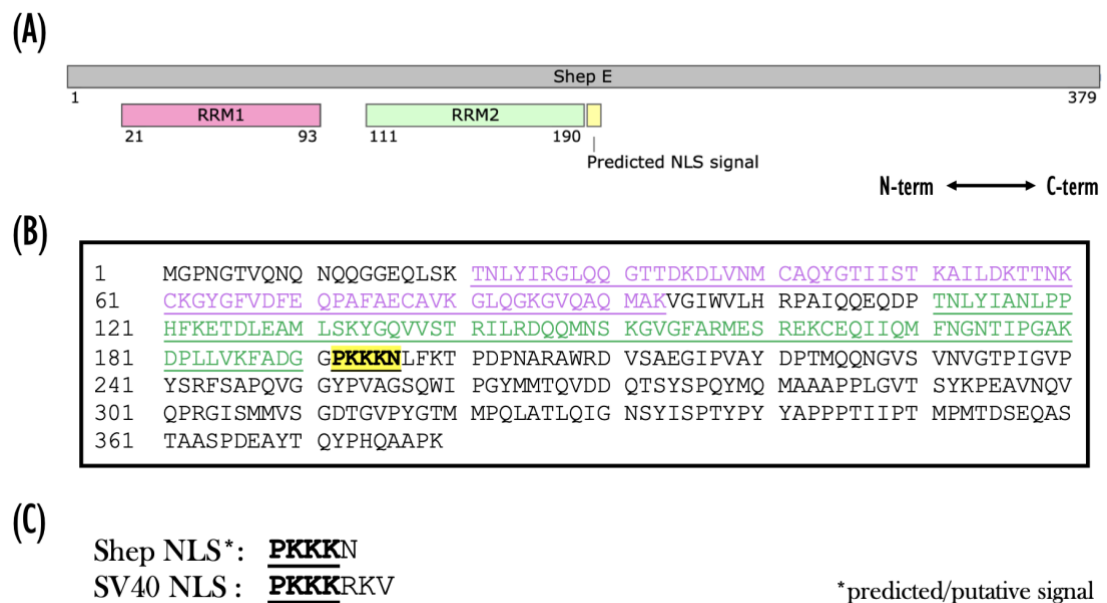


Figure 6.9: Putative nuclear localization signal sequence in Shep protein. (A) Schematic presentation of the *Drosophila* Shep isoform E domain structure. Shep-E is a 40kDa conserved RNA-binding protein of 379 amino acids with two RNA recognition motifs (RRM): RRM1 in pink (21–93aa) and RRM2 in green (111–190aa). The predicted nuclear localization signal (NLS in yellow, 192–196aa) is shown downstream of RRM2 and is present in all isoforms of Shep (not shown). The diagram is depicted to scale. Arabic numbers 1–379 indicate amino acid (aa) positions in the Shep Isoform E protein. (B) The primary amino acid sequence of Shep-E underlining key features depicted in panel ‘A’. The colour-coding of amino acid residues

corresponds to the colours of RNA-binding domains and predicted NLS signal depicted in panel 'A'. **(C)** Comparison of Shep predicted NLS signal sequence to the SV40 large T antigen shows 57% homology between amino acid residues (PKKK, that is proline (P) or lysine (K)), with homologous residues underlined and in bold.

This putative nuclear localization signal of Shep bears high resemblance to the monopartite NLS sequence of the simian virus 40 (SV40) large T antigen, whose NLS is composed of seven amino acids, **PKKKRKV** (compare underlined sequences in text, **Figure 6.9C**), of which five consecutive residues are positively charged (Adam *et al.* 1989). The similarity of the NLS signal between the SV40 large T antigen and Shep suggests that Shep may use a nuclear import mechanism similar to that employed by SV40, leveraging its positively charged residues for recognition and subsequent transport into the nucleus. We propose a model in which the nucleocytoplasmic trafficking of Shep is mediated through the importin α/β heterodimer complex, similar to the mechanism observed with the SV40 large T antigen (Weis 2003). Further experimental investigation is required to confirm and characterize the functional significance of this putative NLS of Shep in nuclear localization and to validate the proposed model. These may include functional assays such as mutagenesis of the putative NLS sequence or perturbation of the importin α/β complex to examine the impact on subcellular localization of Shep. Additionally, co-immunoprecipitation or co-localization experiments with importin α/β and Shep could provide direct evidence of their interaction and involvement in nucleocytoplasmic trafficking. Moreover, it would be interesting to determine whether a Shep mutant for the putative NLS signal we reported here, which is thought to mediate its nuclear translocation, would still translocate into the nucleus across different cell types within the nervous system.

It is worth considering the possibility that alternative mechanisms may be responsible for Shep's nuclear import. Multiple lines of evidence support the hypothesis that some proteins lacking a NLS signal can still be imported into the nucleus through alternative means (Jans *et al.* 1997; Shiota *et al.* 1999; Leslie *et al.* 2004). In such cases, these proteins are unable to directly interact with transport receptors and instead employ a piggyback mechanism by interacting with other proteins that possess a functional NLS signal. This piggyback mechanism has been observed to play a role in facilitating the nuclear localization of an import-defective mutant of the hepatitis D virus antigen (Xia

et al. 1992). Therefore, in the case of Shep, it is possible that it is also indirectly imported to the nucleus by means of this piggyback import mechanism, where Shep protein is "piggybacked" along with NLS-containing proteins or protein complexes that can directly interact with transport receptors and facilitate their nuclear import. This would allow Shep to enter the nucleus despite lacking a direct NLS signal, leveraging the presence of other proteins that possess functional NLS signals.

The mechanism by which Shep translocates to the nucleus in certain cell types but not in others remains unclear. This ambiguity is heightened when considering that isoform E of Shep functions in the nucleus within the nervous system (Matzat *et al.* 2012; Chen *et al.* 2018; Chen *et al.* 2019), yet in the ovary, it remains exclusively cytoplasmic (**Figure 3.6D**). Such observations raise the question of whether alternative mechanisms, such as post-translational modifications or interactions with RNA molecules or proteins acting as decoys, might hinder Shep's nuclear import.

6.9 Future Directions

The data presented in this thesis characterized the expression of Shep in the *D. melanogaster* ovary, demonstrated a role for Shep in the post-transcriptional regulation of both *oskar* and *gurken* mRNAs, and uncovered Shep's interactions with other mRNP complexes. However, it is worth mentioning that follow-up experiments are needed to complement and strengthen our novel findings:

Overexpression of Shep in the female germline, regardless of the isoform, disrupted *oskar* translation at the posterior pole of the oocyte. Additionally, it affected *gurken* expression at the dorsal-anterior cortex of the oocyte, which influenced the development of the egg's dorsal appendages. In spite of our RNA-affinity pulldown assay indicating a direct interaction of Shep with both mRNAs and the supporting *in vivo* evidence, it remains intriguing to determine whether the disruptions observed in *oskar* and *gurken*, caused by the overexpression of various Shep isoforms (A, C, and E), arise from Shep's direct binding to *oskar* and *gurken* mRNAs as opposed to influencing other components within the mRNP complex. To investigate this, we generated several Shep mutant constructs in which three key residues in each RRM

domain were mutated to alanine, disrupting Shep's RNA-binding capacity (Chen *et al.* 2019), as detailed in Section 2.2.1.3 of the 'Material & Methods' Chapter. Unfortunately, given the time constraints, we were unable to analyze these mutants. It will be valuable and insightful to determine if the overexpression of these Shep RRM mutants results in the same disruptions observed with the non-RRM mutant isoforms. Furthermore, it will be interesting to see whether the posterior enrichment of Shep-E requires its capacity to bind RNA. Finally, the analysis of the distribution of these RRM mutants may help our understanding of the mechanism directing the nuclear import of Shep as it is possible that the interaction between Shep and different mRNAs tightly regulated in the cytoplasm during the different stages of oogenesis prevent the nuclear import of Shep within the ovarian cells. In other words, the mRNAs that Shep interacts with in the cytoplasm serve as a decoy that impedes the localization of Shep to the nucleus.

The three RNAcompete-based 7-nucleotide motifs identified by Ray *et al.* (2013) served as good indicators of Shep RNA-binding, particularly motif 1 (WAUWUWD). In many cases we examined, the number of predicted Shep-binding sites within a tested mRNA region correlated with the interactions observed in our *in vitro* RNA-affinity pulldown assays (Figures 5.1B & 5.2). However, no definitive sequence for the Shep RNA-binding site has yet been identified. One possible way to identify the primary sequence mediating Shep RNA-binding interactions involves strategically mutating the binding sites within the mRNA targets we tested. This would be beneficial in understanding the molecular processes underlying the interactions between Shep and its RNA targets.

As part of this project, we defined experimental conditions tailored for the efficient immunoprecipitation of mRNP complexes from *Drosophila* ovaries, leveraging the specific affinity of commercial nanobodies for GFP-tagged proteins. These conditions have allowed for the purification of both mRNAs and proteins that are associated to such complexes. Building on these conditions, it will be interesting to extend our analysis to identify the *in vivo* RNA targets of Shep, and other RNA-binding proteins, using transcriptome- and proteome-wide analyses, respectively. Such results will be

particularly valuable as most RIP-seq studies published in the literature use cell culture lines as opposed to targeting RNP complexes formed *in vivo* in different tissues. Moreover, the same experimental conditions could potentially be adapted to isolate the protein components of these complexes along with Shep, paving the way for the identification of other members of these regulatory RNP complexes using mass spectrometry. The RNA-seq results, in conjunction with the identification of proteins co-immunoprecipitated with Shep, will provide deeper insights into the potential mechanisms underlying Shep's cytoplasmic post-transcriptional control during oogenesis. These insights could also shed light on Shep's roles in other cell types and species.

⌘ Chapter VII ⌘

⌘ Conclusions ⌘

7 Chapter VII – Conclusions:

In conclusion, the results of this PhD research project have revealed novel expression patterns, regulation, function, and interactions of *Drosophila* Shep in the ovaries, a tissue that *shep* had remained largely unexplored prior to our work.

7.1 Characterization of *shep* Gene Expression Pattern in the Ovary

- ✎ Characterizing Shep expression using a GFP protein-trap line shows that Shep is expressed in the germline and in nearly all subpopulations of ovarian somatic follicle cells. During the early stages of oogenesis, the Shep protein is uniformly distributed within the oocyte's cytoplasm. In contrast, Shep localizes to the cell periphery from mid-oogenesis onwards, with enrichment at the posterior pole and anterior cortex. Surprisingly, we failed to detect Shep in the nucleus of any of the ovarian cells.
- ✎ Cross-referencing our data from both RT-PCR and Western blot analyses, we confirmed that of the eight *shep* transcripts annotated in FlyBase, six (A, B, E, F, H, and I) are expressed in the ovary, while D and G are not. Nonetheless, at the protein level, all predicted Shep isoforms are expressed. This is because the coding sequences of transcripts D and B are identical, as are those of transcripts G and E.
- ✎ Analysis of the subcellular distribution patterns of three mCherry-tagged Shep isoforms (A, C, and E) within the female germline revealed that Shep localization to the posterior cortex of the oocyte seems to be isoform-specific. We also hypothesize that the 11-amino acid linker region found between the RRM domains in some Shep isoforms may be critical for its posterior localization. Furthermore, although in the nervous system, Shep functions in the nucleus as a negative regulator of the gypsy chromatin insulator complex, our results show that Gal4-induced expression of various Shep isoforms in ovarian tissues does not lead to its nuclear localization, suggesting a tissue-specific role

for Shep's nuclear function. Additionally, we identified a putative nuclear localization signal, 'PKKKN', in all Shep isoforms based on computational predictions. This signal is located just a few residues downstream of the RRM2 domain and requires further experimental validation.

- ✎ We also showed that *shep* expression in ovarian follicle cells depends on proper intercellular communication between the germline and soma, particularly through the Delta-Notch signalling pathway.

7.2 Biological Role of Shep Protein in the Female Germline

- ✎ Although the downregulation of *shep* levels in the female germline using RNAi hairpins was insufficient to produce *oskar*-related phenotypes, we observed three cell non-autonomous phenotypes, whereby the effects were seen in the neighbouring somatic cells. These phenotypes include disruption of border cell migration, amorphic egg chambers lacking an intact follicular epithelium encapsulating the 16-cell germline syncytium, and downregulation of Lamin C expression in cells within the stalk.
- ✎ Our *in vivo* findings from both loss-of-function and overexpression analyses of *shep* revealed a role for Shep in the post-transcriptional regulation of *oskar* translation. In different genetic backgrounds where Shep activity and levels were compromised, we showed that the Oskar protein appears ectopically and prematurely expressed before the mRNA had reached the oocyte's posterior cortex. Additionally, overexpression of Shep in the female germline, irrespective of the isoform, downregulates Oskar protein levels.
- ✎ Also, *in vivo* overexpression of the different isoforms of Shep in the germline disrupts the development of the dorsal appendages in a cell non-autonomous manner. This biological readout, indicative of a role for Shep in the post-transcriptional regulation of *gurken* mRNP, was reflected in the apparent levels and localization pattern of the Gurken protein at the dorsal-anterior region of the late-stage oocyte.

7.3 Shep Protein Interactions with mRNP Complexes & PTB

- ✎ In the *in vitro* RNA-affinity pulldown assay, we demonstrated that Shep directly binds exclusively to the 3'UTR of *oskar* mRNA and to both untranslated regions of *gurken* mRNA. These findings lend further support to the previously observed post-transcriptional regulatory role of Shep in influencing their cytoplasmic fate. Additionally, we showed an association between Shep and various regions of other mRNAs, including *Delta*, *Notch*, *shotgun* (*Drosophila* DE-Cadherin), *nanos*, *bicoid*, and *K10*. Furthermore, we also revealed novel interactions of the RNA-binding protein PTB with the same set of mRNAs that were tested for Shep.

- ✎ Our *in vivo* co-immunoprecipitation results using the Shep GFP protein-trap line confirmed that Shep and PTB seem to interact indirectly through their associations to common mRNA molecules. Given that the protein-trap line is a null mutant for the small Shep isoforms, the observed RNA-mediated interaction between Shep and PTB applies to the large Shep isoforms. The direct protein-protein interaction between PTB and the small Shep isoforms remain possible and require further characterization.

- ✎ The associations between Shep and the candidate mRNAs tested in the RNA-affinity pulldown assays were confirmed *in vivo* by immunoprecipitating Shep-GFP and its associated mRNP complexes from protein-trap ovaries, followed by RT-PCR analysis. By meeting the recommended requirements, such as RNA concentration and quality, we demonstrated that our protocol generated immunoprecipitated RNA samples compatible with downstream RIP-seq analysis.

References

References:

- Adam, S.A., Lobl, T.J., Mitchell, M.A. and Gerace, L. (1989). Identification of specific binding proteins for a nuclear location sequence. *Nature* **337**:276–279.
- Aguirre-Hernández, R., Hoos, H.H. and Condon, A. (2007). Computational RNA secondary structure design: Empirical complexity and improved methods. *BMC Bioinformatics* **8**:1–16. doi: <https://doi.org/10.1186/1471-2105-8-34>.
- Al-Hashimi, H.M. and Walter, N.G. (2008). RNA dynamics: it is about time. *Current Opinion in Structural Biology* **18**(3):321–329. doi: <https://doi.org/10.1016/j.sbi.2008.04.004>.
- Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K., ... Hunt, T. (2022). *Molecular Biology of the Cell, 7th Ed.* New York, United States: W. W. Norton & Company.
- Allison, L.A. (2021). *Fundamental Molecular Biology, 3rd Ed.* Oxford, UK: Wiley-Blackwell.
- Aloufi, N., Alluli, A., Eidelman, D.H. and Baglole, C.J. (2021). Aberrant post-transcriptional regulation of protein expression in the development of chronic obstructive pulmonary disease. *International Journal of Molecular Sciences* **22**(21). doi: <https://doi.org/10.3390/ijms222111963>.
- Anantharaman, V., Koonin, E. V. and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Research* **30**(7):1427–1464. doi: <https://doi.org/10.1093/nar/30.7.1427>.
- Antel, M. and Inaba, M. (2020). Modulation of Cell-Cell Interactions in *Drosophila* Oocyte Development. *Cells* **9**(2):2–3. doi: <https://doi.org/10.3390/cells9020274>.
- Armstrong, J.D., Texada, M.J., Munjaal, R., Baker, D.A. and Beckingham, K.M. (2006). Gravitaxis in *Drosophila melanogaster*: A forward genetic screen. *Genes, Brain and Behavior* **5**(3):222–239. doi: <https://doi.org/10.1111/j.1601-183X.2005.00154.x>.
- Ashburner, M., Golic, K. and Hawley, R. (2005). *Drosophila: A Laboratory Handbook, 2nd Ed.* New York: Cold Spring Harbor Laboratory Press.
- Ashburner, M. and Roote, J. (2007). *Maintenance of a Drosophila Laboratory: General Procedures.* New York: Cold Spring Harbor Protocol.
- Assa-Kunik, E., Torres, I.L., Schejter, E.D., St Johnston, D. and Shilo, B.Z. (2007). *Drosophila* follicle cells are patterned by multiple levels of Notch signaling and antagonism between the Notch and JAK/STAT pathways. *Development* **134**(6):1161–1169. doi: <https://doi.org/10.1242/dev.02800>.
- Auweter, S.D., Oberstrass, F.C. and Allain, F.H.T. (2006). Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Research* **34**(17):4943–4959. doi: <https://doi.org/10.1093/nar/gkl620>.
- Avilés-Pagán, E.E. and Orr-Weaver, T.L. (2018). Activating embryonic development

- in *Drosophila*. *Seminars in Cell and Developmental Biology* **84**:100–110. doi: <https://doi.org/10.1016/j.semcdb.2018.02.019>.
- Babu, K., Cai, Y., Bahri, S., Yang, X. and Chia, W. (2004). Roles of Bifocal, Homer, and F-actin in anchoring Oskar to the posterior cortex of *Drosophila* oocytes. *Genes and Development* **18**(2):138–143. doi: <https://doi.org/10.1101/gad.282604>.
- Bachmann, A. and Knust, E. (2008). The use of P-element transposons to generate transgenic flies. *Methods in Molecular Biology* **420**:61–77.
- Baksa, K., Parke, T., Dobens, L.L. and Dearolf, C.R. (2002). The *Drosophila* STAT protein, stat92E, regulates follicle cell differentiation during oogenesis. *Developmental Biology* **243**(1):166–175. doi: <https://doi.org/10.1006/dbio.2001.0539>.
- Barresi, M. and Gilbert, S. (2020). *Developmental Biology - 12th Edition*. USA: Oxford University Press.
- Barresi, M. and Gilbert, S. (2023). *Developmental Biology, 13 Ed*. Sunderland, Massachusetts, United States: Sinauer Associates.
- Barwell, T., DeVeale, B., Poirier, L., Zheng, J., Seroude, F. and Seroude, L. (2017). Regulating the UAS/GAL4 system in adult *Drosophila* with Tet-off GAL80 transgenes. *PeerJ* **2017**(12). doi: <https://doi.org/10.7717/peerj.4167>.
- Barwell, T., Geld, S. and Seroude, L. (2023). Comparison of GAL80ts and Tet-off GAL80 transgenes. *microPublication biology* **2023**.
- 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 Reports* **4**(1):220–228. doi: <https://doi.org/10.1016/j.celrep.2013.06.020>.
- Bastock, R. and St Johnston, D. (2008). *Drosophila* oogenesis. *Current Biology* **18**(23):1082–1087. doi: <https://doi.org/10.1016/j.cub.2008.09.011>.
- Bateman, J.R., Lee, A.M. and Wu, C.T. (2006). Site-specific transformation of *Drosophila* via ϕ C31 integrase-mediated cassette exchange. *Genetics* **173**(2):769–777. doi: <https://doi.org/10.1534/genetics.106.056945>.
- Bayer, L. V, Milano, S., Formel, S.K., Kaur, H., Cambeiro, J.A., Slinko, L., ... Diana, P. (2023). Cup is essential for oskar mRNA translational repression during early *Drosophila* oogenesis.
- Bellen, H.J. and Yamamoto, S. (2015). Morgan's Legacy: Fruit Flies and the Functional Annotation of Conserved Genes. *Cell* **163**(1):12–14. doi: <https://doi.org/10.1016/j.cell.2015.09.009>.
- Benoit, P., Papin, C., Kwak, J.E., Wickens, M. and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. *Development* **135**(11):1969–1979. doi: <https://doi.org/10.1242/dev.021444>.
- Benton, R. and Johnston, D.S. (2002). Cell polarity: Posterior Par-1 prevents proteolysis. *Current Biology* **12**(14):479–481. doi: [https://doi.org/10.1016/S0960-9822\(02\)00962-4](https://doi.org/10.1016/S0960-9822(02)00962-4).

- Berg, C.A. (2005). The *Drosophila* shell game: Patterning genes and morphological change. *Trends in Genetics* **21**(6):346–355. doi: <https://doi.org/10.1016/j.tig.2005.04.010>.
- Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of bacteriology* **62**(3):293–300.
- Besse, F., De Quinto, S.L., Marchand, V., Trucco, A. and Ephrussi, A. (2009). *Drosophila* PTB promotes formation of high-order RNP particles and represses oskar translation. *Genes and Development* **23**(2):195–207. doi: <https://doi.org/10.1101/gad.505709>.
- Bhaskaran, H., Rodriguez-Hernandez, A. and Perona, J.J. (2012). Kinetics of tRNA folding monitored by aminoacylation. *Rna* **18**(3):569–580. doi: <https://doi.org/10.1261/rna.030080.111>.
- Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**(6):1513–1523.
- Bischof, J., Maeda, R.K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific ϕ C31 integrases. *Proceedings of the National Academy of Sciences of the United States of America* **104**(9):3312–3317. doi: <https://doi.org/10.1073/pnas.0611511104>.
- Blair, S.S. (2000). Notch signaling: Fringe really is a glycosyltransferase. *Current Biology* **10**(16):608–612. doi: [https://doi.org/10.1016/S0960-9822\(00\)00633-3](https://doi.org/10.1016/S0960-9822(00)00633-3).
- Bloch Qazi, M.C., Heifetz, Y. and Wolfner, M.F. (2003). The developments between gametogenesis and fertilization: Ovulation and female sperm storage in *Drosophila melanogaster*. *Developmental Biology* **256**(2):195–211. doi: [https://doi.org/10.1016/S0012-1606\(02\)00125-2](https://doi.org/10.1016/S0012-1606(02)00125-2).
- Block, C.J., Mitchell, A. V., Wu, L., Glassbrook, J., Craig, D., Chen, W., ... Wu, G. (2021). RNA binding protein RBMS3 is a common EMT effector that modulates triple-negative breast cancer progression via stabilizing PRRX1 mRNA. *Oncogene* **40**(46):6430–6442. doi: <https://doi.org/10.1038/s41388-021-02030-x>.
- Blondel, L., Besse, S., Rivard, E.L., Ylla, G. and Extavour, C.G. (2021). Evolution of a Cytoplasmic Determinant: Evidence for the Biochemical Basis of Functional Evolution of the Novel Germ Line Regulator Oskar. *Molecular Biology and Evolution* **38**(12):5491–5513. doi: <https://doi.org/10.1093/molbev/msab284>.
- Bodenstein, D. (1944). The Induction of Larval Molts in *Drosophila*. *JSTOR* **86**(2):113–124.
- Van De Bor, V., Loreau, V., Malbouyres, M., Cerezo, D., Placenti, A., Ruggiero, F. and Noselli, S. (2021). A dynamic and mosaic basement membrane controls cell intercalation in *Drosophila* ovaries. *Development (Cambridge)* **148**(4):1–13. doi: <https://doi.org/10.1242/dev.195511>.
- Bose, M., Lampe, M., Mahamid, J. and Ephrussi, A. (2022). Liquid-to-solid phase transition of oskar ribonucleoprotein granules is essential for their function in *Drosophila* embryonic development. *Cell* **185**(8):1308–1324.e23. doi: <https://doi.org/10.1016/j.cell.2022.02.022>.

- Boulétreau-Merle, J., Allemand, R., Cohet, Y. and David, J.R. (1982). Reproductive strategy in *Drosophila melanogaster*: Significance of a genetic divergence between temperate and tropical populations. *Oecologia* **53**(3):323–329. doi: <https://doi.org/10.1007/BF00389008>.
- Bram, R.J., Lue, N.F. and Kornberg, R.D. (1986). A GAL family of upstream activating sequences in yeast: roles in both induction and repression of transcription. *The EMBO journal* **5**(3):603–608. doi: <https://doi.org/10.1002/j.1460-2075.1986.tb04253.x>.
- Brameier, M., Krings, A. and MacCallum, R.M. (2007). NucPred - Predicting nuclear localization of proteins. *Bioinformatics* **23**(9):1159–1160. doi: <https://doi.org/10.1093/bioinformatics/btm066>.
- Brand, a H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge, England)* **118**(2):401–15.
- Bratu, D. and McNeil, G. (2015). *Drosophila Oogenesis*. New York, NY, USA: Humana Press.
- Bray, S. and Furriols, M. (2001). Notch pathway: Making sense of suppressor of hairless. *Current Biology* **11**(6):217–221. doi: [https://doi.org/10.1016/S0960-9822\(01\)00109-9](https://doi.org/10.1016/S0960-9822(01)00109-9).
- Brendza, R., Serbus, L., Duffy, J. and Saxton, W. (2000). A Function for Kinesin I in the Posterior Transport of oskar mRNA and Staufen Protein. *Science* **289**(5487):2120–2122.
- Briata, P. and Gherzi, R. (2020). Long non-coding RNA-ribonucleoprotein networks in the post-transcriptional control of gene expression. *Non-coding RNA* **6**(3). doi: <https://doi.org/10.3390/NCRNA6030040>.
- Briggs, R. and King, T.J. (1952). Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proceedings of the National Academy of Sciences* **38**(5):455–463. doi: <https://doi.org/10.1073/pnas.38.5.455>.
- Brischigliaro, M., Fernandez-Vizarra, E. and Viscomi, C. (2023). Mitochondrial Neurodegeneration: Lessons from *Drosophila melanogaster* Models. *Biomolecules* **13**(2):1–22. doi: <https://doi.org/10.3390/biom13020378>.
- Buccitelli, C. and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics* **21**(10):630–644. doi: <https://doi.org/10.1038/s41576-020-0258-4>.
- Buchberger, E., Reis, M., Lu, T.H. and Posnien, N. (2019). Cloudy with a chance of insights: Context dependent gene regulation and implications for evolutionary studies. *Genes* **10**(7). doi: <https://doi.org/10.3390/genes10070492>.
- Bullock, S.L. and Ish-Horowicz, D. (2002). Cell polarity: Oskar seeks PARTner for a stable relationship. *Nature Cell Biology* **4**(5). doi: <https://doi.org/10.1038/ncb0502-e117>.
- Van Buskirk, C., Hawkins, N.C. and Schupbach, T. (2000). Encore is a member of a novel family of proteins and affects multiple processes in *Drosophila* oogenesis.

Development **127**(22):4753–4762. doi: <https://doi.org/10.1242/dev.127.22.4753>.

Buszczak, M. and Cooley, L. (2000). Eggs to die for: Cell death during *Drosophila* oogenesis. *Cell Death and Differentiation* **7**(11):1071–1074. doi: <https://doi.org/10.1038/sj.cdd.4400755>.

Buszczak, M., Paterno, S., Lighthouse, D., Bachman, J., Planck, J., Owen, S., ... Spradling, A.C. (2007). The carnegie protein trap library: A versatile tool for *drosophila* developmental studies. *Genetics* **175**(3):1505–1531. doi: <https://doi.org/10.1534/genetics.106.065961>.

Cáceres, L. and Nilson, L.A. (2005). Production of gurken in the nurse cells is sufficient for axis determination in the *Drosophila* oocyte. *Development* **132**(10):2345–2353. doi: <https://doi.org/10.1242/dev.01820>.

Cai, D., Chen, S.C., Prasad, M., He, L., Wang, X., Choemmel-Cadamuro, V., ... Montell, D.J. (2014). Mechanical feedback through E-cadherin promotes direction sensing during collective cell migration. *Cell* **157**(5):1146–1159. doi: <https://doi.org/10.1016/j.cell.2014.03.045>.

Cavaliere, V., Bernardi, F., Romani, P., Duchi, S. and Gargiulo, G. (2008). Building up the *Drosophila* eggshell: First of all the eggshell genes must be transcribed. *Developmental Dynamics* **237**(8):2061–2072. doi: <https://doi.org/10.1002/dvdy.21625>.

Caygill, E.E. and Brand, A.H. (2016). *The GAL4 System: A Versatile System for the Manipulation and Analysis of Gene Expression*. New York, NY, USA: Humana Press.

Chamberlin, M., Mcgrath, J. and Waskell, L. (1970). New RNA polymerase from *escherichia coli* infected with bacteriophage T7. *Nature* **228**(5268):227–231. doi: <https://doi.org/10.1038/228227a0>.

Chang, J.S., Tan, L. and Schedl, P. (1999). The *Drosophila* CPEB homolog, Orb, is required for oskar protein expression in oocytes. *Developmental Biology* **215**(1):91–106. doi: <https://doi.org/10.1006/dbio.1999.9444>.

Chang, J.S., Tan, L., Wolf, M.R. and Schedl, P. (2001). Functioning of the *Drosophila* orb gene in gurken mRNA localization and translation. *Development* **128**(16):3169–3177. doi: <https://doi.org/10.1242/dev.128.16.3169>.

Chatterjee, S. and Pal, J.K. (2009). Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biology of the Cell* **101**(5):251–262. doi: <https://doi.org/10.1042/bc20080104>.

Chau, J., Kulnane, L.S. and Salz, H.K. (2012). Sex-lethal enables germline stem cell differentiation by down-regulating Nanos protein levels during *Drosophila* oogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **109**(24):9465–9470. doi: <https://doi.org/10.1073/pnas.1120473109>.

Chekulaeva, M., Hentze, M.W. and Ephrussi, A. (2006). Bruno acts as a dual repressor of oskar translation, promoting mRNA oligomerization and formation of silencing particles. *Cell* **124**(3):521–533. doi: <https://doi.org/10.1016/j.cell.2006.01.031>.

Chełkowska-Pauszek, A., Kosiński, J.G., Marciniak, K., Wysocka, M., Bakowska-żywicka, K. and Żywicki, M. (2021). The role of rna secondary structure in regulation

of gene expression in bacteria. *International Journal of Molecular Sciences* **22**(15). doi: <https://doi.org/10.3390/ijms22157845>.

Chen, C.Y.A. and Shyu, A. Bin (1995). AU-rich elements: characterization and importance in mRNA degradation. *Trends in Biochemical Sciences* **20**(11):465–470. doi: [https://doi.org/10.1016/S0968-0004\(00\)89102-1](https://doi.org/10.1016/S0968-0004(00)89102-1).

Chen, D., Brovkina, M., Matzat, L.H. and Lei, E.P. (2019). Shep RNA-binding capacity is required for antagonism of gypsy chromatin insulator activity. *G3: Genes, Genomes, Genetics* **9**(3):749–754. doi: <https://doi.org/10.1534/g3.118.200923>.

Chen, D., Dale, R.K. and Lei, E.P. (2018). Shep regulates Drosophila neuronal remodeling by controlling transcription of its chromatin targets. *Development (Cambridge)* **145**(1):1–11. doi: <https://doi.org/10.1242/dev.154047>.

Chen, D., Gu, T., Pham, T.N., Zachary, M.J. and Hewes, R.S. (2017). Regulatory mechanisms of metamorphic neuronal remodeling revealed through a Genome-Wide modifier screen in drosophila melanogaster. *Genetics* **206**(3):1429–1443. doi: <https://doi.org/10.1534/genetics.117.200378>.

Chen, D. and McKearin, D.M. (2003). A discrete transcriptional silencer in the bam gene determines asymmetric division of the Drosophila germline stem cell. *Development* **130**(6):1159–1170. doi: <https://doi.org/10.1242/dev.00325>.

Chen, D., McManus, C.E., Radmanesh, B., Matzat, L.H. and Lei, E.P. (2021). Temporal inhibition of chromatin looping and enhancer accessibility during neuronal remodeling. *Nature Communications* **12**(1):1–10. doi: <https://doi.org/10.1038/s41467-021-26628-7>.

Chen, D., Qu, C. and Hewes, R.S. (2014). *Neuronal Remodeling during Metamorphosis Is Regulated by the Alan Shepard (Shep) Gene in Drosophila Melanogaster*.

Chen, L. and Kashina, A. (2021). Post-translational Modifications of the Protein Termini. *Frontiers in Cell and Developmental Biology* **9**(July):1–14. doi: <https://doi.org/10.3389/fcell.2021.719590>.

Chen, Y., Tsai, Y.H. and Tseng, S.H. (2016). Regulation of the expression of cytoplasmic polyadenylation element binding proteins for the treatment of cancer. *Anticancer Research* **36**(11):5673–5680. doi: <https://doi.org/10.21873/anticancer.11150>.

Chopra, V.S., Srinivasan, A., Kumar, R.P., Mishra, K., Basquin, D., Docquier, M., ... Mishra, R.K. (2008). Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Developmental Biology* **317**(2):660–670. doi: <https://doi.org/10.1016/j.ydbio.2008.02.008>.

Church, S.H., De Medeiros, B.A.S., Donoughe, S., Márquez Reyes, N.L. and Extavour, C.G. (2021). Repeated loss of variation in insect ovary morphology highlights the role of development in life-history evolution. *Proceedings of the Royal Society B: Biological Sciences* **288**(1950). doi: <https://doi.org/10.1098/rspb.2021.0150>.

Clark, I.E., Wyckoff, D. and Gavis, E.R. (2000). Synthesis of the posterior determinant Nanos is spatially restricted by a novel cotranslational regulatory mechanism. *Current*

Biology 10(20):1311–1314. doi: [https://doi.org/10.1016/S0960-9822\(00\)00754-5](https://doi.org/10.1016/S0960-9822(00)00754-5).

Cléry, A., Blatter, M. and Allain, F.H.T. (2008). RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology* 18(3):290–298. doi: <https://doi.org/10.1016/j.sbi.2008.04.002>.

Clouse, K.N., Ferguson, S.B. and Schüpbach, T. (2008). Squid, Cup, and PABP55B function together to regulate gurken translation in *Drosophila*. *Developmental Biology* 313(2):713–724. doi: <https://doi.org/10.1016/j.ydbio.2007.11.008>.

Cokol, M., Nair, R. and Rost, B. (2000). Finding nuclear localization signals. *EMBO Reports* 1(5):411–415. doi: <https://doi.org/10.1093/embo-reports/kvd092>.

Consortium, T., Roy, S., Ernst, J., Kharchenko, P. V, Kheradpour, P., Negre, N., ... Schwartz, Y.B. (2011). Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE. *Science* 330(6012):1787–1797. doi: <https://doi.org/10.1126/science.1198374>.Identification.

Cook, K.B., Hughes, T.R. and Morris, Q.D. (2015). High-throughput characterization of protein-RNA interactions. *Briefings in Functional Genomics* 14(1):74–89. doi: <https://doi.org/10.1093/bfgp/elu047>.

Corley, M., Burns, M.C. and Yeo, G.W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell* 78(1):9–29. doi: <https://doi.org/10.1016/j.molcel.2020.03.011>.

Craig, N., Green, R., Greider, C., Storz, G. and Wolberger, C. (2021). *Molecular Biology - Principles of Genome Function. 3rd Edition.* 3rd ed. Oxford: Oxford University Press.

Cuevas, M. (2015). *Drosophila* Oogenesis . *eLS*(2006):1–7. doi: <https://doi.org/10.1002/9780470015902.a0001502.pub2>.

Cutler, S.R., Ehrhardt, D.W., Griffiths, J.S. and Somerville, C.R. (2000). Random GFP::cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proceedings of the National Academy of Sciences of the United States of America* 97(7):3718–3723. doi: <https://doi.org/10.1073/pnas.97.7.3718>.

Das, S., Vera, M., Gandin, V., Singer, R.H. and Tutucci, E. (2021). Intracellular mRNA transport and localized translation. *Nature Reviews Molecular Cell Biology* 22(7):483–504. doi: <https://doi.org/10.1038/s41580-021-00356-8>.

Decker, C.J. and Parker, R. (2002). mRNA decay enzymes: Decappers conserved between yeast and mammals. *Proceedings of the National Academy of Sciences of the United States of America* 99(20):12512–12514. doi: <https://doi.org/10.1073/pnas.212518099>.

Deluca, S.Z. and Spradling, A.C. (2018). Efficient expression of genes in the *drosophila* germline using a *uas* promoter free of interference by *hsp70* pinnas. *Genetics* 209(2):381–387. doi: <https://doi.org/10.1534/genetics.118.300874>.

Demeshkina, N., Jenner, L., Yusupova, G. and Yusupov, M. (2010). Interactions of the ribosome with mRNA and tRNA. *Current Opinion in Structural Biology* 20(3):325–332. doi: <https://doi.org/10.1016/j.sbi.2010.03.002>.

- Deng, S., Feng, Y. and Pauklin, S. (2022). 3D chromatin architecture and transcription regulation in cancer. *Journal of Hematology and Oncology* **15**(1):1-23. doi: <https://doi.org/10.1186/s13045-022-01271-x>.
- Deng, W.M., Althausen, C. and Ruohola-Baker, H. (2001). Notch-Delta signaling induces a transition from mitotic cell cycle to endocycle in *Drosophila* follicle cells. *Development* **128**(23):4737-4746. doi: <https://doi.org/10.1242/dev.128.23.4737>.
- Desjardins, P. and Conklin, D. (2010). NanoDrop Microvolume Quantitation of Nucleic Acids. *J. Vis. Exp.*(45). doi: <https://doi.org/10.3791/2565>.
- Dialynas, G., Speese, S., Budnik, V., Geyer, P.K. and Wallrath, L.L. (2010). The role of *Drosophila* Lamin C in muscle function and gene expression. *Development* **137**(18):3067-3077. doi: <https://doi.org/10.1242/dev.048231>.
- Ding, D.Q., Tomita, Y., Yamamoto, A., Chikashige, Y., Haraguchi, T. and Hiraoka, Y. (2000). Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library. *Genes to Cells* **5**(3):169-190. doi: <https://doi.org/10.1046/j.1365-2443.2000.00317.x>.
- Dold, A., Han, H., Liu, N., Hildebrandt, A., Brüggemann, M., Rücklé, C., ... Lasko, P. (2020). *Makorin 1 Controls Embryonic Patterning by Alleviating Bruno1-Mediated Repression of Oskar Translation*.
- Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., ... Burge, C.B. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Molecular Cell* **70**(5):854-867.e9. doi: <https://doi.org/10.1016/j.molcel.2018.05.001>.
- Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews Molecular Cell Biology* **3**(3):195-205. doi: <https://doi.org/10.1038/nrm760>.
- Du, X. and Xiao, R. (2020). An emerging role of chromatin-interacting RNA-binding proteins in transcription regulation. *Essays Biochem* **64**(6):907-918.
- Duan, J., Wang, X. and Kizer, M.E. (2020). *Biotechnological and Therapeutic Applications of Natural Nucleic Acid Structural Motifs*. Springer International Publishing.
- Duffy, J.B. (2002). GAL4 system in *Drosophila*: A fly geneticist's Swiss army knife. *Genesis (United States)* **34**(1-2):1-15. doi: <https://doi.org/10.1002/gene.10150>.
- Duhart, J.C., Parsons, T.T. and Rafferty, L.A. (2017). The repertoire of epithelial morphogenesis on display: Progressive elaboration of *Drosophila* egg structure. *Mechanisms of Development* **148**:18-39. doi: <https://doi.org/10.1016/j.mod.2017.04.002>.
- Elliott, D. and Lodomery, M. (2015). *Molecular Biology of RNA*. Oxford, UK: Oxford University Press.
- ElMaghraby, M.F., Tirian, L., Senti, K.A., Meixner, K. and Brennecke, J. (2022). A genetic toolkit for studying transposon control in the *Drosophila melanogaster* ovary. *Genetics* **220**(1). doi: <https://doi.org/10.1093/genetics/iyab179>.

- Ephrussi, A., Dickinson, L.K. and Lehmann, R. (1991). oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**(1):37–50.
- Ephrussi, A. and Lehmann, R. (1992). Induction of germ cell fate by oskar. *Nature* **358**(July):387–392.
- Fadiga, J. and Nystul, T.G. (2019). The follicle epithelium in the drosophila ovary is maintained by a small number of stem cells. *eLife* **8**:1–17. doi: <https://doi.org/10.7554/eLife.49050>.
- Farina, K. and Singer, R. (2002). The nuclear connection in RNA transport and localization. *Trends in Cell Biology* **12**(10):466–72.
- Fedorova, S.A. and Dorogova, N. V. (2020). Protein trap: a new Swiss army knife for geneticists? *Molecular Biology Reports* **47**(2):1445–1458. doi: <https://doi.org/10.1007/s11033-019-05181-z>.
- Feher, J. (2017). DNA and Protein Synthesis. *Quantitative Human Physiology*:120–129. doi: <https://doi.org/10.1016/b978-0-12-800883-6.00011-2>.
- Fellmeth, J.E. and McKim, K.S. (2022). A Brief History of Drosophila (Female) Meiosis. *Genes* **13**(5). doi: <https://doi.org/10.3390/genes13050775>.
- Ferguson, S.B., Blundon, M.A., Klovstad, M.S. and Schüpbach, T. (2012). Modulation of gurken translation by insulin and TOR signaling in Drosophila. *Journal of Cell Science* **125**(6):1407–1419. doi: <https://doi.org/10.1242/jcs.090381>.
- Fernandes, N. and Buchan, J.R. (2021). RNAs as Regulators of Cellular Matchmaking. *Frontiers in Molecular Biosciences* **8**(April). doi: <https://doi.org/10.3389/fmolb.2021.634146>.
- Fernández-Moreno, M.A., Farr, C.L., Kaguni, L.S. and Garesse, R. (2007). Drosophila melanogaster as a Model System to Study Mitochondrial Biology. *Methods in Molecular Biology* **372**:33–49. doi: <https://doi.org/10.1007/978-1-59745-365-3>.
- Filipowicz, W., Bhattacharyya, S.N. and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nature Reviews Genetics* **9**(2):102–114. doi: <https://doi.org/10.1038/nrg2290>.
- Fischer, F.P., Karge, R.A., Weber, Y.G., Koch, H., Wolking, S. and Voigt, A. (2023). Drosophila melanogaster as a versatile model organism to study genetic epilepsies: An overview. *Frontiers in Molecular Neuroscience* **16**(February). doi: <https://doi.org/10.3389/fnmol.2023.1116000>.
- Franz, A. and Riechmann, V. (2010). Stepwise polarisation of the Drosophila follicular epithelium. *Developmental Biology* **338**(2):136–147. doi: <https://doi.org/10.1016/j.ydbio.2009.11.027>.
- Freitas, N. and Cunha, C. (2013). Mechanisms and Signals for the Nuclear Import of Proteins. *Advances in Genome Science*:77–100. doi: <https://doi.org/10.2174/9781608057566113020005>.
- Fribourg, S., Gatfield, D., Izaurrealde, E. and Conti, E. (2003). A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nature Structural Biology* **10**(6):433–439. doi: <https://doi.org/10.1038/nsb926>.

- Fuke, H. and Ohno, M. (2008). Role of poly (A) tail as an identity element for mRNA nuclear export. *Nucleic Acids Research* **36**(3):1037–1049. doi: <https://doi.org/10.1093/nar/gkm1120>.
- Ganguly, S., Williams, L.S., Palacios, I.M. and Goldstein, R.E. (2012). Cytoplasmic streaming in *Drosophila* oocytes varies with kinesin activity and correlates with the microtubule cytoskeleton architecture. *Proceedings of the National Academy of Sciences of the United States of America* **109**(38):15109–15114. doi: <https://doi.org/10.1073/pnas.1203575109>.
- Ganser, L.R., Kelly, M.L., Herschlag, D. and Al-Hashimi, H.M. (2019). The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **20**(8):474–489. doi: <https://doi.org/10.1038/s41580-019-0136-0>.
- García-Mauriño, S.M., Rivero-Rodríguez, F., Velázquez-Cruz, A., Hernández-Vellisca, M., Díaz-Quintana, A., De la Rosa, M.A. and Díaz-Moreno, I. (2017). RNA binding protein regulation and cross-talk in the control of AU-rich mRNA Fate. *Frontiers in Molecular Biosciences* **4**(OCT):1–9. doi: <https://doi.org/10.3389/fmolb.2017.00071>.
- Gautam, A. (2022). Lithium Chloride-Based Isolation of RNA. *DNA and RNA Isolation Techniques for Non-Experts*. Springer, Cham.
- Ge, J. and Yu, Y.T. (2013). RNA pseudouridylation: New insights into an old modification. *Trends in Biochemical Sciences* **38**(4):210–218. doi: <https://doi.org/10.1016/j.tibs.2013.01.002>.
- Gebauer, F., Preiss, T. and Hentze, M.W. (2012). From cis-regulatory elements to complex RNPs and back. *Cold Spring Harbor Perspectives in Biology* **4**(7):1–14. doi: <https://doi.org/10.1101/cshperspect.a012245>.
- Gerstberger, S., Hafner, M. and Tuschl, T. (2013). Learning the language of post-transcriptional gene regulation. *Genome Biology* **14**(8):6–10. doi: <https://doi.org/10.1186/gb-2013-14-8-130>.
- Giedt, M. and Tootle, T. (2023). *Drosophila Oogenesis*. New York, NY: Humana Press,.
- Giuliani, G., Giuliani, F., Volk, T. and Rabouille, C. (2014). The *Drosophila* RNA-binding protein HOW controls the stability of dgrasp mRNA in the follicular epithelium. *Nucleic Acids Research* **42**(3):1970–1986. doi: <https://doi.org/10.1093/nar/gkt1118>.
- Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **582**(14):1977–1986. doi: <https://doi.org/10.1016/j.febslet.2008.03.004>.
- Godt, D. and Tepass, U. (1998a). *Drosophila* oocyte localisation is mediated by differential cadherin-based adhesion. *Nature* **395**(1996):387–391.
- Godt, D. and Tepass, U. (1998b). *Drosophila* oocyte localization is mediated by differential cadherin-based adhesion. *Nature* **395**(6700):387–391. doi: <https://doi.org/10.1038/26493>.
- Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T. and Adam, S.A. (2004).

Importin α : a multipurpose nuclear-transport receptor. *Trends in Cell Biology* **14**(9):505–514.

Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R. and Choder, M. (2008). Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes and Development* **22**(15):2022–2027. doi: <https://doi.org/10.1101/gad.473608>.

Gong, F.X., Zhan, G., Han, R., Yang, Z., Fu, X. and Xiao, R. (2021). De-dimerization of PTB is catalyzed by PDI and is involved in the regulation of p53 translation. *Nucleic Acids Research* **49**(16):9342–9352. doi: <https://doi.org/10.1093/nar/gkab708>.

Gonsalvez, G.B., Rajendra, T.K., Wen, Y., Praveen, K. and Matera, A.G. (2010). Sm proteins specify germ cell fate by facilitating oskar mRNA localization. *Development* **137**(14):2341–2351. doi: <https://doi.org/10.1242/dev.042721>.

González-Cruz, R.D., Dahl, K.N. and Darling, E.M. (2018). The emerging role of Lamin C as an important LMNA isoform in mechanophenotype. *Frontiers in Cell and Developmental Biology* **6**:1–7. doi: <https://doi.org/10.3389/fcell.2018.00151>.

González-Reyes, A., Elliott, H. and St Johnston, D. (1995). Polarization of both major body axes in drosophila by gurken-torpedo signalling. *Nature* **375**(6533):654–658. doi: <https://doi.org/10.1038/375654a0>.

González-reyes, A. and Johnston, D.S. (1998). Patterning of the follicle cell epithelium along the anterior-posterior axis during *Drosophila* oogenesis. *Development* **125**(15):1–10.

González-Reyes, A. and Johnston, D.S. (1994). Role of Oocyte Position in Establishment of Anterior-Posterior Polarity in *Drosophila*. *Science* **266**(5185):639–642.

Goss, D.J. and Domashevskiy, A. V. (2022). *Messenger RNA (MRNA): The Link Between DNA and Protein*. Elsevier Ltd.

van der Graaf, K., Srivastav, S., Singh, P., McNew, J.A. and Stern, M. (2022). The *Drosophila melanogaster* attP40 docking site and derivatives are insertion mutations of msp-300. *PLoS ONE* **17**(12 December):1–20. doi: <https://doi.org/10.1371/journal.pone.0278598>.

Gramates, L.S., Agapite, J., Attrill, H., Calvi, B.R., Crosby, M.A., dos Santos, G., ... Strelets, V.B. (2022). FlyBase: a guided tour of highlighted features. *Genetics* **220**(4). doi: <https://doi.org/10.1093/genetics/iyac035>.

Grammont, M. and Irvine, K. (2001). fringe and Notch specify polar cell fate during *Drosophila* oogenesis. *Development* **128**(12):2243–2253.

Grammont, M. and Irvine, K.D. (2002). Organizer activity of the polar cells during *Drosophila* oogenesis. *Development* **129**(22):5131–5140. doi: <https://doi.org/10.1242/dev.129.22.5131>.

Greenspan, R. (2004). *Fly Pushing The Theory and Practice of Drosophila Genetics, 2nd Ed.* New York: Cold Spring Harbor Laboratory Press.

Grosjean, H. and Benne, R. (1998). *Modification and Editing of RNA*. Washington,

DC: ASM Press.

Groth, A.C., Fish, M., Nusse, R. and Calos, M.P. (2004). Construction of Transgenic *Drosophila* by Using the Site-Specific Integrase from Phage ϕ C31. *Genetics* **166**(4):1775–1782. doi: <https://doi.org/10.1534/genetics.166.4.1775>.

Guarente, L., Yocum, R.R. and Gifford, P. (1982). A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. *Proceedings of the National Academy of Sciences of the United States of America* **79**(23):7410–7414. doi: <https://doi.org/10.1073/pnas.79.23.7410>.

Guild, G.M., Connelly, P.S., Shaw, M.K. and Tilney, L.G. (1997). Actin filament cables in *Drosophila* nurse cells are composed of modules that slide passively past one another during dumping. *Journal of Cell Biology* **138**(4):783–797. doi: <https://doi.org/10.1083/jcb.138.4.783>.

Gunkel, N., Yano, T., Markussen, F.H., Olsen, L.C. and Ephrussi, A. (1998). Localization-dependent translation requires a functional interaction between the 5' and 3' ends of oskar mRNA. *Genes and Development* **12**(11):1652–1664. doi: <https://doi.org/10.1101/gad.12.11.1652>.

Gurdon, J.B. and Uehlinger, V. (1966). 'Fertile' intestine nuclei. *Nature* **210**(5042):1240–1241. doi: <https://doi.org/10.1038/2101240a0>.

Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. *Mol Cell Biol* **19**(3):1720–1730.

Hachet, O. and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428**(6986):950–955. doi: <https://doi.org/10.1038/nature02483>.

Hales, K.G., Korey, C.A., Larracuenta, A.M. and Roberts, D.M. (2015). Genetics on the fly: A primer on the *Drosophila* model system. *Genetics* **201**(3):815–842. doi: <https://doi.org/10.1534/genetics.115.183392>.

Han, D.D., Stein, D. and Stevens, L.M. (2000). Investigating the function of follicular subpopulations during *Drosophila* oogenesis through hormone-dependent enhancer-targeted cell ablation. *Development* **127**(3):573–583.

Handler, A. and Harrell-Ii, R. (2001). Germline transformation of *Drosophila melanogaster* with the piggyBac transposon vector. *Insect Molecular Biology* **8**(4):449–457.

He, L., Wang, X. and Montell, D.J. (2011). Shining light on *Drosophila* oogenesis: Live imaging of egg development. *Current Opinion in Genetics and Development* **21**(5):612–619. doi: <https://doi.org/10.1016/j.gde.2011.08.011>.

Helfand, S.L. and Rogina, B. (2003). Genetics of Aging in the Fruit Fly, *Drosophila melanogaster*. *Annual Review of Genetics* **37**:329–348.

Hellen, C.U.T. and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes and Development* **15**(13):1593–1612. doi: <https://doi.org/10.1101/gad.891101>.

Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018). A brave new world of

- RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **19**(5):327–341. doi: <https://doi.org/10.1038/nrm.2017.130>.
- Hilgarth, R.S. and Lanigan, T.M. (2020). Optimization of overlap extension PCR for efficient transgene construction. *MethodsX* **7**(December 2019):100759. doi: <https://doi.org/10.1016/j.mex.2019.12.001>.
- Hinnant, T.D., Merkle, J.A. and Ables, E.T. (2020). Coordinating Proliferation, Polarity, and Cell Fate in the *Drosophila* Female Germline. *Frontiers in Cell and Developmental Biology* **8**(February):1–22. doi: <https://doi.org/10.3389/fcell.2020.00019>.
- Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N. and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**:1–16. doi: <https://doi.org/10.1186/1471-2105-12-357>.
- Huynh, J.R., Munro, T.P., Smith-Litière, K., Lepesant, J.A. and St Johnston, D. (2004). The *Drosophila* hnRNP A/B homolog, Hrp48, is specifically required for a distinct step in *osk* mRNA localization. *Developmental Cell* **6**(5):625–635. doi: [https://doi.org/10.1016/S1534-5807\(04\)00130-3](https://doi.org/10.1016/S1534-5807(04)00130-3).
- Huynh, J.R. and St Johnston, D. (2004). The origin of asymmetry: Early polarisation of the *Drosophila* germline cyst and oocyte. *Current Biology* **14**(11):438–449. doi: <https://doi.org/10.1016/j.cub.2004.05.040>.
- Iost, I., Guillerez, J. and Dreyfus, M. (1992). Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *Journal of Bacteriology* **174**(2):619–622. doi: <https://doi.org/10.1128/jb.174.2.619-622.1992>.
- Jackson, R. and Kaminski, A. (1995). Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *RNA* **1**(10):985–1000.
- Jacob, A.G. and Smith, C.W.J. (2017). Intron retention as a component of regulated gene expression programs. *Human Genetics* **136**(9):1043–1057. doi: <https://doi.org/10.1007/s00439-017-1791-x>.
- Jaglarz, M.K., Krzeminski, W. and Bilinski, S.M. (2008). Structure of the ovaries and follicular epithelium morphogenesis in *Drosophila* and its kin. *Development Genes and Evolution* **218**(8):399–411. doi: <https://doi.org/10.1007/s00427-008-0233-0>.
- Jain, R.A. and Gavis, E.R. (2008). The *Drosophila* hnRNP M homolog Rumpelstiltskin regulates nanos mRNA localization. *Development* **135**(5):973–982. doi: <https://doi.org/10.1242/dev.015438>.
- Jans, D.A., Briggs, L.J., Gustin, S.E., Jans, P., Ford, S. and Young, I.G. (1997). The cytokine interleukin-5 (IL-5) effects cotransport of its receptor subunits to the nucleus in vitro. *FEBS Letters* **410**(2–3):368–372. doi: [https://doi.org/10.1016/S0014-5793\(97\)00622-4](https://doi.org/10.1016/S0014-5793(97)00622-4).
- Jarvik, J.W., Adler, S.A., Telmer, C.A., Subramaniam, V. and Lopez, A.J. (1996). CD-tagging: A new approach to gene and protein discovery and analysis. *BioTechniques* **20**(5):896–904. doi: <https://doi.org/10.2144/96205rr03>.

- Jennings, B.H. (2011). *Drosophila - A Versatile Model in Biology & Medicine. Materials Today* **14**(5):190-195. doi: [https://doi.org/10.1016/S1369-7021\(11\)70113-4](https://doi.org/10.1016/S1369-7021(11)70113-4).
- Jeske, M., Bordi, M., Glatt, S., Müller, S., Rybin, V., Müller, C.W. and Ephrussi, A. (2015). The crystal structure of the *Drosophila* germline inducer Oskar identifies two domains with distinct Vasa Helicase- and RNA-binding activities. *Cell Reports* **12**(4):587-598. doi: <https://doi.org/10.1016/j.celrep.2015.06.055>.
- Ji, Y. and Tulin, A. V. (2012). Poly(ADP-ribose) controls DE-cadherin-dependent stem cell maintenance and oocyte localization. *Nature Communications* **3**. doi: <https://doi.org/10.1038/ncomms1759>.
- Jia, D., Soylemez, M., Calvin, G., Bornmann, R., Bryant, J., Hanna, C., ... Deng, W.M. (2015). A large-scale in vivo RNAi screen to identify genes involved in Notch-mediated follicle cell differentiation and cell cycle switches. *Scientific Reports* **5**(February):1-14. doi: <https://doi.org/10.1038/srep12328>.
- Jia, D., Xu, Q., Xie, Q., Mio, W. and Deng, W.M. (2016). Automatic stage identification of *Drosophila* egg chamber based on DAPI images. *Scientific Reports* **6**(November 2015):1-12. doi: <https://doi.org/10.1038/srep18850>.
- Jin, W., Brannan, K.W., Kapeli, K., Park, S.S., Tan, H.Q., Gosztyla, M.L., ... Yeo, G.W. (2023). HyDRA: Deep-learning models for predicting RNA-binding capacity from protein interaction association context and protein sequence. *bioRxiv* **83**(14):2022.12.23.521837. doi: <https://doi.org/10.1016/j.molcel.2023.06.019>.
- Johnson, A.D. (2010). An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics* **26**(10):1386-1389. doi: <https://doi.org/10.1093/bioinformatics/btq098>.
- Johnson, T.K., Henstridge, M.A. and Warr, C.G. (2017). MACPF/CDC proteins in development: Insights from *Drosophila* torso-like. *Seminars in Cell and Developmental Biology* **72**:163-170. doi: <https://doi.org/10.1016/j.semcdb.2017.05.003>.
- Johnson, W.L. and Straight, A.F. (2017). RNA-mediated regulation of heterochromatin. *Current Opinion in Cell Biology* **46**:102-109. doi: <https://doi.org/10.1016/j.ceb.2017.05.004>.
- Johnston, D.S., Beuchle, D. and Nüsslein-Volhard, C. (1991). *staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell* **66**:51-63. doi: [https://doi.org/10.1016/0092-8674\(91\)90138-O](https://doi.org/10.1016/0092-8674(91)90138-O).
- Johnston, D.S. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**(2):201-219. doi: [https://doi.org/10.1016/0092-8674\(92\)90466-P](https://doi.org/10.1016/0092-8674(92)90466-P).
- Johnstone, O. and Lasko, P. (2001). Translational Regulation and RNA Localization in *Drosophila* Oocytes and Embryos. *Annual Review of Genetics* **35**:365-406.
- Jones, C.P. and Ferré-D'Amaré, A.R. (2015). RNA quaternary structure and global symmetry. *Trends in Biochemical Sciences* **40**(4):211-220. doi: <https://doi.org/10.1016/j.tibs.2015.02.004>.

- Kanke, M. and Macdonald, P.M. (2015). Translational activation of Oskar mRNA: Reevaluation of the role and importance of a 5' regulatory element. *PLoS ONE* **10**(5):1-16. doi: <https://doi.org/10.1371/journal.pone.0125849>.
- Kato, Y. and Nakamura, A. (2012). Roles of cytoplasmic RNP granules in intracellular RNA localization and translational control in the *Drosophila* oocyte. *Development Growth and Differentiation* **54**(1):19-31. doi: <https://doi.org/10.1111/j.1440-169X.2011.01314.x>.
- Kerner, P., Degnan, S.M., Marchand, L., Degnan, B.M. and Vervoort, M. (2011). Evolution of RNA-binding proteins in animals: Insights from genome-wide analysis in the sponge amphimedon queenslandica. *Molecular Biology and Evolution* **28**(8):2289-2303. doi: <https://doi.org/10.1093/molbev/msr046>.
- Kim-Ha, J., Kerr, K. and Macdonald, P.M. (1995). Translational regulation of oskar mRNA by Bruno, an ovarian RNA-binding protein, is essential. *Cell* **81**(3):403-412. doi: [https://doi.org/10.1016/0092-8674\(95\)90393-3](https://doi.org/10.1016/0092-8674(95)90393-3).
- Kim-Ha, J., Smith, J.L. and Macdonald, P.M. (1991). oskar mRNA is localized to the posterior pole of the *Drosophila* oocyte. *Cell* **66**(1):23-35.
- Kim, G., Pai, C.I., Sato, K., Person, M.D., Nakamura, A. and Macdonald, P.M. (2015). Region-Specific Activation of oskar mRNA Translation by Inhibition of Bruno-Mediated Repression. *PLoS Genetics* **11**(2):1-22. doi: <https://doi.org/10.1371/journal.pgen.1004992>.
- Kimura, M. and Imamoto, N. (2014). Biological Significance of the Importin- β Family-Dependent Nucleocytoplasmic Transport Pathways. *Traffic* **15**(7):727-748. doi: <https://doi.org/10.1111/tra.12174>.
- King, M.R., Matzat, L.H., Dale, R.K., Lim, J.J. and Lei, E.P. (2014). The RNA-binding protein Rumpelstiltskin antagonizes gypsy chromatin insulator function in a tissue-specific manner. *Journal of Cell Science* **127**(13):2956-2966. doi: <https://doi.org/10.1242/jcs.151126>.
- Klowden, M.J. (2009). Chapter 239 - Spermatheca. In: Resh, V. H. and Cardé, R. T. (eds.). *Encyclopedia of Insects - Second Edition*. Academic Press, pp. 939-940.
- Knapp, J.M., Chung, P. and Simpson, J.H. (2015). Generating customized transgene landing sites and multi-transgene arrays in *Drosophila* using phiC31 integrase. *Genetics* **199**(4):919-934. doi: <https://doi.org/10.1534/genetics.114.173187>.
- Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M. and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importin α . *Journal of Biological Chemistry* **284**(1):478-485. doi: <https://doi.org/10.1074/jbc.M807017200>.
- Kotadia, S., Crest, J., Tram, U., Riggs, B. and Sullivan, W. (2010). Blastoderm Formation and Cellularisation in *Drosophila melanogaster*. *eLS*:1-8. doi: <https://doi.org/10.1002/9780470015902.a0001071.pub2>.
- Kozlov, E., Shidlovskii, Y. V., Gilmutdinov, R., Schedl, P. and Zhukova, M. (2021). The role of CPEB family proteins in the nervous system function in the norm and pathology. *Cell and Bioscience* **11**(1):1-14. doi: <https://doi.org/10.1186/s13578-021->

00577-6.

Kramer, J.M. and Staveley, B.E. (2003). GAL4 causes developmental defects and apoptosis when expressed in the developing eye of *Drosophila melanogaster*. *Genetics and Molecular Research* **2**(1):43-47.

Krauss, J., López de Quinto, S., Nüsslein-Volhard, C. and Ephrussi, A. (2009). Myosin-V Regulates oskar mRNA Localization in the *Drosophila* Oocyte. *Current Biology* **19**(12):1058-1063. doi: <https://doi.org/10.1016/j.cub.2009.04.062>.

Kruys, V., Marinx, O., Shaw, G., Deschamps, J. and Huez, G. (1989). Translational blockade imposed by cytokine-derived UA-rich sequences. *Science* **245**(4920):852-855. doi: <https://doi.org/10.1126/science.2672333>.

Kuchin, S. (2011). Covering All the Bases in Genetics: Simple Shorthands and Diagrams for Teaching Base Pairing to Biology Undergraduates. *Journal of Microbiology & Biology Education* **12**(1):64-66. doi: <https://doi.org/10.1128/jmbe.v12i1.267>.

Kue, C.S. and Kumar, S. (2023). *Nonmammalian Models in Toxicology Screening*. 4th ed. Elsevier Inc.

Kuersten, S. and Goodwin, E.B. (2003). The power of the 3' UTR: Translational control and development. *Nature Reviews Genetics* **4**(8):626-637. doi: <https://doi.org/10.1038/nrg1125>.

Kugler, J.M. and Lasko, P. (2009). Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during *drosophila* oogenesis. *Fly* **3**(1):15-28. doi: <https://doi.org/10.4161/fly.3.1.7751>.

Lai, E.C. (2002). Keeping a good pathway down: Transcriptional repression of Notch pathway target genes by CSL proteins. *EMBO Reports* **3**(9):840-845. doi: <https://doi.org/10.1093/embo-reports/kvf170>.

Lamb, M.C., Kaluarachchi, C.P., Lansakara, T.I., Mellentine, S.Q., Lan, Y., Tivanski, A. V. and Tootle, T.L. (2021). Fascin limits myosin activity within *drosophila* border cells to control substrate stiffness and promote migration. *eLife* **10**:1-29. doi: <https://doi.org/10.7554/eLife.69836>.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E. and Corbett, A.H. (2007). Classical nuclear localization signals: Definition, function, and interaction with importin α . *Journal of Biological Chemistry* **282**(8):5101-5105. doi: <https://doi.org/10.1074/jbc.R600026200>.

Larkin, A., Marygold, S.J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V., ... Thurmond, J. (2020). FlyBase: Updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Research* **49**:D899-D907. doi: <https://doi.org/10.1093/nar/gkaa1026>.

Lasko, P. (2011). Posttranscriptional regulation in *Drosophila* oocytes and early embryos. *Wiley Interdisciplinary Reviews: RNA* **2**(3):408-416.

Lasko, P. (2012). mRNA localization and translational control in *Drosophila* oogenesis. *Cold Spring Harbor Perspectives in Biology* **4**(10):1-15. doi:

<https://doi.org/10.1101/cshperspect.a012294>.

Lasko, P. (2020). Patterning the *Drosophila* embryo: A paradigm for RNA-based developmental genetic regulation. *Wiley Interdisciplinary Reviews: RNA* 11(6):1-19. doi: <https://doi.org/10.1002/wrna.1610>.

Laughon, A. and Gesteland, R.F. (1982). Isolation and preliminary characterization of the GAL4 gene, a positive regulator of transcription in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 79(22 I):6827-6831. doi: <https://doi.org/10.1073/pnas.79.22.6827>.

Lebo, D.P.V. and McCall, K. (2021). Murder on the ovarian express: A tale of non-autonomous cell death in the *drosophila* ovary. *Cells* 10(6). doi: <https://doi.org/10.3390/cells10061454>.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., ... Krause, H.M. (2007). Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell* 131(1):174-187. doi: <https://doi.org/10.1016/j.cell.2007.08.003>.

Lehmann, R. (2016). *Chapter 39 - Germ Plasm Biogenesis-An Oskar-Centric Perspective*. 1st ed. Elsevier Inc.

Lehmann, R. and Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in *drosophila*. *Cell* 47(1):141-152.

LeMosy, E.K. (2003). Pattern formation: The eggshell holds the cue. *Current Biology* 13(13):508-510. doi: [https://doi.org/10.1016/S0960-9822\(03\)00441-X](https://doi.org/10.1016/S0960-9822(03)00441-X).

Leslie, D.M., Zhang, W., Timney, B.L., Chait, B.T., Rout, M.P., Wozniak, R.W. and Aitchison, J.D. (2004). Characterization of Karyopherin Cargoes Reveals Unique Mechanisms of Kap121p-Mediated Nuclear Import. *Molecular and Cellular Biology* 24(19):8487-8503. doi: <https://doi.org/10.1128/mcb.24.19.8487-8503.2004>.

Li, H., Janssens, J., de Waegeneer, M., Kolluru, S.S., Davie, K., Gardeux, V., ... Aerts, S. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* 375(6584). doi: <https://doi.org/10.1126/science.abk2432>.

Li, Y., Minor, N.T., Park, J.K., McKearin, D.M. and Maines, J.Z. (2009). Bam and Bgcn antagonize Nanos-dependent germ-line stem cell maintenance. *Proceedings of the National Academy of Sciences of the United States of America* 106(23):9304-9309. doi: <https://doi.org/10.1073/pnas.0901452106>.

Liao, J.Y., Yang, B., Zhang, Y.C., Wang, X.J., Ye, Y., Peng, J.W., ... Yin, D. (2020). EuRBPDB: A comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Research* 48(D1):D307-D313. doi: <https://doi.org/10.1093/nar/gkz823>.

Di Liegro, C.M., Schiera, G. and Di Liegro, I. (2014). Regulation of mRNA transport, localization and translation in the nervous system of mammals. *International Journal of Molecular Medicine* 33(4):747-762. doi: <https://doi.org/10.3892/ijmm.2014.1629>.

Liljas, A. (2013). Messenger RNA. *Brenner's Encyclopedia of Genetics: Second*

Edition 4:369–370. doi: <https://doi.org/10.1016/B978-0-12-374984-0.00929-3>.

Lim, C.S. and Brown, C.M. (2018). Know your enemy: Successful bioinformatic approaches to predict functional RNA structures in viral RNAs. *Frontiers in Microbiology* **8**:1–19. doi: <https://doi.org/10.3389/fmicb.2017.02582>.

Liu, Y. and Lehmann, M. (2008). A genomic response to the yeast transcription factor GAL4 in *Drosophila*. *Fly* **2**(2):92–98. doi: <https://doi.org/10.4161/fly.6311>.

Llense, F. and Martín-Blanco, E. (2008). JNK Signaling Controls Border Cell Cluster Integrity and Collective Cell Migration. *Current Biology* **18**(7):538–544. doi: <https://doi.org/10.1016/j.cub.2008.03.029>.

Lodish, H., Berk, A., Kaiser, C., Krieger, M., Bretscher, A., Ploegh, H., ... Martin, K. (2016). *Molecular Cell Biology 8th Edition*. New York: W.H. Freeman-Macmillan Learning.

López-Schier, H. and Johnston, D. St. (2001). Delta signaling from the germ line controls the proliferation and differentiation of the somatic follicle cells during *Drosophila* oogenesis. *Genes and Development* **15**(11):1393–1405. doi: <https://doi.org/10.1101/gad.200901.the>.

López-Schier, H. and St. Johnston, D. (2001). Delta signaling from the germ line controls the proliferation and differentiation of the somatic follicle cells during *Drosophila* oogenesis. *Genes and Development* **15**(11):1393–1405. doi: <https://doi.org/10.1101/gad.200901>.

Lowe, N., Rees, J.S., Roote, J., Ryder, E., Armean, I.M., Johnson, G., ... Johnston, D.S. (2014). Analysis of the expression patterns, Subcellular localisations and interaction partners of *drosophila* proteins using a pigp protein trap library. *Development (Cambridge)* **141**(20):3994–4005. doi: <https://doi.org/10.1242/dev.111054>.

Loyer, N., Kolotuev, I., Pinot, M. and Borgne, R. Le (2015). *Drosophila* E-cadherin is required for the maintenance of ring canals anchoring to mechanically withstand tissue growth. *PNAS* **112**(41):12717–12722. doi: <https://doi.org/10.1073/pnas.1814899115>.

Lu, C.-K., Lai, Y.-C., Chen, H.-R. and Chiang, M.-K. (2012). Rbms3, an RNA-binding protein, mediates the expression of Ptf1a by binding to its 3'UTR during mouse pancreas development. *DNA Cell Biol* **31**(7):1245–1251.

Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J. and Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling* **19**(1):1–10. doi: <https://doi.org/10.1186/s12964-021-00741-y>.

Lu, P. and Lu, Y. (2021). Born to Run? Diverse Modes of Epithelial Migration. *Frontiers in Cell and Developmental Biology* **9**(August):1–15. doi: <https://doi.org/10.3389/fcell.2021.704939>.

Lu, W., Lakonishok, M. and Gelfand, V.I. (2021). Gatekeeper function for Short stop at the ring canals of the *Drosophila* ovary. *Current Biology* **31**(15):3207–3220.e4. doi: <https://doi.org/10.1016/j.cub.2021.05.010>.

Lu, W., Lakonishok, M., Liu, R., Billington, N., Rich, A., Glotzer, M., ... Gelfand, V.I.

- (2020). Competition between kinesin-1 and myosin-v defines drosophila posterior determination. *eLife* **9**:1-25. doi: <https://doi.org/10.7554/eLife.54216>.
- Lu, W., Lakonishok, M., Serpinskaya, A.S. and Gelfand, V.I. (2022). A novel mechanism of bulk cytoplasmic transport by cortical dynein in *Drosophila* ovary. *eLife* **11**:1-29. doi: <https://doi.org/10.7554/ELIFE.75538>.
- Lu, W., Winding, M., Lakonishok, M., Wildonger, J. and Gelfand, V.I. (2016). Microtubule-microtubule sliding by kinesin-1 is essential for normal cytoplasmic streaming in *Drosophila* oocytes. *Proceedings of the National Academy of Sciences of the United States of America* **113**(34):E4995-E5004. doi: <https://doi.org/10.1073/pnas.1522424113>.
- Lue, N.F., Chasman, D.I., Buchman, A.R. and Kornberg, R.D. (1987). Interaction of GAL4 and GAL80 gene regulatory proteins in vitro. *Molecular and Cellular Biology* **7**(10):3446-3451. doi: <https://doi.org/10.1128/mcb.7.10.3446>.
- Lunde, B.M., Moore, C. and Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology* **8**(6):479-490. doi: <https://doi.org/10.1038/nrm2178>.
- Lynch, J.A. and Roth, S. (2011). The evolution of dorsal-ventral patterning mechanisms in insects. *Genes and Development* **25**(2):107-118. doi: <https://doi.org/10.1101/gad.2010711>.
- Ma, J. and Ptashne, M. (1987). The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80. *Cell* **50**(1):137-142.
- Mahajan-Miklos, S. and Cooley, L. (1994). Intercellular Cytoplasm Transport during *Drosophila* Oogenesis. *Developmental Biology* **165**(2):336-351.
- Majumdar, S. and Rio, D.C. (2015). P transposable elements in drosophila and other eukaryotic organisms. *Mobile DNA III*:727-752. doi: <https://doi.org/10.1128/9781555819217.ch33>.
- Malik, S., Jang, W., Kim, J.Y. and Kim, C. (2020). Mechanisms ensuring robust repression of the *Drosophila* female germline stem cell maintenance factor Nanos via posttranscriptional regulation. *The FASEB Journal* **34**(9):11421-11430.
- Mansfield, J.H., Wilhelm, J.E. and Hazelrigg, T. (2002). Ypsilon schachtel, a *Drosophila* Y-box protein, acts antagonistically to Orb in the oskar mRNA localization and translation pathway. *Development* **129**(1):197-209. doi: <https://doi.org/10.1242/dev.129.1.197>.
- Mapes, J., Chen, J.T., Yu, J.S. and Xue, D. (2010). Somatic sex determination in *Caenorhabditis elegans* is modulated by SUP-26 repression of tra-2 translation. *Proceedings of the National Academy of Sciences of the United States of America* **107**(42):18022-18027. doi: <https://doi.org/10.1073/pnas.1004513107>.
- Maris, C., Dominguez, C. and Allain, F.H.T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal* **272**(9):2118-2131. doi: <https://doi.org/10.1111/j.1742-4658.2005.04653.x>.
- Markstein, M. (2018). *Drosophila Workers Unite! A Laboratory Manual for Working*

with *Drosophila*. Amherst: University of Massachusetts Amherst.

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E. and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *40*(4). doi: <https://doi.org/10.1038/ng.101>.

Markussen, F.H., Michon, A.M., Breitwieser, W. and Ephrussi, A. (1995). Translational control of oskar generates short OSK, the isoform that induces pole plasm assembly. *Development* **121**(11):3723–3732. doi: <https://doi.org/10.1242/dev.121.11.3723>.

Masuda, A., Kawachi, T. and Ohno, K. (2021). Rapidly growing protein-centric technologies to extensively identify protein-rna interactions: Application to the analysis of co-transcriptional rna processing. *International Journal of Molecular Sciences* **22**(10). doi: <https://doi.org/10.3390/ijms22105312>.

Matoulkova, E., Michalova, E., Vojtesek, B. and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biology* **9**(5):563–576. doi: <https://doi.org/10.4161/rna.20231>.

Matsumoto, K., Toh-e, A. and Oshima, Y. (1978). Genetic control of galactokinase synthesis in *Saccharomyces cerevisiae*: evidence for constitutive expression of the positive regulatory gene *gal4*. *Journal of Bacteriology* **134**(2):446–457. doi: <https://doi.org/10.1128/jb.134.2.446-457.1978>.

Matsumoto, K., Wassarman, K.M. and Wolffe, A.P. (1998). Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO Journal* **17**(7):2107–2121. doi: <https://doi.org/10.1093/emboj/17.7.2107>.

Matzat, L.H., Dale, R.K., Moshkovich, N. and Lei, E.P. (2012). Tissue-Specific Regulation of Chromatin Insulator Function. *PLoS Genetics* **8**(11):1–15. doi: <https://doi.org/10.1371/journal.pgen.1003069>.

Mayr, C. (2019). What are 3' utrs doing? *Cold Spring Harbor Perspectives in Biology* **11**(10). doi: <https://doi.org/10.1101/cshperspect.a034728>.

Mazumder, B., Seshadri, V. and Fox, P.L. (2003). Translational control by the 3'-UTR: The ends specify the means. *Trends in Biochemical Sciences* **28**(2):91–98. doi: [https://doi.org/10.1016/S0968-0004\(03\)00002-1](https://doi.org/10.1016/S0968-0004(03)00002-1).

McCall, K. and Steller, H. (1998). Requirement for DCP-1 Caspase During *Drosophila* Oogenesis. *Science* **279**(5348):230–234.

McClure, C.D., Hassan, A., Aughey, G.N., Butt, K., Estacio-Gómez, A., Duggal, A., ... Southall, T.D. (2022). An auxin-inducible, GAL4-compatible, gene expression system for *Drosophila*. *eLife* **11**. doi: <https://doi.org/10.7554/eLife.67598>.

McDermott, S.M. and Davis, I. (2013). *Drosophila* Hephaestus/Polypyrimidine Tract Binding Protein Is Required for Dorso-Ventral Patterning and Regulation of Signalling between the Germline and Soma. *PLoS ONE* **8**(7). doi: <https://doi.org/10.1371/journal.pone.0069978>.

McGregor, J.R., Xi, R. and Harrison, D.A. (2002). JAK signaling is somatically required for follicle cell differentiation in *Drosophila*. *Development* **129**(3):705–717.

- McLane, L.M. and Corbett, A.H. (2009). Nuclear localization signals and human disease. *IUBMB Life* **61**(7):697–706. doi: <https://doi.org/10.1002/iub.194>.
- Medioni, C., Mowry, K. and Besse, F. (2012). Principles and roles of mRNA localization in animal development. *Development (Cambridge)* **139**(18):3263–3276. doi: <https://doi.org/10.1242/dev.078626>.
- Menendez-Gil, P. and Toledo-Arana, A. (2021). Bacterial 3'UTRs: A Useful Resource in Post-transcriptional Regulation. *Frontiers in Molecular Biosciences* **7**(January):1–12. doi: <https://doi.org/10.3389/fmolb.2020.617633>.
- Micklem, D.R., Adams, J., Grünert, S. and StJohnston, D. (2000). Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. *EMBO Journal* **19**(6):1366–1377. doi: <https://doi.org/10.1093/emboj/19.6.1366>.
- Mignone, F., Gissi, C., Liuni, S. and Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biology* **3**(3):1–10. doi: <https://doi.org/10.1186/gb-2002-3-3-reviews0004>.
- Milas, A. and Telley, I.A. (2022). Polarity Events in the *Drosophila melanogaster* Oocyte. **10**(May):1–13. doi: <https://doi.org/10.3389/fcell.2022.895876>.
- Misawa, K., Nosaka, T., Morita, S., Kaneko, A., Nakahata, T., Asano, S. and Kitamura, T. (2000). A method to identify cDNAs based on localization of green fluorescent protein fusion products. *Proceedings of the National Academy of Sciences of the United States of America* **97**(7):3062–3066. doi: <https://doi.org/10.1073/pnas.97.7.3062>.
- Miyamoto, Y., Yamada, K. and Yoneda, Y. (2016). Importin a: a key molecule in nuclear transport and non-transport functions. *Journal of Biochemistry* **160**(2):69–75. doi: <https://doi.org/10.1093/JB/MVW036>.
- Montell, D.J. (2003). Border-cell migration: The race is on. *Nature Reviews Molecular Cell Biology* **4**(1):13–24. doi: <https://doi.org/10.1038/nrm1006>.
- Montell, D.J., Rorth, P. and Spradling, A.C. (1992). slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes *Drosophila* CEBP. *Cell* **71**(1):51–62.
- Montell, D.J., Yoon, W.H. and Starz-Gaiano, M. (2012). Group choreography: Mechanisms orchestrating the collective movement of border cells. *Nature Reviews Molecular Cell Biology* **13**(10):631–645. doi: <https://doi.org/10.1038/nrm3433>.
- Moore, K.S. and von Lindern, M. (2018). RNA binding proteins and regulation of mRNA translation in erythropoiesis. *Frontiers in Physiology* **9**(JUL):1–17. doi: <https://doi.org/10.3389/fphys.2018.00910>.
- Morais-de-Sá, E., Vega-Rioja, A., Trovisco, V. and StJohnston, D. (2013). Oskar is targeted for degradation by the sequential action of Par-1, GSK-3, and the SCF-Slimb ubiquitin ligase. *Developmental Cell* **26**(3):303–314. doi: <https://doi.org/10.1016/j.devcel.2013.06.011>.
- Morel, V. and Schweisguth, F. (2000). Repression by Suppressor of Hairless and activation by Notch are required to define a single row of single-minded expressing cells

- in the *Drosophila* embryo. *Genes and Development* **14**(3):377–388. doi: <https://doi.org/10.1101/gad.14.3.377>.
- Morgan, T.H. (1910). Sex-Linked Inheritance in *Drosophila*. *Science* **32**(812):120-122.
- Morin, X., Daneman, R., Zavortink, M. and Chia, W. (2001). A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **98**(26):15050–15055. doi: <https://doi.org/10.1073/pnas.261408198>.
- Müller-Mcnicoll, M. and Neugebauer, K.M. (2013). How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics* **14**(4):275–287. doi: <https://doi.org/10.1038/nrg3434>.
- Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**(6301):515–520. doi: <https://doi.org/10.1038/348515a0>.
- Nakai, K. and Horton, P. (1999). PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Sciences* **24**(1):34–35. doi: [https://doi.org/10.1016/S0968-0004\(98\)01336-X](https://doi.org/10.1016/S0968-0004(98)01336-X).
- Neuman-Silberberg, F.S. and Schupbach, T. (1994). Dorsoventral axis formation in *Drosophila* depends on the correct dosage of the gene *gurken*. *Development* **120**(9):2457–2463.
- Nguyen Ba, A.N., Pogoutse, A., Provart, N. and Moses, A.M. (2009). NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10**:1–11. doi: <https://doi.org/10.1186/1471-2105-10-202>.
- Ni, J.Q., Liu, L.P., Binari, R., Hardy, R., Shim, H.S., Cavallaro, A., ... Perrimon, N. (2009). A *drosophila* resource of transgenic RNAi lines for neurogenetics. *Genetics* **182**(4):1089–1100. doi: <https://doi.org/10.1534/genetics.109.103630>.
- Ni, J.Q., Markstein, M., Binari, R., Pfeiffer, B., Liu, L.P., Villalta, C., ... Perrimon, N. (2008). Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. *Nature Methods* **5**(1):49–51. doi: <https://doi.org/10.1038/nmeth1146>.
- Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., ... Perrimon, N. (2011). A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nature Methods* **8**(5):405–407. doi: <https://doi.org/10.1038/nmeth.1592>.
- Nie, L., Wu, G. and Zhang, W. (2006). Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: A multiple regression to identify sources of variations. *Biochemical and Biophysical Research Communications* **339**(2):603–610. doi: <https://doi.org/10.1016/j.bbrc.2005.11.055>.
- Nieuwburg, R., Nashchekin, D., Jakobs, M., Carter, A.P., Trong, P.K., Goldstein, R.E. and St Johnston, D. (2017). Localised dynactin protects growing microtubules to deliver oskar mRNA to the posterior cortex of the *Drosophila* oocyte. *eLife* **6**:1–25. doi: <https://doi.org/10.7554/eLife.27237.001>.
- Niewiadowska, P., Godt, D. and Tepass, U. (1999). DE-cadherin is required for intercellular motility during *Drosophila* oogenesis. *Journal of Cell Biology* **144**(3):533–

547. doi: <https://doi.org/10.1083/jcb.144.3.533>.

Oka, M. and Yoneda, Y. (2018). Importin α : Functions as a nuclear transport factor and beyond. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* **94**(7):259–274. doi: <https://doi.org/10.2183/pjab.94.018>.

Olesnicky, E.C., Antonacci, S., Popitsch, N., Lybecker, M.C., Titus, M.B., Valadez, R., ... Killian, D.J. (2018). Shep interacts with posttranscriptional regulators to control dendrite morphogenesis in sensory neurons. *Developmental Biology* **444**(2):116–128. doi: <https://doi.org/10.1016/j.ydbio.2018.09.022>.

Ong, C., Yung, L.Y.L., Cai, Y., Bay, B.H. and Baeg, G.H. (2015). *Drosophila melanogaster* as a model organism to study nanotoxicity. *Nanotoxicology* **9**(3):396–403. doi: <https://doi.org/10.3109/17435390.2014.940405>.

Osterfield, M., Berg, C.A. and Shvartsman, S.Y. (2017). Epithelial Patterning, Morphogenesis, and Evolution: *Drosophila* Eggshell as a Model. *Developmental Cell* **41**(4):337–348. doi: <https://doi.org/10.1016/j.devcel.2017.02.018>.

Otsuka, H., Fukao, A., Funakami, Y., Duncan, K.E. and Fujiwara, T. (2019). Emerging evidence of translational control by AU-rich element-binding proteins. *Frontiers in Genetics* **10**(MAY):1–10. doi: <https://doi.org/10.3389/fgene.2019.00332>.

Oubridge, C., Ito, N., Evans, P., Teo, C.-H. and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**(6505):432–8.

Panin, V.M., Papayannopoulos, V., Wilson, R. and Irvine, K.D. (1997). Fringe modulates Notch-ligand interactions. *Nature* **387**(6636):908–912. doi: <https://doi.org/10.1038/43191>.

Parisi, M. and Lin, H. (2000). Translational repression: A duet of Nanos and Pumilio. *Current Biology* **10**(2):81–83. doi: [https://doi.org/10.1016/s0960-9822\(00\)00283-9](https://doi.org/10.1016/s0960-9822(00)00283-9).

Pascal, L.E., True, L.D., Campbell, D.S., Deutsch, E.W., Risk, M., Coleman, I.M., ... Liu, A.Y. (2008). Correlation of mRNA and protein levels: Cell type-specific gene expression of cluster designation antigens in the prostate. *BMC Genomics* **9**:1–13. doi: <https://doi.org/10.1186/1471-2164-9-246>.

Passmore, L.A. and Collier, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nature Reviews Molecular Cell Biology* **23**(2):93–106. doi: <https://doi.org/10.1038/s41580-021-00417-y>.

Pearson, J.R., Zurita, F., Tomás-Gallardo, L., Díaz-Torres, A., Díaz de la Loza, M. del C., Franze, K., ... González-Reyes, A. (2016). ECM-Regulator timp Is Required for Stem Cell Niche Organization and Cyst Production in the *Drosophila* Ovary. *PLoS Genetics* **12**(1):1–25. doi: <https://doi.org/10.1371/journal.pgen.1005763>.

Peng, S.S.-Y., Chen, C.-Y.A. and Shyu, A.-B. (1996). Functional Characterization of a Non-AUUUA AU-Rich Element from the c-jun Proto-Oncogene mRNA: Evidence for a Novel Class of AU-Rich Elements. *Molecular and Cellular Biology* **16**(4):1490–1499. doi: <https://doi.org/10.1128/mcb.16.4.1490>.

Peng, Y. and Gavis, E.R. (2022). The *Drosophila* hnRNP F/H homolog Glorund

recruits dFMRP to inhibit nanos translation elongation. *Nucleic Acids Research* **50**(12):7067–7083. doi: <https://doi.org/10.1093/nar/gkac500>.

Perez, C.A.G., Adachi, S., Nong, Q.D., Adhitama, N., Matsuura, T., Natsume, T., ... Watanabe, H. (2021). Sense-overlapping lncRNA as a decoy of translational repressor protein for dimorphic gene expression. *PLoS Genetics* **17**(7):1–18. doi: <https://doi.org/10.1371/journal.pgen.1009683>.

Pérez, I., McAfee, J.G. and Patton, J.G. (1997). Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* **36**(39):11881–11890. doi: <https://doi.org/10.1021/bi9711745>.

Peri, F. and Roth, S. (2000). Combined activities of Gurken and Decapentaplegic specify dorsal chorion structures of the *Drosophila* egg. *Development* **127**:841–850.

Perkins, L.A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., ... Perrimon, N. (2015). The transgenic RNAi project at Harvard medical school: Resources and validation. *Genetics* **201**(3):843–852. doi: <https://doi.org/10.1534/genetics.115.180208>.

Pesole, G., Liuni, S., Grillo, G., Licciulli, F., Mignone, F., Gissi, C. and Saccone, C. (2002). UTRdb and UTRsite: Specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002. *Nucleic Acids Research* **30**(1):335–340. doi: <https://doi.org/10.1093/nar/30.1.335>.

Pisignano, G. and Lodomery, M. (2021). Post-transcriptional regulation through long non-coding RNAs (lncRNAs). *Non-coding RNA* **7**(2):7–8. doi: <https://doi.org/10.3390/ncrna7020029>.

Poukkula, M., Cliffe, A., Changede, R. and Rørth, P. (2011). Cell behaviors regulated by guidance cues in collective migration of border cells. *Journal of Cell Biology* **192**(3):513–524. doi: <https://doi.org/10.1083/jcb.201010003>.

Poulton, J.S. and Deng, W.M. (2007). Cell-cell communication and axis specification in the *Drosophila* oocyte. *Developmental Biology* **311**(1):1–10. doi: <https://doi.org/10.1016/j.ydbio.2007.08.030>.

Prall, W., Sharma, B. and Gregory, B.D. (2019). Transcription Is Just the Beginning of Gene Expression Regulation: The Functional Significance of RNA-Binding Proteins to Post-transcriptional Processes in Plants. *Plant and Cell Physiology* **60**(9):1939–1952. doi: <https://doi.org/10.1093/pcp/pcz067>.

Prasad, M., Jang, A.C.-C., Starz-Gaiano, M., Melani, M. and Montell, D.J. (2007). A protocol for culturing *Drosophila melanogaster* stage 9 egg chambers for live imaging. *Nature Protocols* **2**:2467–2473.

Prasad, M. and Montell, D.J. (2007). Cellular and Molecular Mechanisms of Border Cell Migration Analyzed Using Time-Lapse Live-Cell Imaging. *Developmental Cell* **12**(6):997–1005. doi: <https://doi.org/10.1016/j.devcel.2007.03.021>.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G. and Cormier, M.J. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111**(2):229–233. doi: [https://doi.org/10.1016/0378-1119\(92\)90691-H](https://doi.org/10.1016/0378-1119(92)90691-H).

Pulver, S.R. and Berni, J. (2012). The fundamentals of flying: Simple and inexpensive

strategies for employing *Drosophila* genetics in neuroscience teaching laboratories. *Journal of Undergraduate Neuroscience Education* 11(1).

Quan, H. and Lynch, J.A. (2016). The evolution of insect germline specification strategies. *Current Opinion in Insect Science* 13:99–105. doi: <https://doi.org/10.1016/j.cois.2016.02.013>.

Queenan Anne Marie, Ghabrial Amin and Schüpbach Trudi (1997). Ectopic activation of torpedo/Egfr, a *Drosophila* receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. *Development* 124:3871–3880.

Quellhorst, G. and Rulli, S. (2012). A systematic guideline for developing the best real-time PCR primers Lessons learned from designing assays for more than 14 , 000 genes. *Qiagen*:1–9.

Quinlan, M.E. (2016). Cytoplasmic Streaming in the *Drosophila* Oocyte. *Annual Review of Cell and Developmental Biology* 32:173–195. doi: <https://doi.org/10.1146/annurev-cellbio-111315-125416>.

Quiñones-Coello, A.T., Petrella, L.N., Ayers, K., Melillo, A., Mazzalupo, S., Hudson, A.M., ... Cooley, L. (2007). Exploring strategies for protein trapping in *drosophila*. *Genetics* 175(3):1089–1104. doi: <https://doi.org/10.1534/genetics.106.065995>.

Rajeshkannan, Mahilkar, A. and Saini, S. (2022). GAL Regulon in the Yeast *S. cerevisiae* is Highly Evolvable via Acquisition in the Coding Regions of the Regulatory Elements of the Network. *Frontiers in Molecular Biosciences* 9(March):1–14. doi: <https://doi.org/10.3389/fmolb.2022.801011>.

Ray, D., Ha, K.C.H., Nie, K., Zheng, H., Hughes, T.R. and Morris, Q.D. (2017). RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. *Methods* 118–119:3–15. doi: <https://doi.org/10.1016/j.ymeth.2016.12.003>.RNAcompete.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., ... Hughes, T.R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature Biotechnology* 499(7457):172–177. doi: <https://doi.org/10.1038/nature12311>.

Reiter, L.T., Potocki, L., Chien, S., Gribskov, M. and Bier, E. (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Research* 11(6):1114–1125. doi: <https://doi.org/10.1101/gr.169101>.

Richardson, B.E. and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: Strategies from different organisms. *Nature Reviews Molecular Cell Biology* 11(1):37–49. doi: <https://doi.org/10.1038/nrm2815>.

Ridley, A.J., Schwartz, M.A., Burridge, K., Firtel, R.A., Ginsberg, M.H., Borisy, G., ... Horwitz, A.R. (2003). Cell Migration: Integrating Signals from Front to Back. *Science* 302(5651):1704–1709.

Riechmann, V. and Ephrussi, A. (2001). Axis formation during *Drosophila* oogenesis. *Current Opinion in Genetics and Development* 11:374–383.

Riechmann, V., Gutierrez, G.J., Filardo, P., Nebreda, A.R. and Ephrussi, A. (2002). Par-1 regulates stability of the posterior determinant Oskar by phosphorylation. *Nature*

Cell Biology **4**(5):337–342. doi: <https://doi.org/10.1038/ncb782>.

Ripin, N., Boudet, J., Duszczuk, M.M., Hinniger, A., Faller, M., Krepl, M., ... Allain, F.H.T. (2019). Molecular basis for AU-rich element recognition and dimerization by the HuR C-terminal RRM. *Proceedings of the National Academy of Sciences of the United States of America* **116**(8):2935–2944. doi: <https://doi.org/10.1073/pnas.1808696116>.

Romanelli, M.G., Diani, E. and Lievens, P.M.J. (2013). New insights into functional roles of the polypyrimidine tract-binding protein. *International Journal of Molecular Sciences* **14**(11):22906–22932. doi: <https://doi.org/10.3390/ijms141122906>.

Romanelli, M.G. and Morandi, C. (2002). Importin α binds to an unusual bipartite nuclear localization signal in the heterogeneous ribonucleoprotein type I. *European Journal of Biochemistry* **269**(11):2727–2734. doi: <https://doi.org/10.1046/j.1432-1033.2002.02942.x>.

Rongo, C., Gavis, E.R. and Lehmann, R. (1995). Localization of oskar RNA regulates oskar translation and requires Oskar protein. *Development* **121**(9):2737–2746. doi: <https://doi.org/10.1242/dev.121.9.2737>.

Roote, J. and Prokop, A. (2013). A Rough Guide to Drosophila Mating Schemes. *G3* **3**:353–358. doi: <https://doi.org/10.1038/nmeth0907-675>.

Rorth, P. (2002). Initiating and guiding migration: Lessons from border cells. *Trends in Cell Biology* **12**(7):325–331. doi: [https://doi.org/10.1016/S0962-8924\(02\)02311-5](https://doi.org/10.1016/S0962-8924(02)02311-5).

Rørth, P. (1996). A modular misexpression screen in Drosophila detecting tissue-specific phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **93**(22):12418–12422. doi: <https://doi.org/10.1073/pnas.93.22.12418>.

Rørth, P. (1998). Gal4 in the Drosophila female germline. *Mechanisms of Development* **78**(1–2):113–118. doi: [https://doi.org/10.1016/S0925-4773\(98\)00157-9](https://doi.org/10.1016/S0925-4773(98)00157-9).

Rørth, P., Szabo, K., Bailey, A., Lavery, T., Rehm, J., Rubin, G.M., ... Cohen, S.M. (1998). Systematic gain-of-function genetics in Drosophila. *Development* **125**(6):1049–1057.

Rørth, P., Szabo, K. and Texido, G. (2000). The level of C/EBP protein is critical for cell migration during Drosophila oogenesis and is tightly controlled by regulated degradation. *Molecular Cell* **6**(1):23–30. doi: [https://doi.org/10.1016/S1097-2765\(05\)00008-0](https://doi.org/10.1016/S1097-2765(05)00008-0).

Roth, S. and Lynch, J.A. (2009). Symmetry breaking during Drosophila oogenesis. *Cold Spring Harbor perspectives in biology* **1**(2):1–22. doi: <https://doi.org/10.1101/cshperspect.a001891>.

Rouhana, L., Edgar, A., Hugosson, F., Dountcheva, V., Martindale, M.Q. and Ryan, J.F. (2023). Cytoplasmic Polyadenylation Is an Ancestral Hallmark of Early Development in Animals. *Molecular Biology and Evolution* **40**(6):1–19. doi: <https://doi.org/10.1093/molbev/msad137>.

Routh, S.B. and Sankaranarayanan, R. (2017). *Mechanistic Insights Into Catalytic RNA-Protein Complexes Involved in Translation of the Genetic Code*. 1st ed. Elsevier

Inc.

Sahut-Barnola, I., Dastugue, B. and Couderc, J.L. (1996). Terminal filament cell organization in the larval ovary of *Drosophila melanogaster*: Ultrastructural observations and pattern of divisions. *Roux's Archives of Developmental Biology* **205**(7-8):356-363. doi: <https://doi.org/10.1007/BF00377215>.

Sambrook, J. and Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual. 3rd Edition*. New York: Cold Spring Harbor Laboratory Press.

Santos, A.C. and Lehmann, R. (2004). Germ cell specification and migration in *Drosophila* and beyond. *Current Biology* **14**(14):578-589. doi: <https://doi.org/10.1016/j.cub.2004.07.018>.

Santosh, B., Varshney, A. and Yadava, P.K. (2015). Non-coding RNAs: Biological functions and applications. *Cell Biochemistry and Function* **33**(1):14-22. doi: <https://doi.org/10.1002/cbf.3079>.

Sarikaya, D.P., Belay, A.A., Ahuja, A., Dorta, A., Green, D.A. and Extavour, C.G. (2012). The roles of cell size and cell number in determining ovariole number in *Drosophila*. *Developmental Biology* **363**(1):279-289. doi: <https://doi.org/10.1016/j.ydbio.2011.12.017>.

Sarikaya, D.P. and Extavour, C.G. (2015). The Hippo Pathway Regulates Homeostatic Growth of Stem Cell Niche Precursors in the *Drosophila* Ovary. *PLoS Genetics* **11**(2):1-28. doi: <https://doi.org/10.1371/journal.pgen.1004962>.

Saunders, C. and Cohen, R.S. (1999). The role of oocyte transcription, the 5'UTR, and translation repression and derepression in *Drosophila* gurken mRNA and protein localization. *Molecular Cell* **3**(1):43-54. doi: [https://doi.org/10.1016/S1097-2765\(00\)80173-2](https://doi.org/10.1016/S1097-2765(00)80173-2).

Savant-Bhonsale, S. and Montell, D.J. (1993). torso-like encodes the localized determinant of *Drosophila* terminal pattern formation. *Genes and Development* **7**(12 B):2548-2555. doi: <https://doi.org/10.1101/gad.7.12b.2548>.

Schachtner, L.T., Sola, I.E., Forand, D., Antonacci, S., Postovit, A.J., Mortimer, N.T., ... Olesnický, E.C. (2015). *Drosophila* Shep and *C. elegans* SUP-26 are RNA-binding proteins that play diverse roles in nervous system development. *Development Genes and Evolution* **225**(6):319-330. doi: <https://doi.org/10.1007/s00427-015-0514-3>.

Schäfer, U. and Jäckle, H. (2000). *Drosophila Melanogaster: A Genetic Tool. Comparative Genomics*. Boston, MA: Springer.

Schenborn, E.T. and Mierendorf, R.C. (1985). A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Research* **13**(17):6223-6236.

Schinko, J.B., Weber, M., Viktorinova, I., Kiupakis, A., Averof, M., Klingler, M., ... Bucher, G. (2010). Functionality of the GAL4/UAS system in *Tribolium* requires the use of endogenous core promoters. *BMC Developmental Biology* **10**. doi: <https://doi.org/10.1186/1471-213X-10-53>.

Schnakenberg, S.L., Matias, W.R. and Siegal, M.L. (2011). Sperm-storage defects and

live birth in drosophila females lacking spermathecal secretory cells. *PLoS Biology* 9(11). doi: <https://doi.org/10.1371/journal.pbio.1001192>.

Scott, M.S., Boisvert, F.M., McDowall, M.D., Lamond, A.I. and Barton, G.J. (2010). Characterization and prediction of protein nucleolar localization sequences. *Nucleic Acids Research* 38(21):7388–7399. doi: <https://doi.org/10.1093/nar/gkq653>.

Scott, M.S., Troshin, P. V. and Barton, G.J. (2011). NoD: A Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics* 12. doi: <https://doi.org/10.1186/1471-2105-12-317>.

SenGupta, D. (2013). *RNA-Binding Domains in Proteins*. Elsevier Inc.

Şentürk, M. and Bellen, H.J. (2018). Genetic strategies to tackle neurological diseases in fruit flies. *Current Opinion in Neurobiology* 50:24–32. doi: <https://doi.org/10.1016/j.conb.2017.10.017>.

Serano, T.L., Karlin-McGinness, M. and Cohen, R.S. (1995). The role of fs(1)K10 in the localization of the mRNA of the TGF α homolog gurken within the Drosophila oocyte. *Mechanisms of Development* 51(2–3):183–192. doi: [https://doi.org/10.1016/0925-4773\(95\)00363-0](https://doi.org/10.1016/0925-4773(95)00363-0).

Shiota, C., Coffey, J., Grimsby, J., Grippo, J.F. and Magnuson, M.A. (1999). Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase. *Journal of Biological Chemistry* 274(52):37125–37130. doi: <https://doi.org/10.1074/jbc.274.52.37125>.

Shulman, J.M., Benton, R. and St Johnston, D. (2000). The Drosophila homolog of C. elegans PAR-1 organizes the oocyte cytoskeleton and directs oskar mRNA localization to the posterior pole. *Cell* 101(4):377–388. doi: [https://doi.org/10.1016/S0092-8674\(00\)80848-X](https://doi.org/10.1016/S0092-8674(00)80848-X).

Silver, D.L. and Montell, D.J. (2001). Paracrine signaling through the JAK/STAT pathway activates invasive behavior of ovarian epithelial cells in drosophila. *Cell* 107(7):831–841. doi: [https://doi.org/10.1016/S0092-8674\(01\)00607-9](https://doi.org/10.1016/S0092-8674(01)00607-9).

Simmons, M.J., Haley, K.J. and Thompson, S.J. (2002). Maternal transmission of P element transposase activity in Drosophila melanogaster depends on the last P intron. *Proceedings of the National Academy of Sciences of the United States of America* 99(14):9306–9309. doi: <https://doi.org/10.1073/pnas.142304199>.

Singh, R. and Valcárcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nature Structural and Molecular Biology* 12(8):645–653. doi: <https://doi.org/10.1038/nsmb961>.

Slaidina, M. and Lehmann, R. (2014). Translational control in germline stem cell development. *Journal of Cell Biology* 207(1):13–21. doi: <https://doi.org/10.1083/jcb.201407102>.

Smith, D. (1997). Mini-Exon Epitope Tagging for Analysis of the Protein Coding Potential of Genomic Sequence. *Biotechniques* 23:942–5.

Smith, J.L., Wilson, J.E. and Macdonald, P.M. (1992). Overexpression of oskar directs ectopic activation of nanos and presumptive pole cell formation in Drosophila embryos.

Cell **70**(5):849–859. doi: [https://doi.org/10.1016/0092-8674\(92\)90318-7](https://doi.org/10.1016/0092-8674(92)90318-7).

Spemann, H. (1938). *Embryonic Development and Induction*. New Haven, CT: Yale University Press.

Sponer, J., Bussi, G., Krepl, M., Banas, P., Bottaro, S., Cunha, R.A., ... Otyepka, M. (2018). RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chemical Reviews* **118**(8):4177–4338. doi: <https://doi.org/10.1021/acs.chemrev.7b00427>.

Spracklen, A. and Tootle, T. (2013). The Utility of Stage-specific Mid-to-late Drosophila Follicle Isolation. *Journal of Visualized Experiments*(82).

St Johnston, D. (2023). *Polarity and Axis Formation in the Drosophila Female Germ Line*. Elsevier.

Statello, L., Guo, C.J., Chen, L.L. and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology* **22**(2):96–118. doi: <https://doi.org/10.1038/s41580-020-00315-9>.

Stein, D.S. and Stevens, L.M. (2014). Maternal control of the Drosophila dorsal-ventral body axis. *Wiley Interdisciplinary Reviews: Developmental Biology* **3**(5):301–330. doi: <https://doi.org/10.1002/wdev.138>.

Steinhauer, J. and Kalderon, D. (2006). Microtubule polarity and axis formation in the Drosophila oocyte. *Developmental Dynamics* **235**(6):1455–1468. doi: <https://doi.org/10.1002/dvdy.20770>.

von Stetina, J.R. and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. *Cold Spring Harbor Perspectives in Biology* **3**(10):1–19. doi: <https://doi.org/10.1101/cshperspect.a005553>.

Stonko, D.P., Manning, L., Starz-Gaiano, M. and Peercy, B.E. (2015). A mathematical model of collective cell migration in a three-dimensional, heterogeneous environment. *PLoS ONE* **10**(4):1–19. doi: <https://doi.org/10.1371/journal.pone.0122799>.

Stroebele, E. and Erives, A. (2016). A Notch and Su(H) dependent enhancer complex coordinates expression of nab in Drosophila. **XXX**(September):1–23. doi: <https://doi.org/10.1534/genetics.XXX.XXXXXX>.

Strömnæs, Ö. and Kvelland, I. (1962). Sexual Activity of Drosophila Melanogaster Males. *Hereditas* **48**(3):442–470. doi: <https://doi.org/10.1111/j.1601-5223.1962.tb01826.x>.

Studier, F.W. and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189**(1):113–130. doi: [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2).

Sun, X., Hu, Y., Wu, J., Shi, L., Zhu, L., Xi, P.W., ... Ding, Q. (2018). RBMS2 inhibits the proliferation by stabilizing P21 mRNA in breast cancer. *Journal of Experimental and Clinical Cancer Research* **37**(1):1–14.

Sun, Y., Yolitz, J., Wang, C., Spangler, E., Zhan, M. and Zou, S. (2013). Aging studies in drosophila melanogaster. *Methods in Molecular Biology* **1048**:77–93. doi: https://doi.org/10.1007/978-1-62703-556-9_7.

- Suresh Babu, S., Joladarashi, D., Jeyabal, P., Thandavarayan, R.A. and Krishnamurthy, P. (2015). RNA-stabilizing proteins as molecular targets in cardiovascular pathologies. *Trends in Cardiovascular Medicine* 25(8):676–683. doi: <https://doi.org/10.1016/j.tcm.2015.02.006>.
- Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., ... Burge, C.B. (2016). RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Molecular Cell* 64(2):294–306. doi: <https://doi.org/10.1016/j.molcel.2016.08.035>.
- Tanaka, T. and Nakamura, A. (2011). Oskar-induced endocytic activation and actin remodeling for anchorage of the Drosophila germ plasm. *BioArchitecture* 1(3):122–126. doi: <https://doi.org/10.4161/bioa.1.3.17313>.
- Taormina, G., Ferrante, F., Vieni, S., Grassi, N., Russo, A. and Mirisola, M.G. (2019). Longevity: Lesson from model organisms. *Genes* 10(7). doi: <https://doi.org/10.3390/genes10070518>.
- Tarikere, S., Ylla, G. and Extavour, C.G. (2022). Distinct gene expression dynamics in germ line and somatic tissue during ovariole morphogenesis in Drosophila melanogaster. *G3: Genes, Genomes, Genetics* 12(2). doi: <https://doi.org/10.1093/g3journal/jkab305>.
- Tillery, M.M.L., Blake-Hedges, C., Zheng, Y., Buchwalter, R.A. and Megraw, T.L. (2018). Centrosomal and non-centrosomal microtubule-organizing centers (MTOCs) in Drosophila melanogaster. *Cells* 7(9). doi: <https://doi.org/10.3390/cells7090121>.
- Timmons, A.K., Mondragon, A.A., Meehan, T.L. and McCall, K. (2017). Control of non-apoptotic nurse cell death by engulfment genes in Drosophila. *Fly* 11(2):104–111. doi: <https://doi.org/10.1080/19336934.2016.1238993>.
- Timmons, A.K., Mondragon, A.A., Schenkel, C.E., Yalonetskaya, A., Taylor, J.D., Moynihan, K.E., ... McCall, K. (2016). Phagocytosis genes nonautonomously promote developmental cell death in the Drosophila ovary. *Proceedings of the National Academy of Sciences of the United States of America* 113(9):E1246–E1255. doi: <https://doi.org/10.1073/pnas.1522830113>.
- Torres, I.L., López-Schier, H. and St. Johnston, D. (2003). A notch/delta-dependent relay mechanism establishes anterior-posterior polarity in Drosophila. *Developmental Cell* 5(4):547–558. doi: [https://doi.org/10.1016/S1534-5807\(03\)00272-7](https://doi.org/10.1016/S1534-5807(03)00272-7).
- Traven, A., Jelacic, B. and Sopta, M. (2006). Yeast Gal4: A transcriptional paradigm revisited. *EMBO Reports* 7(5):496–499. doi: <https://doi.org/10.1038/sj.embor.7400679>.
- Trcek, T. and Singer, R.H. (2010). The cytoplasmic fate of an mRNP is determined cotranscriptionally: Exception or rule? *Genes and Development* 24(17):1827–1831. doi: <https://doi.org/10.1101/gad.1972810>.
- Triana-Alonso, F.J., Dabrowski, M., Wadzack, J. and Nierhaus, K.H. (1995). Self-coded 3'-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. *Journal of Biological Chemistry* 270(11):6298–6307. doi: <https://doi.org/10.1074/jbc.270.11.6298>.

- Tsai, K. and Cullen, B.R. (2020). Epigenetic and epitranscriptomic regulation of viral replication. *Nature Reviews Microbiology* **18**(10):559–570. doi: <https://doi.org/10.1038/s41579-020-0382-3>.
- Tsai, Y.S., Gomez, S.M. and Wang, Z. (2014). Prevalent RNA recognition motif duplication in the human genome. *RNA* **20**(5):702–712. doi: <https://doi.org/10.1261/rna.044081.113>.
- Tunstall, N.E., Herr, A., de Bruyne, M. and Warr, C.G. (2012). A screen for genes expressed in the olfactory organs of *Drosophila melanogaster* identifies genes involved in olfactory behaviour. *PLoS ONE* **7**(4). doi: <https://doi.org/10.1371/journal.pone.0035641>.
- Vacher, J., Grosjean, H., Houssier, C. and Buckingham, R.H. (1984). The effect of point mutations affecting *Escherichia coli* tryptophan tRNA on anticodon-anticodon interactions and on UGA suppression. *Journal of Molecular Biology* **177**(2):329–342. doi: [https://doi.org/10.1016/0022-2836\(84\)90460-1](https://doi.org/10.1016/0022-2836(84)90460-1).
- Van Assche, E., Van Puyvelde, S., Vanderleyden, J. and Steenackers, H.P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Frontiers in Microbiology* **6**(MAR):1–16. doi: <https://doi.org/10.3389/fmicb.2015.00141>.
- Vandivier, L.E., Anderson, S.J., Foley, S.W. and Gregory, B.D. (2016). The Conservation and Function of RNA Secondary Structure in Plants. *Annual Review of Plant Biology* **67**:463–488. doi: <https://doi.org/10.1146/annurev-arplant-043015-111754>.
- Vanzo, N., Oprins, A., Xanthakis, D., Ephrussi, A. and Rabouille, C. (2007). Stimulation of Endocytosis and Actin Dynamics by Oskar Polarizes the *Drosophila* Oocyte. *Developmental Cell* **12**(4):543–555. doi: <https://doi.org/10.1016/j.devcel.2007.03.002>.
- Vanzo, N.F. and Ephrussi, A. (2002). Oskar anchoring restricts pole plasm formation to the posterior of the *Drosophila* oocyte. *Development* **129**(15):3705–3714.
- Varani, G. and McClain, W.H. (2000). The G-U wobble base pair diverse biological systems. *EMBO Reports* **1**(1):18–23.
- Vazquez-Pianzola, P. and Suter, B. (2012). Conservation of the RNA transport machineries and their coupling to translation control across eukaryotes. *Comparative and Functional Genomics* **2012**. doi: <https://doi.org/10.1155/2012/287852>.
- Vazquez-Pianzola, P., Urlaub, H. and Suter, B. (2011). Pabp binds to the *osk* 3'UTR and specifically contributes to *osk* mRNA stability and oocyte accumulation. *Developmental Biology* **357**(2):404–418. doi: <https://doi.org/10.1016/j.ydbio.2011.07.009>.
- Veeman, M.T. and McDonald, J.A. (2016). Dynamics of cell polarity in tissue morphogenesis: A comparative view from *Drosophila* and *Ciona*. *F1000Research* **5**. doi: <https://doi.org/10.12688/F1000RESEARCH.8011.1>.
- Villalba, A., Coll, O. and Gebauer, F. (2011). Cytoplasmic polyadenylation and translational control. *Current Opinion in Genetics and Development* **21**(4):452–457. doi: <https://doi.org/10.1016/j.gde.2011.04.006>.

- Vogel, C. and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**(4):227–232. doi: <https://doi.org/10.1038/nrg3185>.
- Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. and Chang, H.Y. (2011). Understanding the transcriptome through RNA structure. *Nature Reviews Genetics* **12**(9):641–655. doi: <https://doi.org/10.1038/nrg3049>.
- Wang, M., Ly, M., Lugowski, A., Laver, J.D., Lipshitz, H.D., Smibert, C.A. and Rissland, O.S. (2017). ME31B globally represses maternal mRNAs by two distinct mechanisms during the *Drosophila* maternal-to-zygotic transition. *eLife* **6**:1–22. doi: <https://doi.org/10.7554/eLife.27891>.
- Wang, X., Adam, J.C. and Montell, D. (2007). Spatially localized Kuzbanian required for specific activation of Notch during border cell migration. *Developmental Biology* **301**(2):532–540. doi: <https://doi.org/10.1016/j.ydbio.2006.08.031>.
- Wang, Z. and Lin, H. (2004). Nanos Maintains Germline Stem Cell Self-Renewal by Preventing Differentiation. *Science* **303**(5666):2016–2019.
- Weis, K. (2003). Regulating access to the genome: Nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**(4):441–451. doi: [https://doi.org/10.1016/S0092-8674\(03\)00082-5](https://doi.org/10.1016/S0092-8674(03)00082-5).
- Weiß, J., Rosemann, J., Krauspe, V., Kappler, M., Eckert, A.W., Haemmerle, M. and Gutschner, T. (2020). RNA-binding proteins as regulators of migration, invasion and metastasis in oral squamous cell carcinoma. *International Journal of Molecular Sciences* **21**(18):1–28. doi: <https://doi.org/10.3390/ijms21186835>.
- Wilhelm, J.E., Buszczak, M. and Sayles, S. (2005). Efficient protein trafficking requires trailer hitch, a component of a ribonucleoprotein complex localized to the ER in *Drosophila*. *Developmental Cell* **9**(5):675–685. doi: <https://doi.org/10.1016/j.devcel.2005.09.015>.
- Wilmut, I., Bai, Y. and Taylor, J. (2015). Somatic cell nuclear transfer: Origins, the present position and future opportunities. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**(1680). doi: <https://doi.org/10.1098/rstb.2014.0366>.
- Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J. and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* **386**(6621):200. doi: <https://doi.org/10.1038/386200a0>.
- Wilson, J.E., Connell, J.E. and Macdonald, P.M. (1996). Aubergine enhances oskar translation in the *Drosophila* ovary. *Development* **122**(5):1631–1639. doi: <https://doi.org/10.1242/dev.122.5.1631>.
- Wolf, M.J. and Rockman, H.A. (2011). *Drosophila*, genetic screens, and cardiac function. *Circulation Research* **109**(7):794–806. doi: <https://doi.org/10.1161/CIRCRESAHA.111.244897>.
- Wolpert, L., Tickle, C. and Arias, A.M. (2019). *Principles of Development*. New York, NY, USA: Oxford University Press.
- Wong, li C. and Schedl, P. (2006). Dissection of *drosophila* ovaries. *Journal of*

Visualized Experiments 8(1):2006. doi: <https://doi.org/10.3791/52>.

Wu, X., Tanwar, P.S. and Raftery, L.A. (2008). Drosophila follicle cells: Morphogenesis in an eggshell. *Seminars in Cell and Developmental Biology* 19(3):271-282. doi: <https://doi.org/10.1016/j.semcd.2008.01.004>.

Xia, Y.P., Yeh, C.T., Ou, J.H. and Lai, M.M. (1992). Characterization of nuclear targeting signal of hepatitis delta antigen: nuclear transport as a protein complex. *Journal of Virology* 66(2):914-921. doi: <https://doi.org/10.1128/jvi.66.2.914-921.1992>.

Xiang, Y., Takeo, S., Florens, L., Hughes L, S.E., Huo, I.J., Gilliland, W.D., ... Hawley, R.S. (2007). The inhibition of polo kinase by matrimony maintains G2 arrest in the meiotic cell cycle. *PLoS Biology* 5(12):2831-2846. doi: <https://doi.org/10.1371/journal.pbio.0050323>.

Xiao, R., Chen, J.Y., Liang, Z., Luo, D., Chen, G., Lu, Z.J., ... Fu, X.D. (2019). Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* 178(1):107-121.e18. doi: <https://doi.org/10.1016/j.cell.2019.06.001>.

Xu, J. and Gridley, T. (2012). Notch Signaling during Oogenesis in Drosophila melanogaster . *Genetics Research International* 2012(Type 1):1-10. doi: <https://doi.org/10.1155/2012/648207>.

Yamada, T., Nagahama, M. and Akimitsu, N. (2018). Interplay between Transcription and RNA Degradation. *Gene Expression and Regulation in Mammalian Cells - Transcription From General Aspects*. doi: <https://doi.org/10.5772/intechopen.71862>.

Yang, N., Yu, Z., Hu, M., Wang, M., Lehmann, R. and Xu, R.M. (2015). Structure of Drosophila Oskar reveals a novel RNA binding protein. *Proceedings of the National Academy of Sciences of the United States of America* 112(37):11541-11546. doi: <https://doi.org/10.1073/pnas.1515568112>.

Yang, Y. and Wang, Z. (2019). IRES-mediated cap-independent translation, a path leading to hidden proteome. *Journal of Molecular Cell Biology* 11(10):911-919. doi: <https://doi.org/10.1093/jmcb/mjz091>.

Yano, T., de Quinto, S.L., Matsui, Y., Shevchenko, Anna, Shevchenko, Andrej and Ephrussi, A. (2004). Hrp48, a Drosophila hnRNPA/B homolog, binds and regulates translation of oskar mRNA. *Developmental Cell* 6(5):637-648. doi: [https://doi.org/10.1016/S1534-5807\(04\)00132-7](https://doi.org/10.1016/S1534-5807(04)00132-7).

Yarunin, A., Harris, R.E., Ashe, M.P. and Ashe, H.L. (2011). Patterning of the Drosophila oocyte by a sequential translation repression program involving the d4EHP and Belle translational repressors. *RNA Biology* 8(5):904-912. doi: <https://doi.org/10.4161/rna.8.5.16325>.

Yu, J., Poulton, J., Huang, Y.C. and Deng, W.M. (2008). The Hippo pathway promotes Notch signaling in regulation of cell differentiation, proliferation, and oocyte polarity. *PLoS ONE* 3(3). doi: <https://doi.org/10.1371/journal.pone.0001761>.

Yuan, Z., Praxenthaler, H., Tabaja, N., Torella, R., Preiss, A., Maier, D. and Kovall, R.A. (2016). Structure and Function of the Su(H)-Hairless Repressor Complex, the Major Antagonist of Notch Signaling in Drosophila melanogaster. *PLoS Biology*

14(7):1-26. doi: <https://doi.org/10.1371/journal.pbio.1002509>.

Zhang, X.B., Dong, W., Li, K.X., Wang, J.J., Shen, J., Moussian, B. and Zhang, J.Z. (2020). Flexible manipulation of Omb levels in the endogenous expression region of *Drosophila* wing by combinational overexpression and suppression strategy. *Insect Science* **27**(1):14-21. doi: <https://doi.org/10.1111/1744-7917.12705>.

Zhao, B.S., Roundtree, I.A. and He, C. (2016). Post-transcriptional gene regulation by mRNA modifications. *Nature Reviews Molecular Cell Biology* **18**(1):31-42. doi: <https://doi.org/10.1038/nrm.2016.132>.

Zhao, T., Graham, O.S., Raposo, A. and StJohnston, D. (2012). Growing microtubules push the oocyte nucleus to polarize the *Drosophila* dorsal-ventral axis. *Science* **336**(6084):999-1003. doi: <https://doi.org/10.1126/science.1219147>.

Zhao, T., Xiao, Y., Huang, B., Ran, M.J., Duan, X., Wang, Y.F., ... Yu, X.Q. (2022). A dual role of lola in *Drosophila* ovary development: regulating stem cell niche establishment and repressing apoptosis. *Cell Death and Disease* **13**(9):1-9. doi: <https://doi.org/10.1038/s41419-022-05195-9>.

Zhu, L., Xi, P.W., Li, X.X., Sun, X., Zhou, W. Bin, Xia, T.S., ... Wei, J.F. (2019). The RNA binding protein RBMS3 inhibits the metastasis of breast cancer by regulating Twist1 expression. *Journal of Experimental and Clinical Cancer Research* **38**(1):1-11. doi: <https://doi.org/10.1186/s13046-019-1111-5>.

Zimyanin, V., Lowe, N. and St Johnston, D. (2007). An Oskar-Dependent Positive Feedback Loop Maintains the Polarity of the *Drosophila* Oocyte. *Current Biology* **17**(4):353-359. doi: <https://doi.org/10.1016/j.cub.2006.12.044>.

Zubiaga, A.M., Belasco, J.G. and Greenberg, M.E. (1995). The Nonamer UUAUUUAUU Is the Key AU-Rich Sequence Motif That Mediates mRNA Degradation. *Molecular and Cellular Biology* **15**(4):2219-2230. doi: <https://doi.org/10.1128/mcb.15.4.2219>.

⌘ Appendices ⌘

Appendices:

Appendix 1: List of Primer Sequences Designed

List of oligos and primers sequence designed for PCR amplification, cloning, and sequencing of constructs. Underlined sequences (NNN) indicate endonuclease recognition site designed within primers for cloning purposes.

N°	Oligo(s)	Sequence (5' → 3')	Uses
Oligonucleotide to generate <i>shep</i> TRIP RNAi constructs for transgenic flies:			
1	Shep-TRiP_Top strand	CTAGCAGTCAAGTTGTTTCGACCAGA ATATAGTTATATTCAAGCATATATTC TGGTCGAAACAACCTGGCG	For cloning of Shep TRiP hairpin into pWALIU ^M -20 and -22
2	Shep-TRiP_Bottom strand	AATTCGCCAAGTTGTTTCGACCAGAA TATATGCTTGAATATAACTATATTCT GGTCGAAACAACCTGACTG	

Primers to generate pBlueScript-II constructs for <i>in vitro</i> Transcription:			
3	shg_5'UTR_fwd_XbaI	AAA <u>CTAGAT</u> GTTGCATTGTGTTTTT GCT	For cloning of shotgun (shg; DE-Cadherin) 5'UTR into pBS-II KS ⁺
4	shg_5'UTR_rev_XhoI	AA <u>ACTCGAGACT</u> CCTCTAGCGGCGATC	
5	shg_CDS_fwd_XbaI	AAA <u>CTAGA</u> AATGTCCACCAGTGTCC	For cloning of shg CDS into pBS-II KS ⁺
6	shg_CDS_rev_EcoRI	AA <u>GAAATTC</u> GATGCGCCAGCCCTG	
7	shg_3'UTR_fwd_XbaI	AA <u>CTAGAT</u> AGGAATCTTCGCCAGC	For cloning of shg 3'UTR into pBS-II KS ⁺
8	shg_3'UTR_rev_XhoI	TT <u>ACTCGAGAT</u> TCGATTTTATTATAT TTTTTATTTTAACACAATTG	
9	grk_5'UTR_fwd_XbaI	AA <u>CTAGA</u> AGTCTCTTTTCCGACGT GC	For cloning of gurken (grk) 5'UTR into pBS-II KS ⁺
10	grk_5'UTR_rev_EcoRI	AA <u>GAAATTC</u> TGAAAACGCTTGGG	
11	grk_CDS_fwd_XbaI	AA <u>CTAGA</u> AATGATGCAAATCCCATT TAC	For cloning of grk CDS into pBS-II KS ⁺
12	grk_CDS_rev_EcoRI	AA <u>GAAATTC</u> TAGCATCTGACAAAAA AGC	
13	grk_3'UTR_fwd_XbaI	AA <u>CTAGAGAT</u> TTAGAATTTGATTT GGAAAC	For cloning of grk 3'UTR into pBS-II KS ⁺
14	grk_3'UTR_rev_EcoRI	AA <u>GAAATTC</u> GAAATGTTTTAATTCTGT TTC	

Primers to generate mCh-tagged Shep constructs for transgenic flies:			
15	Fwd-Shep-RA-BglII	AAA <u>AGATCT</u> ATGCACCCACGATACAG	Paired with primer #17 for cloning of Shep-A CDS into pTIGER-mCh
16	Fwd-Shep-REC-BglII	AAA <u>AGATCT</u> ATGGGTCCCAACGGC	For cloning of Shep -C/-E CDS into pTIGER-mCh
17	Rev-Shep-XbaI	AA <u>CTAGACT</u> ATTTGGGAGCGGC	

Appendices

Primers to identify expression of <i>shep</i> transcripts using RT-PCR:			
18	RT-PCR_ShepA/F_fwd	CGTTGCGCTTTCTTCGTTTTTTTTTCGGC	For amplification of <i>shep</i> transcript A (Amplicon size 893 bp)
19	RT-PCR_Shep.A_rev	GGCACTCGCTGACCGTACGTGG	
20	RT-PCR_Shep.B_fwd	AATTAGTTGAACTTCGTTGCTCGACC CGACG	Paired with primer #28 for amplification of <i>shep</i> transcript B (Amplicon size 1,179 bp)
21	RT-PCR_Shep.Dn_fwd	GCAACATCGAAAGGAGCAACATCCGC	Paired with primer #28 for amplification of <i>shep</i> transcript D (Amplicon size 1,190 bp)
22	RT-PCR_ShepHIE_fwd	GGTTAGTTCGCAGTTCGCAGTTCCC	For amplification of <i>shep</i> transcript E (Amplicon size 1,048 bp)
23	RT-PCR_Shep.E_rev	CAAATTTGTGGGATCCTGTTCCCTGTTG	
24	RT-PCR_Shep.F_fwd	GTAAGTTTTCTAAGCCAACCTGTGGACTATTATTGG	Paired with primer #27 for amplification of <i>shep</i> transcript F (Amplicon size 2,196 bp)
25	RT-PCR_Shep.G_fwd	CTCAGTCAGCATCCGACATCGTTTCGAGC	Paired with primer #27 for amplification of <i>shep</i> transcript G (Amplicon size 2,541 bp)
26	RT-PCR_Shep.Hi_fwd	GGGTTACATGATGACTCAGGTAGATGATCAGTAC	For amplification of <i>shep</i> transcript H (Amplicon size 1,267 bp)
27	RT-PCR_ShepALL1_rev	ATCATCATCGAAATGGATTTTGTTCGTTTTTATTTGCTCTG	
28	RT-PCR_ShepALL2_rev	GTTGTTTTATCTAAAATAGCCTTGGT TGATATAATTGTTCCG	Paired with primer #22 for amplification of <i>shep</i> transcripts -H, -I, & -E (Amplicon size 880 bp)
29	RT-PCR_Shep.In_fwd	CAGGTAGATGATCAGACTTCGTATTCTCCACAG	2nd part of nested PCR - Paired with primer #30 for amplification of <i>shep</i> transcripts -I (Amplicon size 1,342 bp)
30	RT-PCR_Shep.In_rev	CGCTTTCTTCGTTCTTCTGGGGG	1st part of nested PCR - Paired with primer #22 for amplification of <i>shep</i> transcripts -H & -I (Amplicon size 2,800 bp)

Appendices

Primers used for colony PCR screenings and sequencing of pWalium constructs: *: Asterisks indicate primers used to sequence inserts in pWalium constructs.			
31	pWalium20_fwd*	ACCAGCAACCAAGTAAATCAAC	Paired with primer #33 to screen for insert in pWalium-20 vector
32	pWalium22_fwd*	GGTGATAGAGCCTGAACCAG	To screen of insert in pWalium-22 vector.
33	pWalium_rev	TAATCGTGTGTGATGCCTACC	

Primers to generate epitope tagged Shep constructs:			
34	Shep-RA_EcoRI_fwd	AAAGAATTCATGCACCCACGATACAG	Paired with primer #36 for cloning of Shep-A CDS into pSpot-ev1
35	Shep-RE/C_EcoRI_fwd	AAAGAATTCATGGGTCCCAACGGC	For cloning of Shep -C/-E CDS into pSpot-ev1
36	Shep_HindIII_rev	AAAAGCTTCTATTTGGGAGCGGC	
37	Su(var)2-10.CDS_EcoRI_fwd	AAAGAATTCATGCGAAAGACCCGCTCT	For cloning of Su(var)2-10 CDS into pSpot-ev1
38	Su(var)2-10.CDS_XhoI_rev	AAACTCGAGTTAGCCGACGTTTGGGC	
39	Shep-XhoI_rev	AAACTCGAGCTATTTGGGAGCGGC	For cloning of Shep CDS into pNEBExpress
40	Shep_noSTOP_XhoI_rev	AAACTCGAGTTTGGGAGCGGCTTG	For cloning of Shep CDS lacking a stop codon into pNEBExpress
41	TopOligo_N-termSpotTag_BamHI	GATCCATGCCGGATCGCGTGCGCGCATCTCTCACTGGAGCAGCG	For cloning of N-terminal Spot-tag CDS into pNEBExpress
42	BottomOligo_N-termSpotTag_BamHI	GATCCGCTGCTCCAGTGAGAGACTGCGCGCACGGATCCGGCATG	
43	TopOligo_C-termSpotTag_XhoI	TCGAGTCTGCCGGCACCCCGGATCGGTGCGCGCAGTCTCTCACTGGAGCAGCTAAC	For cloning of C-terminal Spot-tag CDS into pNEBExpress
44	BottomOligo_C-termSpotTag_XhoI	TCGAGTTAGCTGCTCCAGTGAGAGACTGCGCGCACGGATCCGGGGTCCCGGCAGAC	

Primers to sequence constructs listed in Appendix 3: Asterisks (*) indicate primers part of the 'Standard Primers' provided by Eurofins Genomics.			
45	T7*	TAATACGACTCACTATAGGG	Sequence 5' end of all pBSII-KS and pSpot-ev1 constructs
46	T3*	AATTAACCCTCACTAAAGGG	Sequence 3' end of all pBSII-KS constructs
47	pBSII_fwd	TGCGGGCCTCTTCGCTATTACG	Sequence pBSII constructs from T7 promoter into the MCS
48	M13uni-21*	TGTA AACGACGGCCAGT	Sequence 5' end of all pBSII-KS constructs

Appendices

49	M13rev-29*	CAGGAAACAGCTATGACC	Sequence 3' end of all pBSII-KS constructs
50	pTIGER_fwd	CCGGCAAGGGTCGAGTC	Sequence 5' end of all pTIGER constructs
51	pTIGER_down	TCAAAGGCAGAAATGTTTACTCTTGACC	Sequence 3' end of all pTIGER constructs
52	mCherry-Ct.2_fwd	CGTGGAACAGTACGAACGC	Sequence of 5' end of recombinant proteins N-terminally fused to mCh.
53	VAL-20_fwd	ACCAGCAACCAAGTAAATCAAC	Sequence 5' end of all Valium-20 constructs
54	VAL-22_fwd	GGTGATAGAGCCTGAACCAG	Sequence 5' end of all Valium-22 constructs
55	pSpot_Rev	TCCGCCAAAACAGCCAAGC	Sequence 3' end of all pSpot-eV1 constructs
56	NEBExpressVector_rev	AGCCAACCTCAGCTTCCTTTTCG	Sequence 3' end of all pNEBExpress constructs

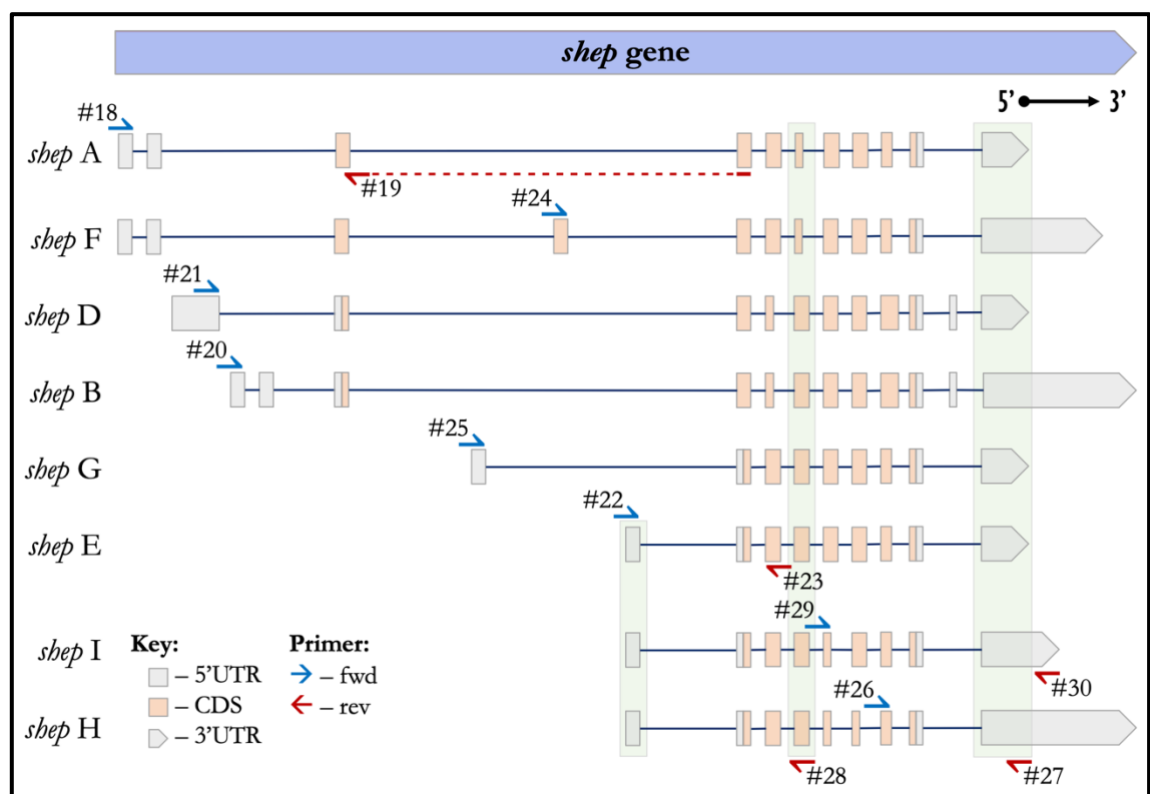
Primers used and generated for quantification of transcript expression level: *: Asterisks indicate genes used as reference/housekeeping gene. †: cross indicate primers generated for this project.			
57	H2A(CG31618)_fwd*	CCGTGCCGGTCTTCAATTCCCTG	Quantification of <i>his2A</i> mRNA levels
58	H2A(CG31618)_rev*	CGAGAACCTCAGCGGCCAGAT	
59	PolIII_fwd*	CCACCCGGCCACGTAAG	Quantification of RNA <i>polIII</i> mRNA levels
60	PolIII_rev*	AAGAGGGAGAAACACTCGGC	
61	Rp49/RpL32_fwd*	GCTAAGCTGTGCACAAATGGCG	Quantification of <i>rp49</i> mRNA levels
62	Rp49/RpL32_rev*	TCCGGTGGGCAGCATGTG	
63	Oskar_fwd	GCAACTATATATCCGTGCGCG	Quantification of <i>osk</i> mRNA levels
64	Oskar_rev	CCCGTCAGTTTTTCGATATTCAC	
65	Gurken_fwd	CCCGCGCTTGCTGCTC	Quantification of <i>grk</i> mRNA levels
66	Gurken_rev	CACACTTGCATCTCCTTGTGG	
67	Bicoid_fwd	GACCTGCGCCATCGCCGTT	Quantification of <i>bcd</i> mRNA levels
68	Bicoid_rev	ACCCTTCAAAGGCTCCAAGATCTGTAGC	
69	Shotgun_fwd†	GACGTTTGCACCTTCAACGTTACC	Quantification of <i>shg</i> mRNA levels
70	Shotgun_rev†	CCGCAGAATCTCGTATTCGACC	
71	Nanos_fwd†	CACCGCCAATTCGCTCCTTATG	Quantification of <i>nos</i> mRNA levels
72	Nanos_rev†	GCTGGTGA CTGCACTAGC	
73	Delta_fwd	GCAGGGCGATTACTGTCACATACC	Quantification of <i>Dl</i> mRNA levels
74	Delta_rev	GCATTGATTGGGCTTGTCGCAATG	
75	Notch_fwd	AACACCCACGGATCTTACCAGTGC	Quantification of <i>N</i> mRNA levels
76	Notch_rev	AGACCGTTTCGACCGGCAAATTC	

Primers to construct Shep-A and -E RRM mutants for the generation of transgenic flies:			
78	Shep.RRM1mut_fwd. n	AAAGGTGCCGGCGCCGTCGACGCCGAGCAGCCAGCCTTCG	To generate Shep-A RRM1 mutant construct

Appendices

79	Shep.RRM1mut_rev_n	CGAAGGCTGGCTGCTCGGCGTCGACG GCGCCGGCACCTTT	lacking RNA-binding capacity
80	Shep.RRM2_mut_fw_d	ATGAACTCCAAGGGCGCTGGCGCTGC CGCCATGGAGAGTCGCGAG	To generate Shep-E RRM2 mutant construct lacking RNA-binding capacity
81	Shep.RRM2_mut_rev_v	CTCGCGACTCTCCATGGCGGCAGCGC CAGCGCCCTTGGAGTTCAT	
82	Shep-E_mut_rev	AATGGCCGGCCTATGAAGCACCCAGA TACCCACTTTGGCCATCTGAGCTTGC ACGC	To generate Shep-E RRM mutant construct lacking RNA-binding capacity
83	Shep-E_mut_fwd	GTGGGTATCTGGGTGCTTCATAGGCC GGCCATTCAACAGGAACAGGATCCC	

Appendix 1.1 – Primers design for the identification of *shep* transcripts using qualitative RT-PCR.



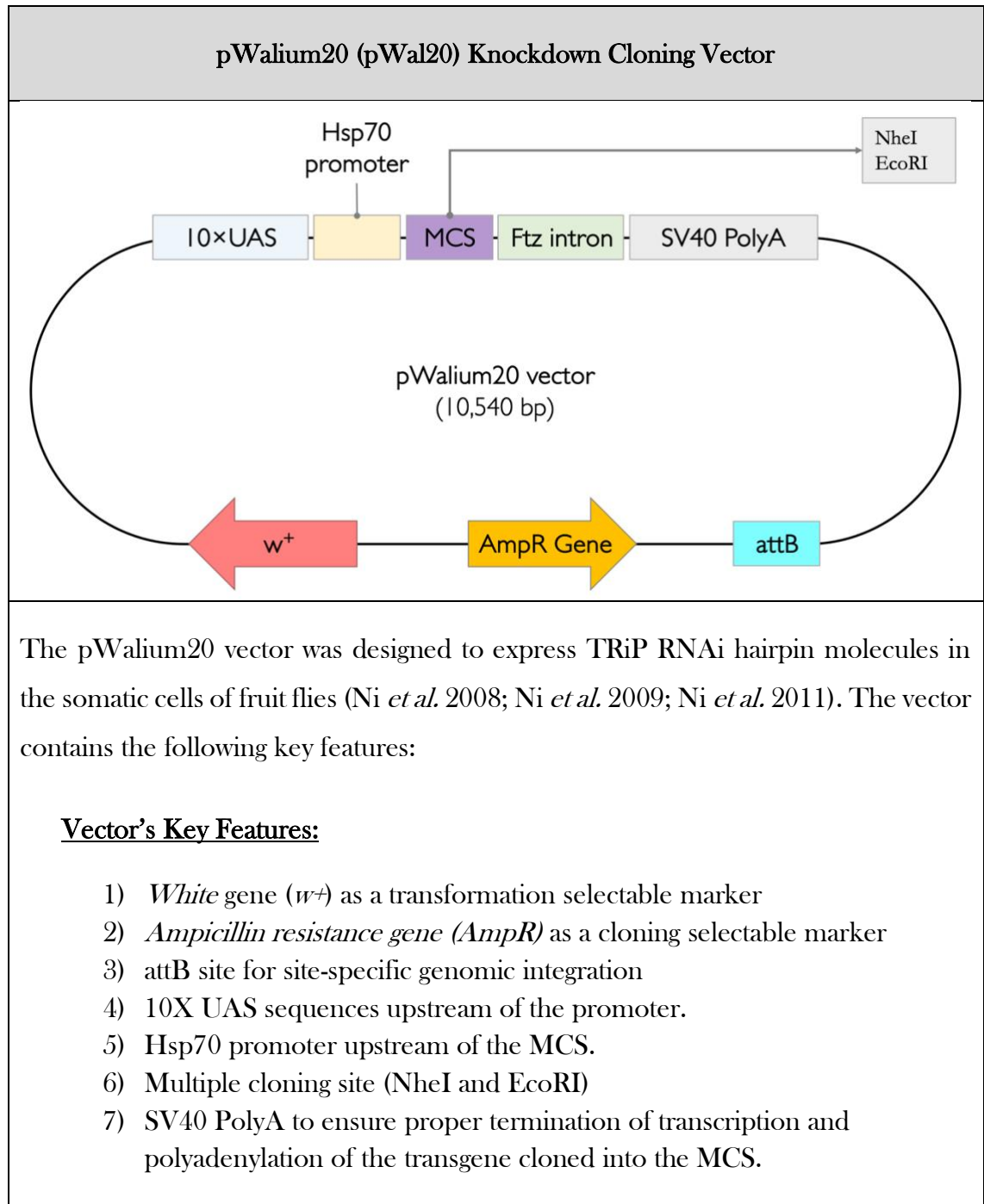
The above diagram presents a schematic representation of the *shep* gene model, as detailed in FlyBase. It depicts the genomic organization of eight *shep* transcripts (A, B, D, E, F, G, H, and I). Each exon is colour-coded to distinguish between untranslated regions (in gray) and coding sequences (in orange), with lines connecting them representing introns. Arrows, coloured blue for forward primers and red for reverse primers, indicate the approximate annealing sites used in identifying expressed *shep* transcripts. Additionally, regions highlighted in green denote sites where primers, capable of recognizing multiple or all transcripts, were used in various reactions for transcript identification. Note that the diagram is not to scale. For detailed information on the primers, refer to Appendix 1, using the corresponding reference numbers (e.g., see primer #18 on page 314).

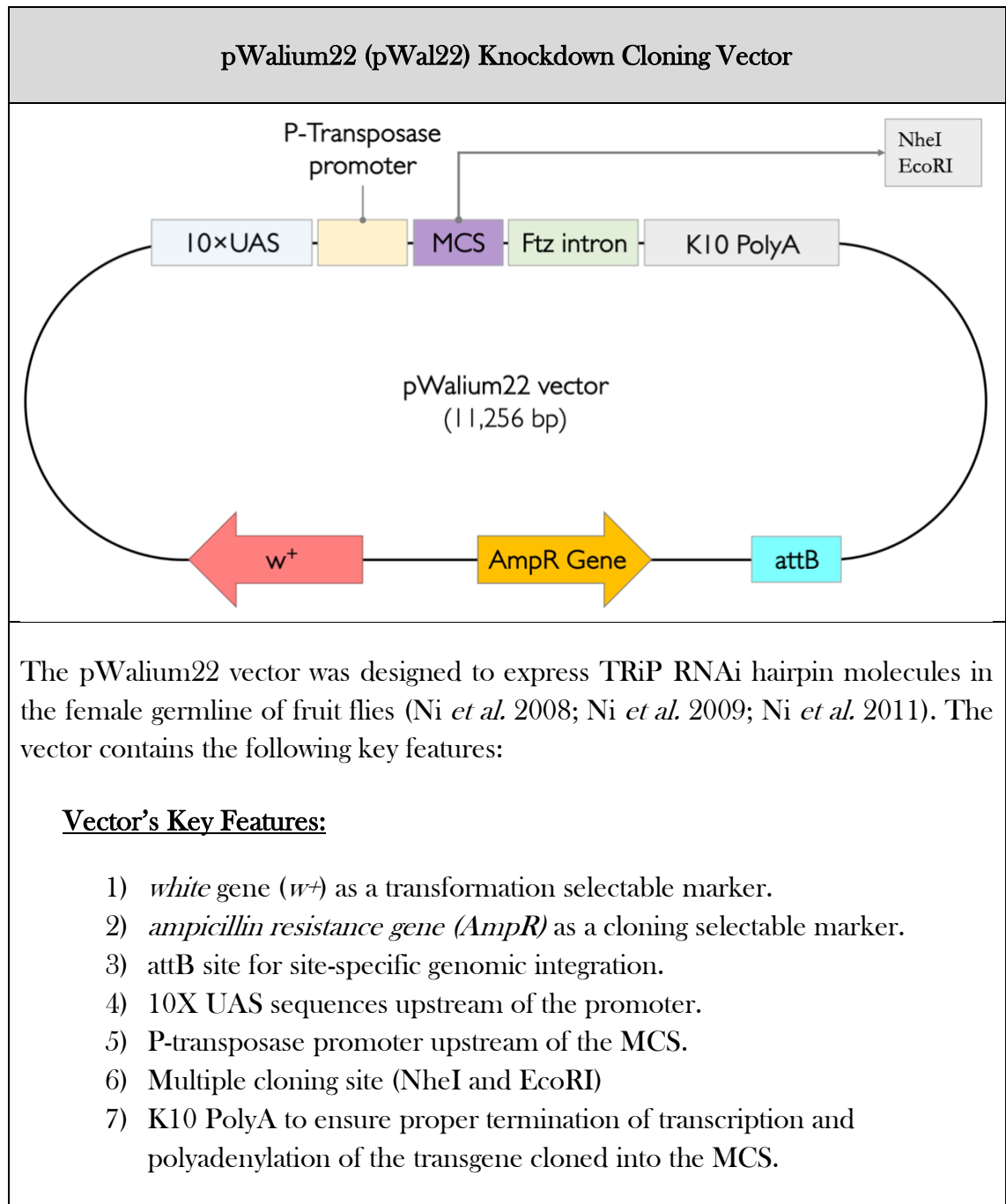
Appendix 2: List of FastDigest® Enzymes

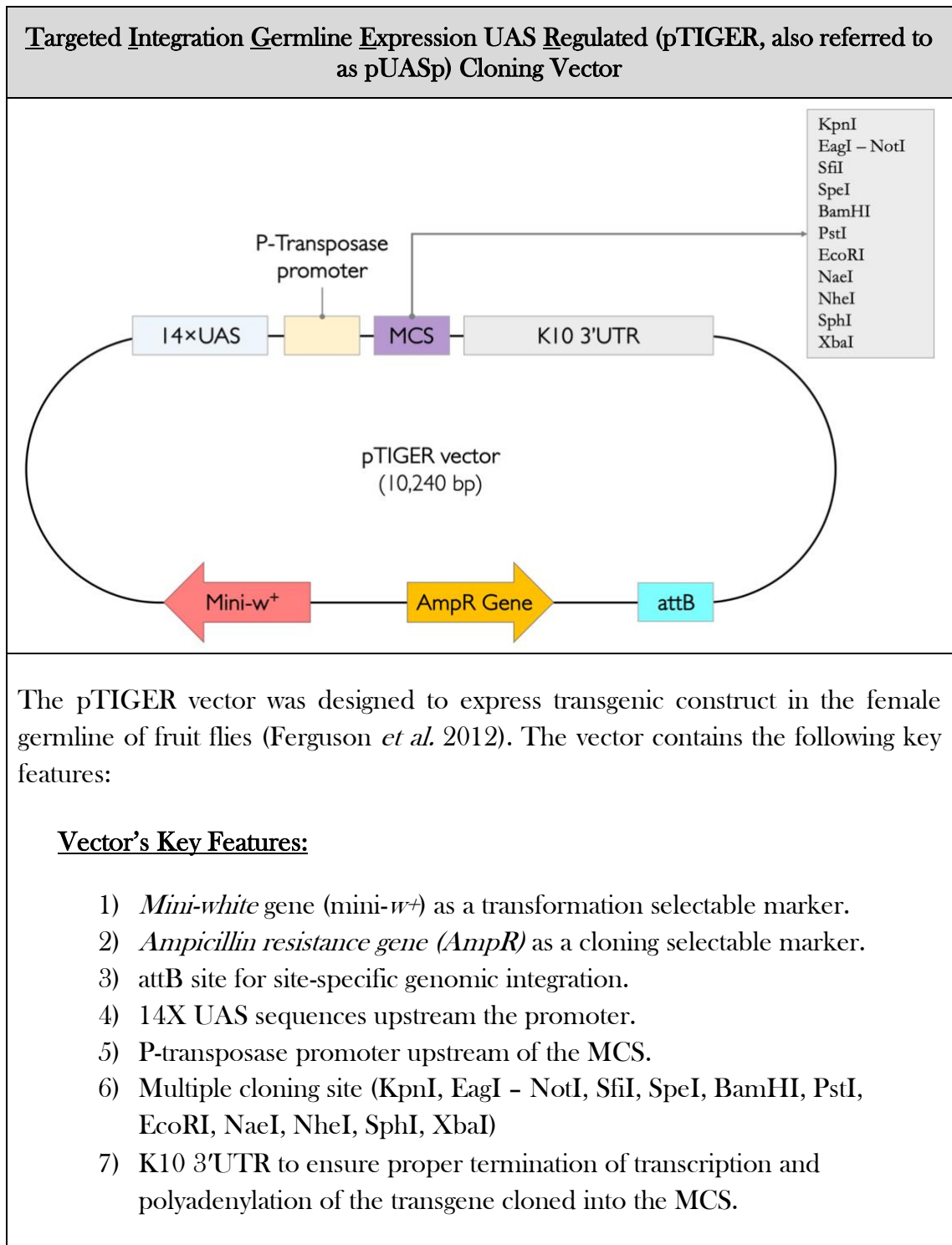
List of FastDigest® enzymes used in this project. *: Asterisks indicate restriction enzymes which exhibit star activity during short incubations of <1 hour at 37°C.

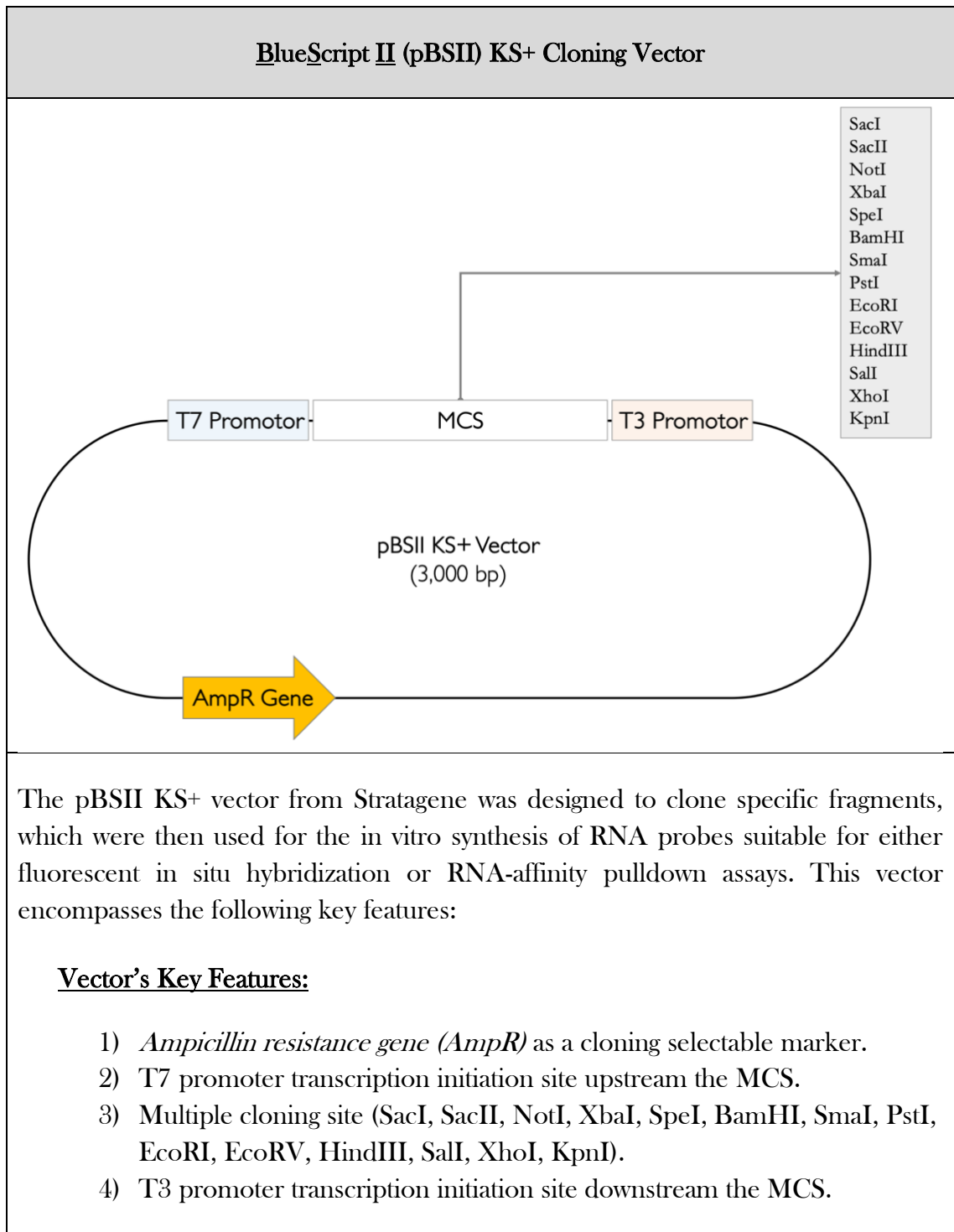
FastDigest Restriction Enzyme	Recognition Sequence	Overhang/End Produced	Cat. #
<i>Bam</i> HI*	5' G↓GATCC 3' 3' CCTAG↑G 5'	5'	FD0054
<i>Bgl</i> II	5' A↓GATCT 3' 3' TCTAG↑A 5'	5'	FD0084
<i>Eco</i> RII*	5' G↓AATTC 3' 3' CTAA↑G 5'	5'	FD0274
<i>Hind</i> III	5' A↓AGCTT 3' 3' TTCGA↑A 5'	5'	FD0504
<i>Kpn</i> I	5' GGTAC↓C 3' 3' C↑CATGG 5'	3'	FD0524
<i>Nae</i> I (also known as PdiI)	5' GCC↓GGC 3' 3' CGG↑CCG 5'	Blunt	FD1524
<i>Nhe</i> I	5' G↓CTAGC 3' 3' CGATC↑G 5'	5'	FD0973
<i>Not</i> I	5' GC↓GGCCGC 3' 3' CGCCGG↑CG 5'	5'	FD0593
<i>Sma</i> I	5' CCC↓GGG 3' 3' GGG↑CCC 5'	Blunt	FD0663
<i>Spe</i> I	5' A↓CTAGT 3' 3' TGATC↑A 5'	5'	FD1253
<i>Xba</i> I	5' T↓CTAGA 3' 3' AGATC↑T 5'	5'	FD0684
<i>Xho</i> I	5' C↓TCGAG 3' 3' GAGCT↑C 5'	5'	FD0695

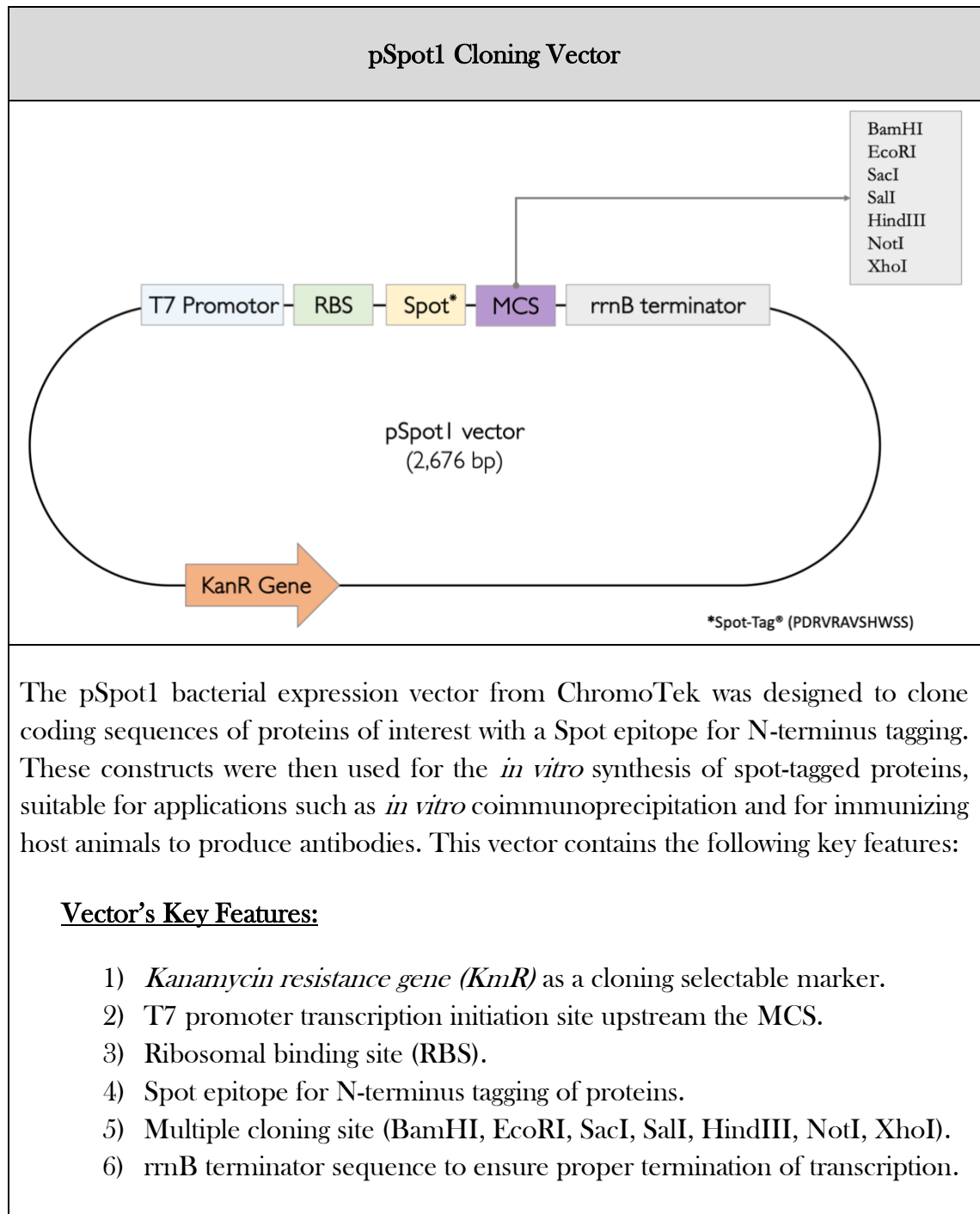
Appendix 3: Maps of Empty Vectors Used During Cloning of Constructs Listed in Appendix 4

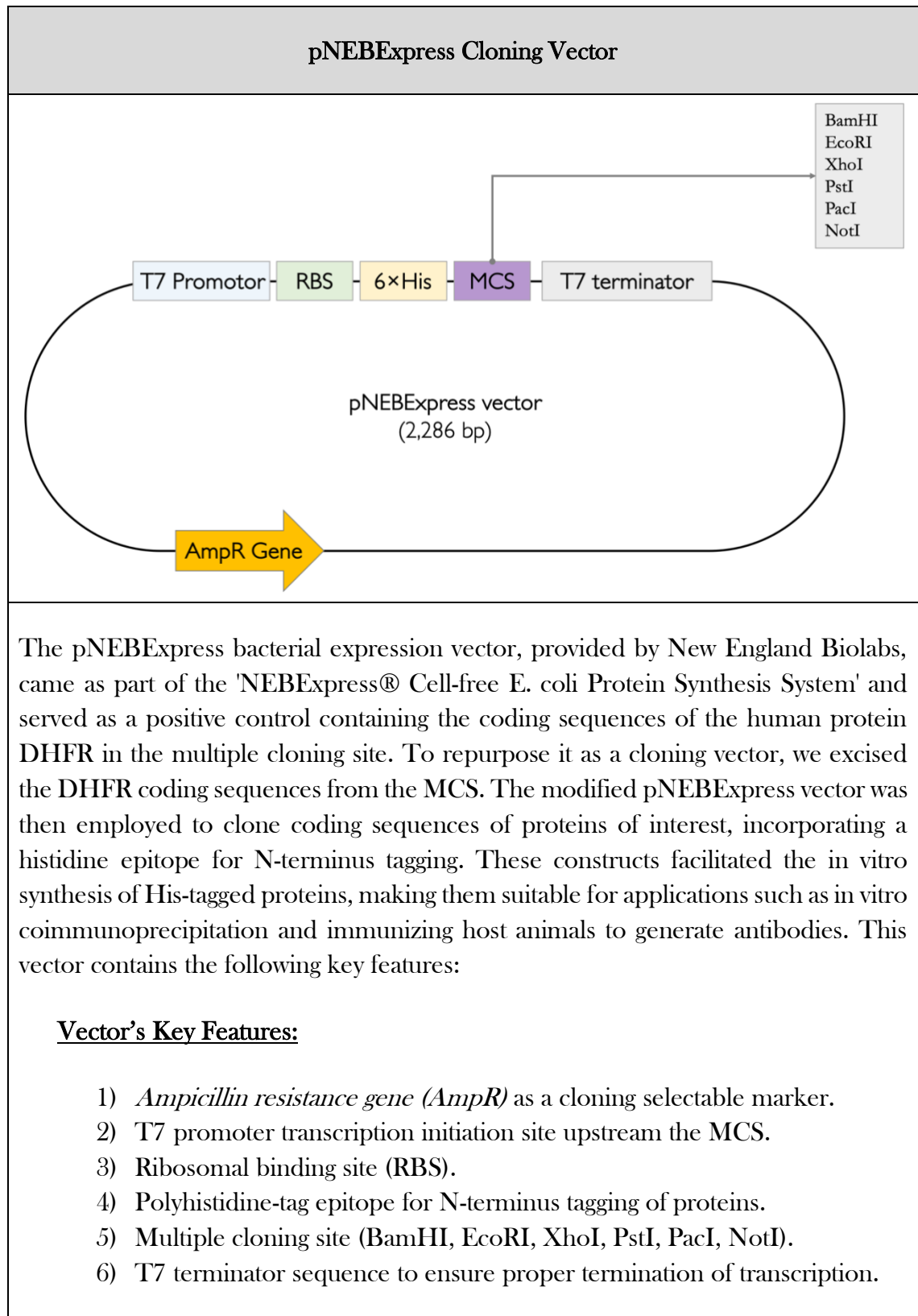












Appendix 4: Recombinant Constructs Generated

Appendix 4.1 – Constructs generated for the synthesis of RNA probes.

List of pBlueScript II constructs for *in vitro* transcription of biotinylated sense RNA probes via T7 promoter for Pull down assay. Asterisks (*) indicate constructs generated in Dr. Sonia Lopez De Quinto's Lab where this PhD project was conducted.

pBlueScript II Clone	pBSII Variant	Linearized by*	Fragment length (nt.)	Reference
y14 (tsunagi) CDS	-	BamHI	550	Besse et al. 2009
oskar M1M2 (i.e. 5'UTR)	KS (+)	HindIII	430	
Short oskar (i.e. CDS)	SK (+)	BamHI	1,400	
oskar 3'UTR	KS (+)	HindIII	1,100	
delta 5'UTR	SK (+)	EcoRI	680	SLQ* Lab
delta CDS	SK (+)	XbaI	2,500	
delta 3'UTR	SK (+)	Scal	2,100	
notch 5'UTR	SK (+)	XbaI	800	
notch CDS-1/2	SK (+)	XbaI	4,000	
notch CDS-2/2	SK (+)	XbaI	4,000	
notch 3'UTR	KS (+)	XhoI	1,240	
Shotgun (DE-Cadherin) 5'UTR	KS (+)	XhoI	725	This PhD Project
shotgun CDS	KS (+)	EcoRI	4,525	
shotgun 3'UTR	KS (+)	XhoI	1,150	
gurken 5'UTR	KS (+)	EcoRI	374	
gurken CDS	KS (+)	EcoRI	888	
gurken 3'UTR	KS (+)	EcoRI	456	SLQ Lab
nanos 3'UTR	KS (+)	HindIII	880	
bicoid 3'UTR	KS (+)	XhoI	910	
K10 3'UTR	KS (+)	HindIII	1,490	

Appendix 4.2 – Constructs generated for protein expression

List of recombinant DNA constructs generated in this PhD project for fusion protein expression.

Plasmid DNA Construct	Vector	Transgene Expression	Tagging	Application
pSpot-ShepA	pSpot1	Shep A	N-terminus	Protein expression in a Bacterial or cell-free system
pSpot-ShepC	pSpot1	Shep C	N-terminus	
pSpot-ShepE	pSpot1	Shep E	N-terminus	
pSpot-PTB	pSpot1	PTB	N-terminus	
pSpot-Bruno	pSpot1	Bruno	N-terminus	
pSpot-Su(var)2-10	pSpot1	Su(var2-10)	N-terminus	
pET15b-ShepE	pET15b	Shep E	N-terminus	
pNEBExpress-His:ShepE	pNEBExpress	Shep E	N-terminus	
pNEBExpress-His:Spot:ShepE	pNEBExpress	Shep E	N-terminus	
pNEBExpress-His:ShepE:Spot	pNEBExpress	Shep E	C-terminus	

Appendices

Appendix 4.3 – Constructs generated for fly transgenesis




List of DNA constructs designed and generated in this PhD project for fly transgenesis. †: Chromosomal location of attP landing sites: attP40 (on 2nd chromosome) and attP2 (on 3rd chromosome).

Plasmid DNA Construct	Vector	Transgene Expression	PhiC31 attP Landing site †	Application	
VAL22-Shep	pWALIU22	dsRNA	attP2	RNA interference (RNAi) Knockdown of Shep	
VAL20-Shep	pWALIU20	dsRNA	attP40		
VAL20-Shep	pWALIU20	dsRNA	attP2		
UASp-attB-mCh:Shep-A	pTIGER	Fusion Protein	attP40	Fluorescent labelling of Shep, overexpression, and ectopic expression.	
UASp-attB-mCh:Shep-C	pTIGER	Fusion Protein	attP40		
UASp-attB-mCh:Shep-E	pTIGER	Fusion Protein	attP40		
UASp-attB-mCh:Shep-E	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam	pTIGER	Fusion Protein	attP2	DNA adenine methyltransferase Identification of binding sites in the genome	
UASp-attB-LT3-mCh-Dam:PolII	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam:Brm	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam:PC	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam:HP1a	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam:H1	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam:PTB	pTIGER	Fusion Protein	attP2		
UASp-attB-LT3-mCh-Dam:Su(var)2-10	pTIGER	Fusion Protein	attP2		
UASp-attB-mCh:Shep-A ^{RRM}	pTIGER	Fusion Protein	attP40		Investigate the effects of Shep with impaired RNA-binding capacity
UASp-attB-mCh:Shep-C ^{RRM}	pTIGER	Fusion Protein	attP40		
UASp-attB-mCh:Shep-E ^{RRM}	pTIGER	Fusion Protein	attP40		
UASp-attB-mCherry	pTIGER	Fluorescent Protein	attP40	Control for gain-of-function experiments	

Appendix 5: Computational Scoring of Predicted Shep Binding Sites

In a comprehensive analysis using RNAcompete, Ray et al. (2013) reported RNA-binding motifs for 205 distinct genes from 24 diverse eukaryotes. Their study also revealed a positive correlation between the motifs identified *in vitro* and actual *in vivo* RNA-binding data. Among the RBPs investigated was the *Drosophila melanogaster* Shep protein, for which three distinct RNA-binding motifs were identified.

Summary of *Drosophila* Shep Binding Motif preferences based on RNAcompete data obtained from Ray et al. (2013). * Refer to Appendix 5.6 for IUPAC nucleotide code.

Gene	Species	ID	IUPAC*	Binding Motif
<u>SHEP</u>	<i>Drosophila melanogaster</i>	RNCMPT00174	WAUWUWD	
		RNCMPT00175	WUAUWWA	
		RNCMPT00068	AUAUWD	

As part of the sequence analysis conducted for this project, we quantified the number of Shep RNA-binding motifs present within each mRNA using the SnapGene genomic viewer tool. Refer to the subsequent sections for scoring details of the three distinct motifs (see Sections 5.1–5.3).

Appendices

Appendix 5.1 – Shep RNA-Binding Predictions Based on Motif 1.

Motif #1 – WAUWUWD (see Appendix 5.6 for IUPAC nucleotide code)

#	mRNA	ID	N° of predicted binding sites per region		
			5'UTR	CDS	3'UTR
1	<i>oskar</i>	FBgn0003015	0	1	6
2	<i>delta</i>	FBgn0003015	0	0	33
3	<i>notch</i>	FBgn0004647	1	3	22
4	<i>shotgun (DE-Cad)</i>	FBgn0003391	1	1	29
5	<i>tsunagi (Y14)</i>	FBgn0033378	0	0	1
6	<i>gurken</i>	FBgn0001137	3	0	6
7	<i>nanos</i>	FBgn0002962	0	0	39
8	<i>bicoid</i>	FBgn0000166	0	0	5
9	<i>fs(1) K10</i>	FBgn0000810	2	0	20

Appendix 5.2 – Shep RNA-Binding Predictions Based on Motif 2.

Motif #2 – WUAUWWA (Refer to Appendix 5.6 for IUPAC nucleotide code)

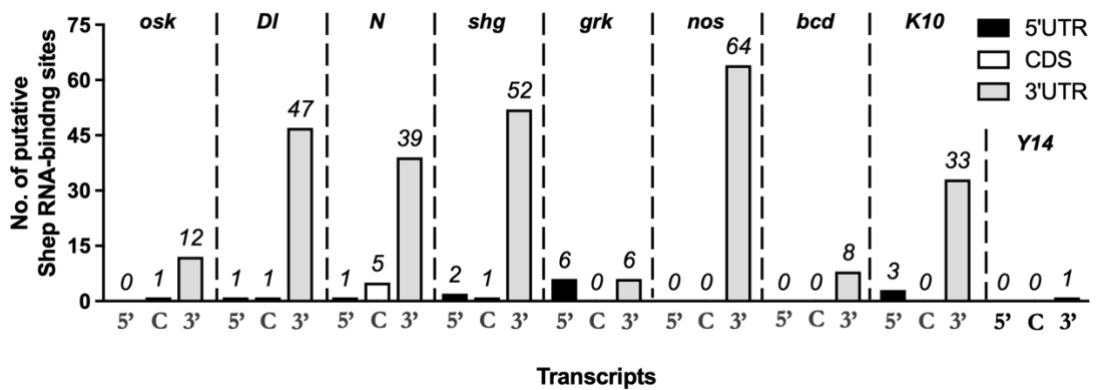
#	mRNA	ID	N° of predicted binding sites per region		
			5'UTR	CDS	3'UTR
1	<i>oskar</i>	FBgn0003015	0	0	5
2	<i>delta</i>	FBgn0003015	0	1	7
3	<i>notch</i>	FBgn0004647	0	1	15
4	<i>shotgun (DE-Cad)</i>	FBgn0003391	1	0	18
5	<i>tsunagi (Y14)</i>	FBgn0033378	0	0	0
6	<i>gurken</i>	FBgn0001137	2	0	0
7	<i>nanos</i>	FBgn0002962	0	0	24
8	<i>bicoid</i>	FBgn0000166	0	0	1
9	<i>fs(1) K10</i>	FBgn0000810	0	0	11

Appendix 5.3 – Shep RNA-Binding Predictions Based on Motif 3.

Motif #3 – AUAUUWD (see Appendix 5.6 for IUPAC nucleotide code)

#	mRNA	ID	N ^o of predicted binding sites per region		
			5'UTR	CDS	3'UTR
1	<i>oskar</i>	FBgn0003015	0	0	1
2	<i>delta</i>	FBgn0003015	1	0	7
3	<i>notch</i>	FBgn0004647	0	1	2
4	<i>shotgun (DE-Cad)</i>	FBgn0003391	0	0	5
5	<i>tsunagi (Y14)</i>	FBgn0033378	0	0	0
6	<i>gurken</i>	FBgn0001137	1	0	0
7	<i>nanos</i>	FBgn0002962	0	0	1
8	<i>bicoid</i>	FBgn0000166	0	0	2
9	<i>fs(1) K10</i>	FBgn0000810	1	0	2

Appendix 5.4 – Shep Binding Putative Sites Plotted Based on All Three RNAcompete Motifs (Raw Data).



The graph shows the combined number of putative Shep RNA-binding sites identified in each transcript based on the three motifs. For each transcript, the mRNA is split into three distinct regions: the 5' untranslated region (5'UTR, 5'), coding sequence (CDS, C), and 3' untranslated region (3'UTR, 3'). For the exact values of the binding sites for each motif, refer to Appendix 5. Abbreviations – *osk*: *oskar*, *Dl*: *Delta*, *N*: *Notch*, *shg*: *shotgun*, *grk*: *gurken*, *nos*: *nanos*, *bcd*: *bicoid*, *K10*: *female sterile (1) K10*, *Y14*: *tsunagi*.

Appendix 5.5 – Lengths of mRNA Candidates Investigated in Motif Analysis.

The length values listed below were used to normalize the number of putative Shep-binding sites identified in each mRNA region, as detailed in Appendix 5.1 – 5.4, On a per kilobase (Kb) basis. The normalized data is presented in **Figure 5.1** of Chapter 5.

#	mRNA	ID	Transcript	Length in base pairs (bp)		
				5'UTR	CDS	3'UTR
1	<i>oskar</i>	FBgn0003015	C	434	1407	1028
2	<i>delta</i>	FBgn0003015	A	679	2502	2097
3	<i>notch</i>	FBgn0004647	A	798	8112	1240
4	<i>shotgun (DE-Cad)</i>	FBgn0003391	A	724	4524	1152
5	<i>tsunagi (Y14)</i>	FBgn0033378	A	97	498	96
6	<i>gurken</i>	FBgn0001137	A	374	888	456
7	<i>nanos</i>	FBgn0002962	A	263	1206	880
8	<i>bicoid</i>	FBgn0000166	D	195	1470	834
9	<i>fs(1) K10</i>	FBgn0000810	A	171	1392	1489

Appendix 5.6 – The Nucleotide Code: IUPAC Nomenclature for Nucleic Acids

List of nucleic acids IUPAC nomenclatures. The IUPAC nucleotide code, which represents DNA sequence degeneracy, is adapted from Aguirre-Hernández *et al.* (2007) and Johnson (2010).* mnemonic to serve as a useful aide-mémoire.

Nucleotide Symbol or Code	Meaning	Description of Designation origin*
G	Guanine	<u>G</u> uanine
A	Adenine	<u>A</u> denine
T	Thymine	<u>T</u> hymine
C	Cytosine	<u>C</u> ytosine
R	G or A	pu <u>R</u> ine
Y	T or C	p <u>Y</u> rimidine
M	A or C	a <u>M</u> ino group on base
K	G or T	<u>K</u> etone group on base
S	G or C	<u>S</u> trong interaction (3 H bonds)
W	A or T	<u>W</u> eak interaction (2 H bonds)
H	A or C or T	not-G, H follows G in the alphabet
B	G or T or C	not- A, B follows A
V	G or C or A	not-T (not-U), V follows U
D	G or A or T	not-C, D follows C
N	G or A or T or C	a <u>N</u> y
. or –	None	Gap

Appendices

Appendix 6: List of Fly Stocks

List of *Drosophila melanogaster* stocks used in this project. †: Overtime stocks might become homozygous for allele(s) of interest. Therefore, balancers (incl. markers) in brackets indicate their potential segregation within a stock, but if any, at a lower frequency. ††: Abbreviations refer to *Drosophila* Stock Centres from which stocks were purchased. BL – Bloomington *Drosophila* Stock Center. SLQ indicate stocks generated in the lab. * Refer to Appendix 6.1 for list of genes affected by genomic deletion in the deficiency stocks.

Fly Line	Genotype [†]	Ref. ^{††}
Wildtype Control Stocks		
<i>white</i> (<i>w¹¹⁸</i>)	<i>w¹¹⁸</i> ;;	Lab stock
eGFP control	<i>w</i> ; <i>osk</i> -Gal4:VP16/CyO; UASp-eGFP/TM3,Sb	Combined
Protein Trap Stocks		
PTB	<i>w¹¹⁸</i> ;;PTB:GFP ^{dsRed} /TM6B,Tb,Hu	(Besse <i>et al.</i> 2009)
Yps	<i>yw</i> ;; Yps:GFP/TM3,Sb	CA06791 (Buszczak <i>et al.</i> 2007)
Shep #1	<i>w</i> ;; Shep:GFP ^{CC00236} /(TM3,Sb,Ser)	BL# 51578
Shep #2	<i>w</i> ;; Shep:GFP ^{MI06349} /TM3,Sb,Ser	BL# 51207
Shep #3	<i>w</i> ;; Shep:GFP ^{MI03592} /TM6C,Sb,Tb	BL# 61769
Germline Gal4 Drivers Stocks		
Double <i>oskar</i> driver	<i>w</i> ; <i>osk</i> -Gal4:VP16/CyO; <i>osk</i> -Gal4:VP16/TM3,Sb	Combined
Single <i>oskar</i> driver (2 nd Chr.)	<i>w¹¹⁸</i> ; <i>osk</i> -Gal4:VP16/CyO;	BL# 44241
Single <i>oskar</i> driver (3 rd Chr.)	<i>w¹¹⁸</i> ;; <i>osk</i> -Gal4:VP16/TM3,Sb	BL# 44242
Maternal Triple Driver (MTD)	<i>otu</i> -Gal4:VP16; <i>nos</i> -Gal4 ; <i>nos</i> -Gal4:VP16	BL# 31777
<i>nanos</i> driver (X Chr.)	<i>w¹¹⁸</i> , <i>nos</i> -Gal4; ;	BL# 7303
<i>nanos</i> driver (3 rd Chr.)	<i>w¹¹⁸</i> ; ; <i>nos</i> -Gal4:VP16	BL# 4937
<i>oskar</i> driver & Shep-PT	<i>w</i> ; <i>osk</i> -Gal4:VP16/CyO; Shep:GFP ^{CC00236}	Combined
Trans-heterozygous <i>osk-nos</i> driver	<i>w</i> ; <i>osk</i> -Gal4:VP16/CyO; <i>nos</i> -Gal4:VP16/TM3,Sb	Combined
Trans-heterozygous <i>osk-bam</i> driver	<i>w</i> ; <i>osk</i> -Gal4:VP16/CyO; <i>bam</i> -Gal4:VP16/TM3,Sb	Combined
Trans-heterozygous <i>osk</i> driver & Deficiency 1	<i>w</i> ; <i>osk</i> -Gal4:VP16/CyO; Df(3L) ^{ED210} /TM3,Sb	Combined

Appendices

Trans-heterozygous <i>nos</i> driver & Shep-GFP	<i>nos</i> -Gal4; ; Shep:GFP ^{CC00236}	Combined
Trans-heterozygous <i>nos-osk</i> driver	<i>nos</i> -Gal4; ; <i>osk</i> -Gal4:VP16/TM3,Sb	Combined
Soma Gal4 Drivers Stocks		
<i>tub</i> > Gal80	<i>w</i> ; ; <i>tub</i> -Gal80ts/TM3,Sb	BL# 7018
<i>traffic jam (tj)</i> driver	<i>w</i> ; <i>tj</i> -Gal4/CyO;	Gift from Julius Brennecke, IMBA, Vienna (Austria) (Flybase ID FBtp0089190)
<i>tj</i> & <i>tub</i>	<i>w</i> ; <i>tj</i> -Gal4/CyO; <i>tub</i> -Gal80ts/TM3,Sb	Combined
<i>slbo</i> driver	<i>w</i> ; <i>slbo</i> -Gal4/CyO;	Rørth <i>et al.</i> (1998)
<i>slbo</i> reporter line	<i>w</i> ; ; <i>slbo</i> -LifeAct:GFP/TM3,Sb	BL# 58364
Trans-heterozygous <i>slbo</i> driver & reporter	<i>w</i> ; <i>slbo</i> -Gal4/CyO; <i>slbo</i> -LifeAct:GFP/TM3,Sb	Combined
UASp Stocks (for Germline Expression)		
Untagged eGFP	<i>w</i> ; ; UASp-eGFP/TM3,Sb	SLQ unpublished
mCherry-tagged Shep-A	<i>w</i> ; UASp-mCh:Shep-A/CyO;	Generated for this project
mCherry-tagged Shep-C	<i>w</i> ; UASp-mCh:Shep-C/CyO;	Generated for this project
mCherry-tagged Shep-E (2 nd Chr.)	<i>w</i> ; UASp-mCh:Shep-E/CyO;	Generated for this project
mCherry-tagged Shep-E (3 rd Chr.)	<i>w</i> ; ; UASp-mCh:Shep-E/TM3,Sb	Generated for this project
mCherry-tagged Shep-A RRM mutant	<i>w</i> ; UASp-mCh:Shep-A ^{RRM} /CyO;	Generated for this project
mCherry-tagged Shep-A RRM mutant	<i>w</i> ; UASp-mCh:Shep-C ^{RRM} /CyO;	Generated for this project
mCherry-tagged Shep-A RRM mutant	<i>w</i> ; UASp-mCh:Shep-E ^{RRM} /CyO;	Generated for this project
RNAi Stocks		
Valium22- <i>shep</i>	<i>w</i> ; Valium22- <i>shep</i> ^{TRIP.GL00659} /CyO;	BL# 38218
Walium22- <i>shep</i>	<i>w</i> ; ; Valium22- <i>shep</i> ^{TRIP.GL00659} /TM3,Sb	Generated for this project

Appendices

Double V/Walium22- <i>shep</i>	<i>w</i> ; Valium22- <i>shep</i> ^{TRIP.GL00659} /CyO; Valium22- <i>shep</i> ^{TRIP.GL00659} /TM3,Sb	Combined
Double Walium22- <i>ptb</i>	<i>w</i> ; Walium22- <i>ptb</i> /CyO; Walium22- <i>ptb</i> /TM3,Sb	SLQ unpublished
Trans-heterozygous V/Walium22-Shep- <i>ptb</i>	<i>w</i> ; Valium22- <i>shep</i> ^{TRIP.GL00659} /CyO; Walium22- <i>ptb</i> /TM3,Sb	Combined
Valium22- <i>eGFP</i>	<i>w</i> ⁺ ; Valium22- <i>eGFP</i> ^{shRNA.1} /CyO;	BL# 41557
Valium22- <i>eGFP</i>	<i>w</i> ⁺ ; Valium22- <i>eGFP</i> ^{shRNA.1} /TM3,Sb	BL# 41558
Double Valium22- <i>eGFP</i>	<i>w</i> ; Valium22- <i>eGFP</i> ^{shRNA.1} /CyO; Valium22- <i>eGFP</i> ^{shRNA.1} /TM3,Sb	Combined
Valium22- <i>khc</i>	<i>w</i> ⁺ ; Valium22- <i>khc</i> ^{TRIP.GL00330} /TM3,Sb	BL# 35409
Valium22- <i>Notch</i>	<i>w</i> ⁺ ; Valium22- <i>N</i> ^{TRIP.GL00092} /TM3,Sb	BL# 35213
Valium22- <i>Delta</i>	<i>w</i> ⁺ ; Valium22- <i>DI</i> ^{FL00520} /CyO;	BL# 36784
Valium20- <i>shep</i>	<i>w</i> ; Valium20- <i>shep</i> ^{TRIP.HMS00959} /TM3,Sb	BL# 33996
Walium20- <i>shep</i>	<i>w</i> ; Walium20- <i>shep</i> ^{TRIP.GL00659} /CyO;	Generated for this project
Walium20- <i>shep</i>	<i>w</i> ; Walium20- <i>shep</i> ^{TRIP.GL00659} /TM3,Sb	Generated for this project
Double V/Walium20- <i>shep</i>	<i>w</i> ; Walium20- <i>shep</i> ^{TRIP.GL00659} /CyO; Valium20- <i>shep</i> ^{TRIP.HMS00959} /TM3,Sb	Combined
Valium20- <i>eGFP</i>	<i>w</i> ⁺ ; Valium20- <i>eGFP</i> ^{shRNA.1} /CyO;	BL# 41555
Valium20- <i>eGFP</i>	<i>w</i> ⁺ ; Valium20- <i>eGFP</i> ^{shRNA.1} /TM3,Sb	BL# 41556
Double Valium20- <i>eGFP</i>	<i>w</i> ; Valium20- <i>eGFP</i> ^{shRNA.1} /CyO; Valium20- <i>eGFP</i> ^{shRNA.1} /TM3,Sb	Combined
Valium20- <i>Notch</i>	<i>w</i> ; Valium20- <i>N</i> ^{TRIP.HMS00001} /TM3,Sb	BL# 33611
Valium20- <i>N</i> and Gal80ts	<i>w</i> ; <i>tub</i> -Gal80ts; Valium20- <i>N</i> ^{TRIP.HMS00001} /TM3,Sb	Combined
Valium20- <i>DI</i>	<i>w</i> ; Valium20- <i>DI</i> ^{TRIP.HMS01309} /TM3,Sb	BL# 34322
Valium20- <i>DI</i> and Gal80ts	<i>w</i> ; <i>tub</i> -Gal80ts; Valium20- <i>DI</i> ^{TRIP.HMS01309} /TM3,Sb	Combined
Large Genomic Deletion Covering Shep Locus *		
<i>Deficiency 1</i> *	<i>w</i> ; Df(3L) ^{ED210} /TM3,Sb	BL# 8061
<i>Deficiency 2</i> *	<i>w</i> ; Df(3L) ^{Exc6103} /TM3,Sb	BL# 7582
<i>Deficiency 3</i> *	<i>w</i> ; Df(3L) ^{Exc6104} /TM3,Sb	BL# 7583
Mutants - Transgenic Insertion Stocks		
<i>shep</i> P-element insertion 1	<i>w</i> ¹¹⁸ ; <i>shep</i> ^{KG10149}	BL# 15231
<i>shep</i> P-element insertion 2	<i>w</i> ¹¹⁸ ; <i>shep</i> ^{BG00836}	BL# 12513
<i>shep</i> P-element insertion 3	<i>w</i> ¹¹⁸ ; <i>shep</i> ^{EY04794}	BL# 16366

Appendices

<i>shep Gal4</i> <i>insertion 1</i>	$w^{118};; shep^{C522}$	BL# 3747
<i>shep Gal4</i> <i>insertion 2</i>	$w^{118};; shep^{BG02468}$	BL# 12693

Appendix 6.1 – Genetic Background of Deficiency Fly Stocks

List of genes affected by genomic deletion in the deficiency stocks. * Information regarding chromosomal breakpoints and affected genes is based on the current annotation and iteration of FlyBase (Gramates *et al.* 2022) and the Bloomington Drosophila Stock Center (Abbreviated below as BL).

Stock (Ref)	Genotype & Chromosomal Breakpoints *	Genes Disrupted*
Deficiency 1 (BL# 8061)	$w^-;; Df(3L)^{ED210}/TM3,Sb$ (64B9-64C13, 3L:4544234-5355342)	73 genes: AlaRS-m , LeuRS-m , axo , CG4597 , Yod1 , CG4611 , Gdap1 , CG4669 , CG7509 , CG7514 , Fitm , Dhrs4 , Tcs5 , CG10674 , CG11357 , ATPsynCF6L , CG12493 , hoka , CG13705 , CG13707 , CG13708 , CG13711 , CG13712 , CG13713 , CG13716 , CG15212 , CG15213 , CG15876 , CG17030 , Dnah3 , CG18418 , CG18808 , t , Grip128 , CG32235 , CG32236 , CG32237 , TTL11B , CG32246 , CG33777 , CG34266 , CG42540 , CHMP2B , Con , lncRNA:CR34047 , lncRNA:CR43627 , lncRNA:CR43879 , lncRNA:CR43881 , lncRNA:CR43882 , lncRNA:CR43884 , Dhc64C , RhoGEF64C , Gen , lama , Rh50 , HDAC1 , shep , Src64B , Srp54k , tRNA:Ser-AGA-2-5 , lncRNA:CR44528 , asRNA:CR45160 , lncRNA:CR45438 , lncRNA:CR45685 , lncRNA:CR45740 , lncRNA:CR45741 , lncRNA:CR45742 , lncRNA:CR45821 , asRNA:CR45830 , lncRNA:CR45983 , lncRNA:CR45984 , lncRNA:CR45985 , lncRNA:CR46198 , CG46456
Deficiency 2 (BL# 7582)	$w^-;; Df(3L)^{Exel6103}/TM3,Sb$ (64C4-64C8, 3L:4976403-5184796)	25 genes: AlaRS-m , CG4597 , Yod1 , CG4611 , Gdap1 , Fitm , Dhrs4 , Tcs5 , CG10674 , ATPsynCF6L , CG15212 , CG15213 , CG17030 , t , Grip128 , CHMP2B , Con , lncRNA:CR34047 , lncRNA:CR43884 , Gen , shep , Srp54k , asRNA:CR45160 , lncRNA:CR45685 , lncRNA:CR45985 , lncRNA:CR46198
Deficiency 3 (BL# 7583)	$w^-;; Df(3L)^{Exel6104}/TM3,Sb$ (64C8-64C13, 3L:5184796-5366062)	11 genes: CG4669 , Cvt-c1 , Klp64D , lama , shep , tRNA:Val-TAC-1-1 , tRNA:Val-TAC-1-2 , tRNA:Ser-AGA-2-5 , Uev1A , lncRNA:CR45741 , lncRNA:CR45742 , CG46456

Appendices

Appendix 7: Alignment of Shep Isoforms

Multiple Sequence Alignment		
<p>A multiple sequence alignment of the different Shep isoforms was performed using CLUSTAL Omega (version 1.2.4). Key features of the protein such as RRM1 (NNN in green), RRM2 (NNN in blue), linker region between RRM domains (NNN in yellow highlight), and 7aa at the C-terminus (NNN in red) are color-coded, see below:</p>		
Shep-F	MHPRYSPAPPPQQQQQMGPPHQQQGGGGGGVSMRGPSNAQQQLPPQIPRSQNYSNNGSSS	60
Shep-A	MHPRYSPAPPPQQQQQMGPPHQQQGGGGGGVSMRGPSNAQQQLPPQIPRSQNYSNNGSSS	60
Shep-C	-----	0
Shep-D	-----	0
Shep-B	-----	0
Shep-H	-----	0
Shep-G	-----	0
Shep-E	-----	0
Shep-I	-----	0
Shep-F	SAAAAPLTSRSAFFPGAPLTASAVALKGALPQRPPAMTSPAAAAAGAALAAGAPYRGAASW	120
Shep-A	SAAAAPLTSRSAFFPGAPLTASAVALKGALPQRPPAMTSPAAAAAGAALAAGAPYRGAASW	120
Shep-C	-----	0
Shep-D	-----MTSPAAAAAGAALAAGAPYRGAASW	25
Shep-B	-----MTSPAAAAAGAALAAGAPYRGAASW	25
Shep-H	-----	0
Shep-G	-----	0
Shep-E	-----	0
Shep-I	-----	0
Shep-F	TPQGYAPAAAAAAVAQQAAYRYTAPLPQPAYAAYTPHTATTPATTTVSFLSQPVDDYYW	180
Shep-A	TPQGYAPAAAAAAVAQQAAYRYTAPLPQPAYAAYTPHTATTPATT-----T	168
Shep-C	-----	0
Shep-D	TPQGYAPAAAAAAVAQQAAYRYTAPLPQPAYAAYTPHTATTPATT-----T	73
Shep-B	TPQGYAPAAAAAAVAQQAAYRYTAPLPQPAYAAYTPHTATTPATT-----T	73
Shep-H	-----	0
Shep-G	-----	0
Shep-E	-----	0
Shep-I	-----	0
Shep-F	YGQRVPTAASPSNTNSSSSSNTGSQSGTLSTLSLNTTNTNTNMGPNGTVQNQNQQGGGQL	240
Shep-A	YGQRVPTAASPSNTNSSSSSNTGSQSGTLSTLSLNTTNTNTNMGPNGTVQNQNQQGGGQL	228
Shep-C	-----MGPNGTVQNQNQQGGGQL	18
Shep-D	YGQRVPTAASPSNTNSSSSSNTGSQSGTLSTLSLNTTNTNTNMGPNGTVQNQNQQGGGQL	133
Shep-B	YGQRVPTAASPSNTNSSSSSNTGSQSGTLSTLSLNTTNTNTNMGPNGTVQNQNQQGGGQL	133
Shep-H	-----MGPNGTVQNQNQQGGGQL	18
Shep-G	-----MGPNGTVQNQNQQGGGQL	18
Shep-E	-----MGPNGTVQNQNQQGGGQL	18
Shep-I	-----MGPNGTVQNQNQQGGGQL	18

Shep-F	SKT N LYIRGLQ Q GT T DKDLV N MC A Q Y GT I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	300
Shep-A	SKT N LYIRGLQ Q GT T DKDLV N MC A Q Y GT I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	288
Shep-C	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	78
Shep-D	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C --Y G F V D F EQ P A F A E C A	191
Shep-B	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C --Y G F V D F EQ P A F A E C A	191
Shep-H	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	78
Shep-G	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	78
Shep-E	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	78
Shep-I	SK T N L Y I R G L Q Q G T T D K D L V N M C A Q Y G T I IST K A I LD K TT N K C K G Y G F V D F EQ P A F A E C A	78

Appendices

Shep-F	VKGLQKGVQAQMAK-----QQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	349
Shep-A	VKGLQKGVQAQMAK-----QQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	337
Shep-C	VKGLQKGVQAQMAK-----QQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	127
Shep-D	VKGLQKGVQAQMAKVGIWVLRHPAIQQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	251
Shep-B	VKGLQKGVQAQMAKVGIWVLRHPAIQQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	251
Shep-H	VKGLQKGVQAQMAKVGIWVLRHPAIE--QDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	136
Shep-G	VKGLQKGVQAQMAKVGIWVLRHPAIQQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	138
Shep-E	VKGLQKGVQAQMAKVGIWVLRHPAIQQEQDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	138
Shep-I	VKGLQKGVQAQMAKVGIWVLRHPAIE--QDP	TNLYIANLPPHFKETDLEAML	SKYGGVV	136
***** : *****				
Shep-F	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		409
Shep-A	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		397
Shep-C	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		187
Shep-D	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		311
Shep-B	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		311
Shep-H	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		196
Shep-G	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		198
Shep-E	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		198
Shep-I	STRILRDQOMNSKGVGFARMESREKCEQIQMFNGNTIPGAKDLLVKFADGG	PKKNLF		196

Shep-F	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			469
Shep-A	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			457
Shep-C	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			247
Shep-D	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			371
Shep-B	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			371
Shep-H	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			256
Shep-G	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			258
Shep-E	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			258
Shep-I	KTPDPNARAWRDVSAEGIPVAYDPTMQQNGSVNVGTPIGVPYSRFSAPQVGGYPVAGSQ			256

Shep-F	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			529
Shep-A	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			517
Shep-C	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			307
Shep-D	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			431
Shep-B	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			431
Shep-H	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			310
Shep-G	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			318
Shep-E	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			318
Shep-I	WIPGYMMTQVDDQTSYSPQYMMAAAPPLGVTSYKPEAVNQVQPRGISMMVSGDTGVPYG			316

Shep-F	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			582
Shep-A	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			570
Shep-C	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			360
Shep-D	TMPQLATLQIGNS NFSPSLQ YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			491
Shep-B	TMPQLATLQIGNS NFSPSLQ YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			491
Shep-H	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			363
Shep-G	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			371
Shep-E	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			371
Shep-I	TMPQLATLQIGNS-----YISPTYPPYAPPPTIIPTMPMTDSEQASTAASPDEAYTQ			369

Shep-F	YPHQAAPK	590		
Shep-A	YPHQAAPK	578		
Shep-C	YPHQAAPK	368		
Shep-D	YPHQAAPK	499		
Shep-B	YPHQAAPK	499		
Shep-H	YPHQAAPK	371		
Shep-G	YPHQAAPK	379		
Shep-E	YPHQAAPK	379		
Shep-I	YPHQAAPK	377		
