A structure-function analysis of Vasa in *Drosophila* oogenesis and germ cell formation

Mehrnoush Dehghani

Department of Biology, McGill University
Montreal, QC, Canada

May 2016

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Table of content

Table of content I
Acknowledgments VI
Abstract VII
Résumé VIII
Preface X
Contributions of authors XII
List of abbreviations XIII
List of figures and tables XVII

Chapter 1: Literature review 1
1.1 Overview 1
1.2 Inductive or epigenetic germ cell specification 1
1.3 Preformation mode of germ cell development 5
1.3.1 *Caenorhabditis elegans* 5
1.3.2 *Xenopus laevis* 6
1.3.3 Zebrafish (*Danio rerio*) 7
1.3.4 *Drosophila melanogaster* 8
1.4 Structure and components of pole plasm in *Drosophila* 11
1.4.1 *oskar* 12
1.4.2 *nanos* 13
1.4.3 Tudor 14
1.4.4 Vasa 15
1.4.5 *germ-cell less* and *polar granule component* 17
1.5 DEAD-box family of RNA helicases 19
1.5.1 The conserved helicase core 19
1.5.2 Diverse flanking regions 21
1.5.3 Cellular functions of DEAD-box proteins 22
mRNA splicing: 22
mRNA export: 23
Translation initiation: 23
mRNA decay: 23

Ribosome biogenesis: 24
Transcription: 24
1.6 Translational regulation in *Drosophila* oogenesis 24
1.6.1 Eukaryotic translation initiation 25
1.6.2 Oskar translation 26
1.6.3 Nanos translation 27
1.6.4 Translation of Caudal and Hunchback 27
1.6.5 Gurken translation: role of Vas as a translational activator 28
1.7 Vas role in piRNA pathway 30
1.7.1 piRNA pathway protects genome integrity in the germline 30
1.7.2 Vas function in piRNA pathway 34
1.8 Vas function in regulating mitosis and meiosis 36
1.9 Vas orthologs in other species 38
1.9.1 *C. elegans* 38
1.9.2 Sea urchin 39
1.9.3 Zebrafish 41
1.9.4 *Xenopus* 42
1.9.5 Mouse 43
1.9.6 Human 44
1.10 Vas function in totipotent stem cells 45
1.11 Vas implication in cancer 46
1.12 Vas structural analyses; preferential binding to mRNA targets 47
1.13 Rationale for experiments 48

Chapter 2: *In vivo* mapping of the functional regions of the DEAD-box helicase Vasa 51
Abstract 51
Introduction 52
Materials and Methods 54
Fly stocks and transgenic lines 54
Western blots 54
Egg-laying assay 55
Hatching assay 55
Immunostaining and *in situ* hybridization 55

Reverse transcription quantitative PCR (RT-qPCR) 55
Live imaging 56
Statistical analyses 56
Results 56
A series of eGFP-Vas proteins with mutations in the N-terminus, conserved helicase domains and the short C-terminus 56
Most non-null mutations of *vas*, including some affecting residues highly conserved in DEAD-box helicases, can support egg production. 59
Vas requires the conserved DEAD-box domains as well as the acidic C-terminus for the translational activation of Grk. 62
The N-terminal and C-terminal domains as well as the conserved core region contain important elements for Vas localization. 64
The role of Vas in repressing transposon-encoded mRNAs requires its helicase function and is also dependent on some N-terminal and C-terminal motifs. 67
In addition to its core helicase domain, Vas requires motifs in the N-terminus and C-terminus regions to support germ cell specification. 69
Posterior patterning activity of Vas is abolished by removing the entire N-terminal domain or by point mutations in conserved DEAD-box helicase domains 71
Discussion 75
Mutant forms of Vas with abrogated DEAD-box motifs can support oogenesis and can localize to the posterior pole of the stage-10 oocyte, but RNA helicase activity of Vas is required for most of its cellular and developmental functions 75
A highly acidic Vas-specific C-terminal motif is required for many Vas functions 76
The rapidly-evolving N-terminal region contributes to Vas function 77
Acknowledgments 78
Supplementary material 79
Appendix A 85
A.1 Mapping the structural domain in the N-terminus of Vas for sequence-specific binding to RNA 85
Plasmid construction and UV cross-linking assays 86
Results: the *in vitro* assays map a sequence-specific RNA binding motif to amino acids 183-200 of Vas. 86
A.2 *In vivo* mapping of functional domains in the N-terminus of Vas using a site-specific transgenesis system 89
A.3 Further analysis of $vas^{\Delta 636\text{-}646}$ for its defects in germ cell formation 92
Connecting text 97

Chapter 3: Studying the functional role of an invariant tryptophan residue in the C-terminus of Vas 98
Abstract 98
Introduction 99
Materials and Methods 101
Endogenous *vas* alleles 101
Transgenic *vas* alleles 102
in vivo analysis 102
Yeast two-hybrid assays 102
Results 103
Vas orthologs from different species contain an invariant tryptophan in their C-terminal region 103
Substitution of Trp660 with Glu abolishes germ cell formation and embryonic viability 103
Vas localization is independent of Trp660 106
Mutation of Trp660 to Glu reduces Vas function in piRNA pathway 108
vas^{W660E} females produce a similar number of embryos to wild type; these embryos however, have defects in dorsal appendages. 110
Phenylalanine in position of Trp660 supports germ cell formation, abdominal segmentation and dorsal-ventral patterning. 112
The direct binding of Vas to its known interacting partners is independent of Trp660. 115
Discussion 115
Trp660 has a fundamental role in Vas activity: a non-conservative substitution of this residue abolishes germ cell formation and embryonic viability. 116
A *vas* transgene encoding Phe instead of Trp660 could restore Vas functions in the mutant background 117
Acknowledgments 118
Supplementary figures 119
Appendix B 122
Introduction 122
Materials and methods 122
Co-IP 122
Yeast two-hybrid 123
Plasmids: 123
Protein expression test: 123
Interaction assays: 123

Library screen: 124
Results 124
The co-immunoprecipitation approach to identify Vas-interacting partners is impeded by weak or transient nature of the interactions. 124
Y2H confirmed previously known interactions of Vas with Osk, Gus and eIF5B, but did not indicate direct interactions with the other 13 partners 131
A Y2H screen identifies new putative interacting partners of Vas 133
Discussion 140
Examining Vas interactions through a co-IP approach 140
The Y2H screen suggests new functions for Vas in miRNA pathway and ribosome biogenesis 143

Chapter 4: Conclusion and future directions 145
Additional quantitative analyses on the role of RGG motifs 145
The role of Vas in pole plasm concentration 147
Post-translational modifications of Vas 150
Further investigation on the role of Vas in cancer 151
Vas role in inter-specific hybrid dysgenesis 152

References 154

Acknowledgments

Firstly, I would like to thank my supervisor, Dr. Paul Lasko, whose support and encouragement during this project provided me great confidence as a researcher. He gave me the freedom to explore my ideas and at the same time the guidance I needed over the course of this work. I am grateful for the funds that he allocated to the costs of this project including the reagents and my stipend.

I would like to gratefully acknowledge my supervisory committee members, Dr. Laura Nilson and Dr. Richard Roy for all of their guidance through this process. I truly benefited from their constructive feedbacks, and the scientific discussions we had.

I learned a great deal from many people in Lasko lab in particular the previous research associates, Dr. Sanjay Ghosh, Dr. Greco Hernández and Dr. Chiara Gamberi who provided me with technical inputs, and helped me to develop critical thinking through interactive discussions. I also thank the current research associate, Dr. Niankun Liu, for teaching me the *in vitro* cross-linking assay when I first started my PhD, and all the other members of lasko lab for so many memorable moments that we shared. My special thanks go to Beili Hu for helping me with the microinjection of more than thirty five constructs. I also appreciate the useful feedbacks from people in Dr. Nam-Sung Moon, Dr. Laura Nilson and Dr. Frieder Schöck's labs, during our joint fly meetings. I am indebted to Elke Küster-Schöck, the previous member of Cell Imaging and Analysis Network (CIAN) at McGill, for microscopy training and technical assistance. These acknowledgments would not be complete without thanking Dr. Fillip Port for his technical advice on CRISPR-Cas9 gene editing, Vogel lab and Dr. Anja Katzemich for sharing some protocols and reagents used in yeast two-hybrid assays and Josée Houde for proofreading my French abstract.

I express my heart-felt gratitude to my family who continuously encouraged and supported me to persevere toward my goals. In particular, I must thank my uncle Bahram and his family here in Canada, who have always been there for me during happy or difficult times. I also want to take a moment and thank all my previous professors in Baha'i Institute for Higher Education (BIHE), whose courage and sacrifice has been continuously inspiring me in my academic path.

Abstract

The DEAD-box RNA helicase, Vasa (Vas), is found in the germline of all metazoans tested, and plays central roles in germ cell development. In *Drosophila* Vas has a wide range of functions in oogenesis, germ cell formation and embryonic abdominal patterning. Vas sequence contains a core region, which is highly conserved among different DEAD-box proteins and is sufficient for its helicase activity *in vitro*. Like many other DEAD-box proteins the core domain of Vas is flanked by divergent N- and C-terminal sequences, whose functional significance has not been well studied. The research presented in this thesis applies a systematic mutagenesis approach to examine different motifs in these flanking regions for their roles in various functions of Vas. We show that interval 140-200 of the N-terminus, proximal to the core domain, contains essential elements for Vas localization to the nuage and pole plasm. This partly accounts for impaired function of $Vas^{\Delta 3\text{-}200}$, in transposon silencing, Gurken translation, abdominal patterning of embryos and germ cell formation. We further show that replacing RGG motifs with AGG in the N-terminus significantly decreases Vas activity in germ cell formation.

Our analyses on the C-terminal region reveal that motif 636-646, which is conserved among Vas orthologs from closely related *Drosophila* species, is required for condensed localization of pole plasm components at the posterior region. Embryos from $vas^{\Delta 636\text{-}646}$ females produce pole buds, which however mostly regress due to insufficient amounts of pole plasm components. In addition, most Vas orthologs terminate with a highly acidic motif, which also contains a conserved tryptophan, Trp660 in the *Drosophila* protein. Deleting motif 655-661 altogether or mutating Trp660 to glutamic acid abolishes Vas activity in germ cell formation, and severely impairs Grk translation, transposon silencing and embryonic abdominal patterning. We also observed that, despite being highly conserved, Trp660 could be replaced with phenylalanine, another hydrophobic aromatic residue, for most Vas functions in oogenesis and germ cell formation.

Together these analyses indicate that various functions of Vas could be uncoupled through mutating different motifs. Biochemical or genetic approaches, such as yeast two-hybrid assay, which was applied in this study, could explore the effect of these mutations on Vas-associated ribonucleoprotein complexes.

Résumé

L'ARN hélicase Vasa (Vas) se trouve dans la lignée germinale de tous les métazoaires testés et joue un rôle central dans le développement des cellules germinales. Chez la *Drosophila*, Vas possède un large éventail de fonctions dans l'ovogenèse, dans la formation des cellules germinales et dans la segmentation abdominale embryonnaire. La séquence de Vas contient une région central qui est hautement conservée entre les différentes protéines DEAD-box et qui est suffisante pour son activité hélicase *in vitro*. Comme beaucoup d'autres protéines DEAD-box, le domaine central de Vas est flanqué en N- et C-terminal par des séquences divergentes, dont les fonctions n'ont pas été bien étudiées. La recherche présentée dans cette thèse applique une approche de mutagenèse systématique pour examiner les rôles des différents motifs présents dans ces régions flanquantes. Nous démontrons que l'intervalle 140-200 de la région N-terminale, près du domaine central, contient des éléments essentiels pour la localisation de Vas au nuage et au plasm polaire. Cela explique en partie la fonction perturbée de $Vas^{\Delta 3\text{-}200}$ dans le silençage des transposons, la traduction de Gurken, la segmentation abdominale des embryons et la formation des cellules germinales. En outre, nous démontrons que le remplacement du motif RGG par le motif AGG dans la région N-terminale diminue significativement l'activité de Vas dans la formation des cellules germinales.

Nos analyses sur la région C-terminale révèlent qu le motif 636-646 est nécessaire pour la localisation condensée de composants du plasm polaire à la région postérieure des embryons. Ce motif est conservé parmi les orthologues de Vas provenant d'espèces de *Drosophila* étroitement liées. Les embryons des femelles $vas^{\Delta 636\text{-}646}$ produisent des bourgeons polaires dont la plupart régressent à cause de quantités insuffisantes de composants du plasm polaire. Par ailleurs, la plupart des orthologues de Vas se terminent par un motif fortement acide contenant également un tryptophane conservé, Trp660. L'élimination complète de ce motif 655-661 ou le changement du Trp660 pour un acide glutamique abolit l'activité de Vas dans le développement des cellules germinales. De plus, ceci compromet gravement la traduction de Gurken, le silençage des transposons et la segmentation abdominale embryonnaire. Nous avons également observé qu'en dépit d'être hautement conservé, Trp660 pourrait être remplacé par une phénylalanine, un autre résidu aromatique hydrophobe, pour la plupart des fonctions de Vas dans l'ovogenèse et dans la formation des cellules germinales.

Ensemble, ces analyses indiquent qu'il est possible d'analyser les différentes fonctions de Vas par des mutations de motifs spécifiques. Les approches biochimiques ou génétiques, telle que le criblage double-hybride qui a aussi été appliqué dans cette étude, pourraient explorer l'effet de ces mutations sur les complexes de ribonucléoprotéines associés à Vas.

Preface

The germ cell lineage is responsible for producing new individuals in most multi-cellular organisms, and thus provides a cellular continuity between generations. A complex network of regulatory genes, which in many cases are conserved across species, functions to safeguard germ cell specification, maintenance, and migration. One highly conserved component, whose function is required at multiple stages of germ cell development in various species, is the RNA helicase, Vasa (Vas). We used *Drosophila*, as a versatile genetic system, to study different aspects of Vas function in germline. The first chapter overviews common mechanisms of germ cell specification across the metazoans, with a particular focus on *Drosophila*. Distinct functions of Vas in *Drosophila* oogenesis, germ cell formation and embryonic patterning are discussed. In addition, Vas orthologs in other animal models and in human are reviewed. This chapter also summarizes our knowledge of the functional elements in Vas structure, and the questions that remain to be answered.

Chapter 2 presents data obtained from a systematic mutagenesis approach that we applied to target various motifs in the poorly studied N- and C-terminal regions as well as the highly conserved core helicase domain of Vas. These mutations were characterized for their effects on Vas localization and functions during oogenesis, and early embryogenesis. Through this study several elements in the N- and C-terminal regions of Vas were specifically linked to different functions of Vas in piRNA biogenesis, Gurken translation and germ cell specification.

The structure-function analyses presented in chapter 2 revealed that a C-terminal motif (655-661), containing several acidic residues and a highly conserved tryptophan, Trp660, is essential for Vas function in germ cell formation. Interestingly, the Trp residue is specifically found in Vas orthologs and, unlike conserved residues in the helicase domain, is not shared among DEAD-box proteins. Thus we subsequently focused our studies on the role of Trp660 in various Vas functions, through non-conservative or conservative substitutions. Results of these analyses are included in a manuscript presented in chapter 3, which we are currently finalizing. In addition, we applied biochemical and genetic approaches, particularly a yeast two-hybrid assay, to explore Vas protein interactome in response to deletion of motif 655-661 or mutation of Trp

660. Results of these experiments are briefly presented in chapter 3, and discussed with more details in appendix B.

Chapter 4 further discusses some of the results presented in the previous chapters, and suggests follow-up experiments. Potential approaches to overcome the technical challenges that we faced in some of our experiments, and possible directions for future studies of Vas are also discussed in this chapter.

Contributions of authors

All the data presented in chapters 2 and 3, and the appendices are my original works. I prepared all sections of this thesis, including the illustrations and the figures, in partial fulfillment of the degree requirements. Dr. Paul Lasko provided me with valuable inputs in designing the experiments and analyzing the data. He reviewed the thesis, edited the text and added helpful comments.

Chapter 2 includes Dehghani and Lasko (2015). I performed all the experiments, prepared the figures and wrote the initial draft of the manuscript. Dr. Lasko supervised my work, provided the funding, edited the manuscript and communicated with the journal. I prepared a draft of responses to the reviewers' comments, which was finalized by Dr. Lasko.

The manuscript presented in Chapter 3 will be submitted as Dehghani and Lasko (2016). All the experiments described in this section were performed by me. I prepared the figures and the first draft of the manuscript. Dr. Lasko helped me with data analyses, edited the manuscript and will communicate with the journal.

List of abbreviations

(U)-rich	Uridine-rich
A	Adenine
a.a.	amino acid
AD	Activation domain
aDMA	asymmetrically dimethylated arginines
ago3	*argonaut 3*
AP	Anterior posterior
aret	*arrest*
Arg	Arginine
ATP	Adenosine triposphate
Aub	Aubergine
Bam	Bag-of-marbles
Barr	Barren
bcd	*bicoid*
BD	DNA binding domain
Bel	Belle
BicD	Bicaudal D
Bio-ID	Proximity-dependent biotin identification
BMP	Bone morphogenetic protein
brat	*brain tumor*
BRE	Bru response elements
Bru	Bruno
Btz	Barentsz
C. elegans	*Caenorhabditis elegans*
cad	*caudal*
CB	Cystoblast
CDS	Coding DNA sequence
co-IP	co-Immunoprecipitation
CPEB	Cytoplasmic polyadenylation element binding
CRISPR	Clustered regularly-interspaced short palindromic repeats
Csul	Capsuleen
CTD	C-terminal domain
CTE	C-terminal extension
C-terminal	Carboxyl-terminal
Cuff	Cutoff
D	Aspartic acid
D. melanogaster	*Drosophila melanogster*
Dazl	Deleted in azoospermia-like

DCR-1	Dicer-1
DD	Dimerization domain
DDO	Double dropout
DDO/A	DDO/aureobasidin A
DDO/A/X	DDO/ aureobasidin A /X-α-Gal
Dlic	Dynein light intermediate chain
dpc	days post coitum
dRMTs	*Drosophila* arginine methyltransferases
DSB	Double-strand breaks
DSP	Dithiobis [succinimidyl propionate]
dsRNA	double-stranded RNA
DTC	Distal tip cell
DTT	Dithiothreitol
DV	Dorsal ventral
E	Glutamic acid
E 10.5	Embryonic day 10.5
E. coli	*Escherichia coli*
eGFP	enhanced green fluorescent protein
EGFR	Epidermal growth factor receptor
Egl	Egalitarian
eIF	eukaryotic initiation factor
EJC	Exon junction complex
ExE	Extraembryonic ectoderm
Exu	Exuperantia
F	Phenylalanine
FAF	Fat facets
FL	Full length
FSN	F-box synaptic protein
ftz	*fushi tarazu*
G	Glycine
gcl	*germ cell-less*
GLH-1	Germ line helicase-1
Glo	Glorund
Glu	Glutamic acid
Gly	Glycine
grk	*gurken*
GSC	Germline stem cell
GST	Glutathione S-transferase
Gus	Gustavus
HA	Hemagglutinin
hb	*hunchback*

hESCs	Human embryonic stem cells
HP1	Heterochromatin protein 1
IgG	Immunoglobulin G
iPSCs	induced pluripotent stem cells
K	lysine
kb	kilobase
kDa	kilodalton
L	Leucine
L(3)mbt	Lethal (3) malignant brain tumor
LC-MS/MS	Liquid chromatography coupled with tandem mass spectrometry
Mago	Mago nashi
miRNA	microRNA
smms	small micromeres
MT	Microtubule
MTOC	Microtubule organizing center
MVH	Mouse Vas Homolog
N	Asparagine
NMD	nonsense-mediated mRNA decay
Nos	Nanos
NPC	Nuclear pore complex
N-terminal	NH_2-terminal
ORF	Open reading frame
OrR	*Oregon R*
Osk	Oskar
P bodies	Processing bodies
PABP	Poly(A)-binding protein
PBS	Phosphate-buffered saline
pgc	*polar granule component*
PGCs	Primordial germ cells
Phe	Phenylalanine
PIC	Preinitiation complex
piRISC	piRNA-induced RISC complex
piRNA	Piwi-interacting RNA
piwi	*p-element induced wimpy testis*
PSC	Pluripotent stem cells
PTM	Post-translational modification
Pum	Pumilio
Q	Glutamine
QDO	Quadruple dropout
R	Arginine

RBD	RNA-binding domain
RBP	RNA-binding proteins
RecA	Recombinase A
Rhi	Rhino
RNAi	RNA interference
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
SD medium	Synthetic defined medium
sDMA	symmetrically dimethylated arginines
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of mean
Ser2	Serine 2
SF2	Superfamily2
siRNA	Small interfering RNA
smms	Small micromeres
snRNP	Small nuclear ribonucleoprotein
spn-E	Spindle-E
ssRNA	single stranded RNA
Stau	Staufen
T	Threonine
TDO	Triple dropout
Tej	Tejas
Trp	Tryptophan
RT-qPCR	Quantitative reverse transcription *PCR*
Tub	Tubulin
Tud	Tudor
UTR	Untranslated region
UV	Ultraviolet
(U)-rich	Uridine-rich
V	Valine
Vas	Vasa
VCP	Vasa core protein
VE	Visceral endoderm
W	Tryptophan
wt	*wild type*
XVLG1	*Xenopus* vasa-like gene 1
X-α-Gal	X-α-galactosidase
X-β-Gal	X-β-galactosidase
Y2H	Yeast two-hybrid
Zuc	Zucchini

List of figures and tables

Fig. 1.1 Germ cell specification in different model organisms 4

Fig. 1.2 Conserved structure and mechanism of function of DEAD-box proteins..... 20

Fig. 1.3 piRNA transcripts are derived from transposon-rich heterochromatin, called piRNA clusters, in the pericentromeric or subtelomeric regions. 33

Fig. 2.1 Summary of the deletions and point mutations examined in this study. 58

Fig. 2.2 Fecundity of *vas*^{PH165} females carrying different *egfp-vas* constructs, and expression levels from those constructs. 61

Fig. 2.3 Dorsal-ventral patterning in eggs produced by *vas*^{PH165} females carrying different *egfp-vas* constructs. 63

Fig. 2.4 Localization of eGFP-Vas proteins in *vas*^{1} ovaries. 66

Fig. 2.5 HeT-A expression in ovaries of *vas*^{PH165} females carrying different *egfp-vas* constructs 68

Fig. 2.6 Germ cell formation in embryos from *vas*^{1} females expressing different eGFP-Vas proteins. 70

Fig. 2.7 Time course of pole cell development in eGFP-Vas^{+} and eGFP-Vas$^{\Delta 636\text{-}646}$ expressing embryos. 72

Fig 2.8 The ability of various *egfp-vas* constructs to restore abdominal segmentation in *vas*^{1} embryos. 73

Fig. 2.S1 The same western blot as in Fig. 2.2 B stained with anti-GFP to compare expression levels of eGFP-Vas proteins in ovaries from *vas*^{1}/+ females carrying the different constructs... 79

Fig. 2.S2 Representative confocal images of Grk immunostaining in the subset of stage 8 egg chambers that were positively stained in each genotype. 80

Fig. 2.S3 A comparison between eGFP-Vas^{+}, eGFP-Vas$^{\Delta 17\text{-}110,\ 3xAGG}$ and eGFP-Vas$^{\Delta 3\text{-}139}$ for their stability in 0-2 h embryos from *vas*^{1}/+ females. 81

Fig. 2.S4 *vas*^{1}*; egfp-vas*$^{\Delta 15\text{-}75}$ and *vas*^{1}*; egfp-vas*$^{\Delta 636\text{-}646}$ are compared for their variability in the number of pole cells per stage 4-5 embryos. 82

Fig. A.1 Mapping sequence-specific RNA-binding domain in Vas through an *in vitro* cross-linking assay..... 88

Fig. A.2 Functional comparisons between different *egfp-vas* alleles substituting RGG with AGG or KGG..... 91

Fig. A.3 When examined in a *wt* background eGFP-Vas$^{\Delta 636\text{-}646}$ segregation into the germ cells and its stability is comparable to eGFP-Vas$^{\Delta 15\text{-}75}$ (*wt* control). 93

Fig. A.4 Localization of germ cell components in the posterior region of *vas*$^{\Delta 636\text{-}646}$ embryos. .. 95

Fig. 3.1 Vas orthologs share a conserved tryptophan (W) in their C-terminal end, which is not common between the other DEAD-box proteins.. 104

Fig. 3.2 Substitution of tryptophan (W) with glutamic acid (E), which is an abundant residue in the C-terminus of Vas, is associated with the failure in germ cell formation and abolished viability. 105

Fig. 3.3 eGFP-Vas^{W660E} mimics the localization pattern of eGFP-Vas^{+} in ovaries..... 107

Fig. 3.4 Substitution of Trp660 with Glu impairs Vas function in transposable element silencing. 109

Fig. 3.5 *vas*^{W660E} females produce a wild type number of embryos. The W660E mutation, however, has a significant effect on the formation of dorsal appendages ……111

Fig. 3.6 Conservative substitution of Trp with Phe does not affect the *vas* functions in germ cell formation, embryonic patterning or Grk translation ……. 114

Fig. 3.S1 Interactions of Vas with Osk, Gus and eIF5B are not abolished by Δ655-661 or W660E. 119

Fig. B.1 A comparison between a Vas antibody (A), a GFP antibody (B) and GFP-Trap beads (C) for their efficiency to immunoprecipitate eGFP-Vas from embryos..... 126

Fig. B.2 A co-IP approach to isolate Vas associated protein complexes from *Drosophila* ovaries and embryos. 127

Fig. B.3 an *egfp-vas* allele with mutation in DEAD motif (D399L) was used to form stable interactions in the RNP complexes..... 130

Fig. B.4 Candidate genes from *Drosophila* were N-terminally fused to the activation domain (AD: 12.5 kDa) of Gal4 and expressed in yeast. 132

Fig. B.5 Direct interactions between candidate genes (Table 3.S1) and *vas*$^{+}$ were tested through a yeast two-hybrid assay. 134

Fig. B.6 Three different assays were used to confirm positive interactions identified in the yeast two-hybrid screen using *vas*$^{+}$ as the bait and a *Drosophila* cDNA library as the prey. 136

Fig. B.7 The cDNA library was tested for the presence of *vas* and its known interacting partners. 139

Fig. B.8 Among the genes identified to interact with *vas* through my yeast two-hybrid screens, nine that are expressed in the ovaries (*mRpl44, Rpp20, CG2678, pasha, CG2091, CG1239, Pka-C1, eas* and *cip4*) were isolated from yeast, amplified in *E. coli* and transformed back into Y187 yeast strain to subsequently test their interactions with *vas* mutant proteins. 141

Table 2.S1 Results of the statistical analyses **……** 83

Table 3.S1 The list of the proteins that have been previously found to associate with Vas through different assays, including yeast two-hybrid (Y2H), GST pull-down or co-immunoprecipitation (co-IP). 120

Table B.1 The top 27 proteins identified by LC/MS/MS analyses in the co-IP elutes from *OrR* (*wt* control) and *egfp-vas* embryos using GFP-Trap beads..... 129

Table B.2 list of the genes identified in a yeast two-hybrid screen using the full-length *vas* as the bait and a *Drosophila* cDNA library (Clontech Inc.) as the prey 137

Chapter 1: Literature review

1.1 Overview

In sexually reproducing animals a population of cells, called germ cells, is set aside during early development, to give rise to gametes in the adult gonads. Various mechanisms, which in many aspects are similar between different species, ensure specification and maintenance of germ cells, their ability to migrate to the gonads, and to undergo meiosis and sex-specific differentiations. Our knowledge of these mechanisms will help to understand developmental processes relevant to human infertility, or other pathological conditions that result from ectopic activation of germline-specific pathways.

The DEAD-box RNA helicase, Vasa (Vas), is expressed in the germline of various organisms and plays critical roles in germ cell development. *vas* has been studied most extensively in *Drosophila melanogaster*, where its functions are required in different stages of oogenesis and germ cell formation. These studies show that Vas associates with different ribonucleoprotein complexes, and regulates translation of germ cell-specific mRNAs; however, the exact mechanisms of Vas functions are still incompletely understood.

In this literature review I first address the key events during germ cell development in the context of the best-studied animal models, and then outline different functions of Vas orthologs in germ cell development. In addition many studies, which will be reviewed here, have reported that Vas functions in somatic stem cells and is also ectopically expressed in several types of cancer. I will further discuss structural features and cellular functions of DEAD-box proteins, and more specifically focus on the few previous studies that provide insights into Vas structure and its interactions with other proteins or mRNA.

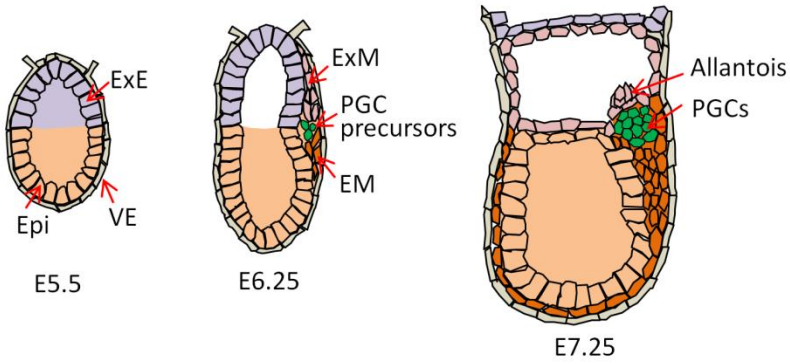
1.2 Inductive or epigenetic germ cell specification

In mice and many other mammals, primordial germ cells (PGCs) are specified after implantation of the blastocyst. A small group of cells in the posterior proximal region of the epiblast receive inductive signals from extraembryonic ectoderm (ExE) and visceral endoderm

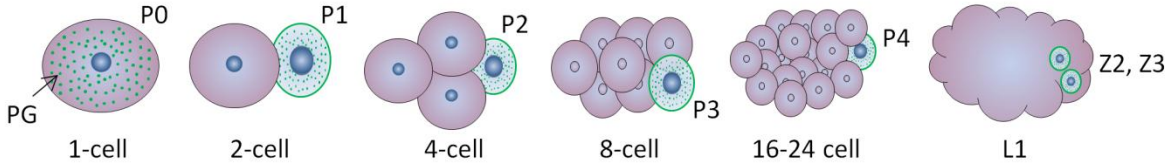
(VE), instructing them to adopt a PGC fate (Fig. 1.1 A; Hayashi et al., 2007; Tsang et al., 2001). Members of the bone morphogenetic protein (BMP) family, such as BMP4, BMP8b and BMP2, which are secreted from ExE and VE, are identified as some of these inductive signals (Lawson et al., 1999; Ying et al., 2000; Ying and Zhao, 2001). The BMP signal transducers, Smads, are expressed in the epiblast at the same time and are necessary for germ cell formation (Chang and Matzuk, 2001; Saitou et al., 2002; Tremblay et al., 2001). In a dose-dependent manner Smad-mediated BMP signaling induces expression of the transcription repressor, *Blimp1/Prdm1*, in a few cells in posterior proximal epiblast (Arnold et al., 2006; Kurimoto et al., 2008). Blimp1 is critical for lineage-restriction of PGCs through suppressing expression of somatic genes such as *Hox* genes (Ohinata et al., 2005). In complex with PRMT5, the arginine histone methyltransferase, Blimp1 further orchestrates specific epigenetic modifications in germ cells to maintain pluripotency during their migration into gonads (Hayashi et al., 2007). The epigenetic reprogramming of epiblastic cells into PGCs has been recapitulated in mouse pluripotent stem cells (PSCs), through a transient state called epiblast-like cells (EpiLCs), by either addition of BMP4 into the culture or over-expression of Blimp1 (Hayashi et al., 2011).

Upon their migration to the gonadal ridges, at E10.5-E11.5, PGCs start to express factors, such as the RNA-binding proteins Dazl and Vasa (Vas), which are required for their maturation and survival (Castrillon et al., 2000; Gill et al., 2011). Over-expression of Dazl and Vas promotes meiotic progression in pluripotent human embryonic stem cells (hESCs) *in vitro* (Medrano et al., 2012). It is proposed that the divergent RNA binding proteins such as Vas, Nanos (Nos), Pumilio (Pum) and Piwi, which are up-regulated in germ cells, are involved in translational control of meiosis-related genes. Shortly after gonadal colonization, PGCs acquire sex-specific properties, which include a specific DNA methylation profile and, in females, X-inactivation based on the microenvironment of the gonadal niche (Martinez-Arroyo et al., 2014). Pluripotency ends by this stage although some markers associated with pluripotency continue to be expressed even in the gonial stem cells (Leitch and Smith, 2013). Germ cells that enter fetal ovaries start meiosis but are arrested in late prophase of meiosis I until puberty, whereas in the fetal testis, as a result of meiosis inhibitors, germ cells do not start meiosis and instead remain in a mitotic arrest (McLaren, 1995).

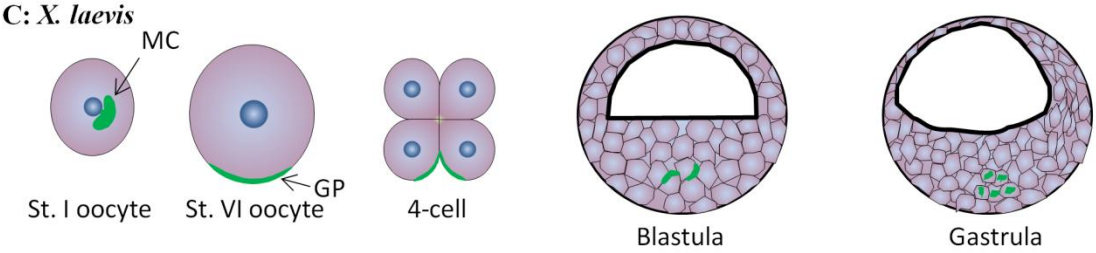
A: *M. musculus*



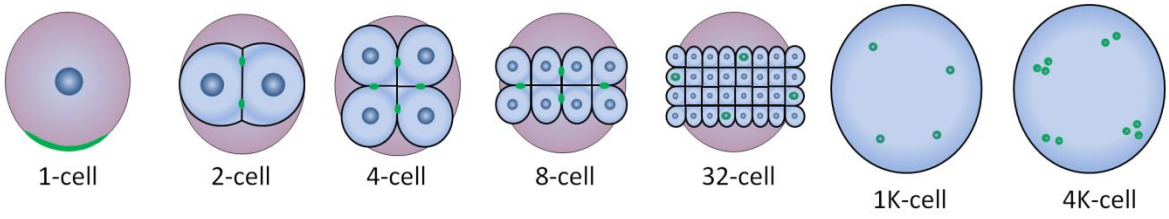
B: *C. elegans*



C: *X. laevis*



D: *D. rerio*



E: *D. melanogaster*

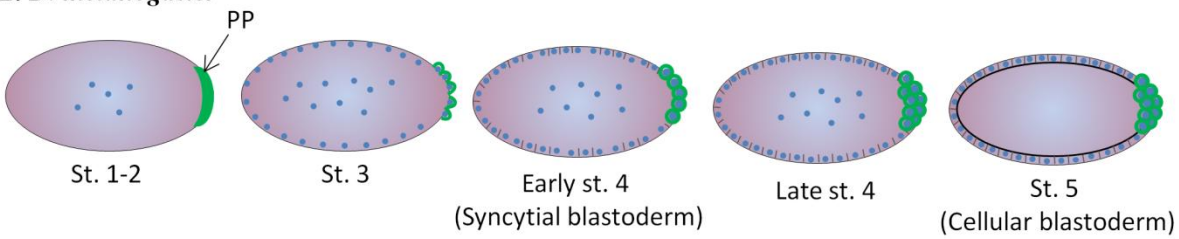


Fig. 1.1 Germ cell specification in different model organisms, **A.** In mice precursors of primordial germ cells (PGCs) first appear in the posterior proximal epiblast (epi) at ~E6.25, in response to signals from extra-embryonic ectoderm (ExE) and visceral endoderm (VE). A cluster of about 40 PGCs then derive from these cells at the base of allantois at ~E7.25 (Saitou & Yamaji, 2012). **B**. P granules (PGs) in *C. elegans* are first distributed throughout the 1-cell embryos, but become asymmetrically segregated into germline (P1, P2, P3, P4) blastomeres during the first four cell divisions. P4 divides only once during embryogenesis to produce the two PGCs, Z2 and Z3 (Hubbard and Greenstein, 2005). **C**. Germ plasm (GP) in *Xenopus* originates within the "mitochondrial cloud" (MC) in previtellogenic oocytes, and becomes localized at the vegetal pole at the onset of vitellogenesis. With embryonic cleavage germ plasm becomes incorporated into a small number of blastomeres, which give rise to PGCs at the gastrula stage (King, 2014). **D**. In Zebrafish (*Danio rerio*) germ plasm components, which are initially anchored to the vegetal cortex translocate to the animal pole and localize at the distal ends of cleavage furrows in 2-cell and 4-cell embryos. With subsequent cell divisions germ plasm becomes asymmetrically segregated into only four cells until 1K-cell stage. During 4K-cell stage germ plasm spreads in the cytoplasm of the cells, which divide and produce PGCs in four different positions within the embryo (Raz, 2003). **E**. In *Drosophila* pole plasm (PP) is localized at the posterior region of early embryos. At stage 3 pole cells bud off from the posterior pole, and their nuclei undergo two more divisions before they cellularize at stage 4. Cellularization of the somatic cells occurs during stage 5 (Santos and Lehmann, 2004).

1.3 Preformation mode of germ cell development

In many organisms, including *Drosophila, C. elegans, Xenopus* and zebrafish, there is a physical continuity of germline across generations, meaning that a specific cytoplasmic region in the oocyte, termed germ plasm, contains maternally inherited mRNA-protein complexes that segregate into subsequently developing PGCs (Lesch and Page, 2012). Work in *Drosophila* shows that transplantation of germ plasm, which is formed at the posterior pole and thus is often called pole plasm, to the anterior region of oocyte induces ectopic germ cells, providing evidence that germ plasm contains all the necessary factors for germ cell specification (Illmensee and Mahowald, 1974). In the next sections characteristics of predetermined germ cell formation that have been studied in model organisms will be discussed.

1.3.1 *Caenorhabditis elegans*

In *C. elegans* non–membrane-bound organelles composed of maternally encoded protein and RNAs, referred as P granules, are dispersed in the cytoplasm of the one-cell embryo (Fig. 1.1 B). During four subsequent asymmetric cell divisions, the mitotic spindles are displaced toward one side of the cell, where P granules also accumulate; thus the smaller daughter cell inherits most of the germ cell determinants and develops into germline (P) lineage. P4 divides only once to form two PGCs, Z2 and Z3, which remain arrested in G2 until the mid-L1 stage (Gönczy and Rose, 2005; Strome, 2005).

There is evidence in *C. elegans* for three key mechanisms that mediate germ cell specification and maintenance: 1) transcriptional inhibition of somatic fate-inducing genes, 2) translational regulation of germ cell specific mRNA, and 3) establishment of appropriate chromatin structure required for transcriptional silencing. One of the P granule components is the CCCH-type zinc finger protein, PIE-1, which acts as a transcription repressor (Mello et al., 1996; Seydoux et al., 1996). PIE-1 inhibits phosphorylation of the RNA polymerase II C-terminal domain (CTD) by CDK9, and thus suppresses transcription of somatic genes at the elongation step (Zhang et al., 2003). P granules are also enriched for numerous RNA-binding proteins, several of which are implicated in translational control (Lee et al., 2006). *nos-1* and *nos-2*, required for PGC development in *C. elegans*, are related to *Drosophila nanos* (*nos*), which has known functions in translational regulation during germline development (Parisi and Lin, 2000;

Subramaniam and Seydoux, 1999). Translation of *nos-2* is suppressed in the oocyte and in the germline blastomeres until the 28-cell embryo stage by several maternal RNA-binding proteins, namely OMA-1, OMA-2, MEX-3 and SPN-4, and then is specifically activated in P_4 by POS-1 (D'Agostino et al., 2006; Jadhav et al., 2008). The transcriptionally repressed state of germline precursors, Z2 and Z3, is also governed by a compact configuration of numerous chromatin domains (Wang and Seydoux, 2013). Four chromatin regulating genes, mes-2, 3, 4 and 6, were identified in a screen for maternal factors involved in fertility (Capowski et al., 1991). The unique chromatin architecture is associated with a decrease of the active chromatin mark, H3K4me2, and is a conserved feature of germ cells (Furuhashi et al., 2010; Schaner et al., 2003).

Z2 and Z3 begin proliferating in mid-L1 concurrent with division of somatic precursors, which produces the somatic gonad primordium and the two distal tip cells (DTCs), each capping one gonad arm (Pazdernik and Schedl, 2013). A robust proliferation of germ cells occurs in L4, in the distal mitotic zone of the gonads in response to signals from DTC. At the same time germ cells that are some distance away from the DTC enter the meiotic cycle, which is regulated by a complex network of genes (Kimble and Crittenden, 2005). For example, mRNAs such as *gld-1*, *2*, *3* and *nos-3,* which promote meiosis progression, are repressed in the proximal cells to DTC by the RNA-binding proteins, FBF1 and FBF2, expressed in response to Notch signalling from DTC (Crittenden et al., 2002; Lamont et al., 2004). In hermaphrodites, spermatogenesis occurs in L4 and oogenesis in the adult stage. Oocyte arrests during meiotic prophase, and its maturation, which in turn stimulates ovulation, is triggered by the signals from sperm (Hubbard and Greenstein, 2005). On the contrary, sperms complete meiosis without any arrest, while their final maturation as motile spermatozoa occurs after they enter into the spermatheca.

1.3.2 *Xenopus laevis*

Maternal mRNA and protein localization in *Xenopus,* which takes place during the course of oocyte differentiation and growth, is critical for germ cell formation and embryonic patterning (Kloc et al., 2001). This process depends on the polarized cytoskeleton, the trans-acting factors, and for mRNAs the cis-acting elements in the 3′ UTR (Deshler et al., 1998; Forristall et al., 1995; Havin et al., 1998; King et al., 2005). Germ plasm first appears within the mitochondrial cloud, a distinctive cytoplasmic structure enriched for mitochondria and electron-dense granulofibrillar material in the previtellogenic oocyte, and contains a large group of germ

cell-specific maternal factors (Fig. 1.1 C; Heasman et al., 1984; Ikenishi, 1998; King, 2014) Mitochondrial cloud disperses after vitellogenesis, and fragments of germ plasm become anchored to the vegetal cortex. After fertilization and during cleavage germ plasm segregates into the vegetal blastomeres which will develop into PGCs (Strome and Updike, 2015; Whitington and Dixon, 1975). PGCs then move laterally toward the presumptive gonadal ridge and remain quiescent until three weeks of age (Kataoka et al., 2006; Wylie and Heasman, 1976). For the first time in vertebrates it was shown that germ plasm ectopically transplanted into animal hemisphere of *Xenopus* embryos could induce formation of PGC-like cells, which however fail to migrate from the ectopic position towards the genital ridge (Tada et al., 2012). When gonads become sexually differentiated PGCs start their mitotic divisions and differentiate into spermatogonia and oogonia. Primary oogonia multiply to produce groups of 16 interconnected oogonia called the cysts. The cells in each cyst enter meiosis in synchrony and produce oocytes, which in late pachytene become separated from each other, and individually surrounded by follicle cells. Oocytes arrest in diplotene and wait for the permissive hormonal condition, which may take years, before meiosis maturation. All the vegetally localized RNAs, which are critical for PGC formation, mesoderm induction and axis determination, are already anchored to their final destination in the meiotically arrested oocyte (Hausen and Riebesell, 1991; Kloc et al., 2001).

1.3.3 Zebrafish (*Danio rerio*)

The first confirmation of predetermined germ cell specification in zebrafish came from *in situ* hybridization experiments in whole mount early embryos, showing that the germline-specific mRNA, *vas*, is present in the embryos as early as in the two-cell stage, and localizes along the cleavage planes (Fig. 1.1 D; Yoon et al., 1997). With consecutive cell divisions and until the 1000-cell stage the *vas*-containing germ plasm is segregated to only four PGCs, which then start to divide and produce around 25–50 PGCs before 4000-cell stage. Furthermore, ablation of the cytoplasm at the distal ends of cleavage furrows at four-cell stage, where *vas* and *nos1* mRNAs are localized, results in significantly decreased number of PGCs, indicating the essential role of maternally supplied factors for germ cell specification (Hashimoto et al., 2004).

Large scale screens to find tissue-specifically expressed genes in zebrafish identified several factors that are expressed in the germline (Koprunner et al., 2001; Kudoh et al., 2001). Among

these a *nanos*-related gene (*nos1*) is specifically expressed in PGCs, and its translational inhibition using morpholinos results in severe defects in PGC development (Koprunner et al., 2001). Furthermore, in zebrafish PGC migration is unique in that germ cells are specified in four different positions within the embryo before gastrulation, and PGCs need to find their migratory path to the gonads regardless of their starting point. This complex phenomenon, which involves active movement of PGCs, relies on directional cues from somatic tissues (Paksa and Raz, 2015; Weidinger et al., 1999; Weidinger et al., 2002). In screens to identify genes involved in PGC migration zebrafish has been used as a more practical vertebrate model than mouse, partly due to its extrauterine development and the large number of offspring (Patton and Zon, 2001). These studies have led to discovering critical genes such as *cxcr4b*, which is a chemokine receptor expressed in PGCs during their migration (Knaut et al., 2003). Cxcr4 and its ligand Sdf1/Cxcl12 are also involved in neuronal-cell migration, tumor cell metastasis and stem cell mobilization in mammals (Chatterjee et al., 2014; Mukherjee and Zhao, 2013; Petit et al., 2002; Tiveron and Cremer, 2008; Yang et al., 2015a).

1.3.4 *Drosophila melanogaster*

Due to numerous experimental advantages, *Drosophila* has been used for the most comprehensive studies on predetermined mode of germ cell specification. *Drosophila* heavily relies on maternally supplied proteins and mRNA for early development, and thus requires posttranscriptional regulation mainly at the level of mRNA localization and translation (Becalska and Gavis, 2009; Lasko, 2012). Nurse cell-produced materials including germ cell determinants are unidirectionally transported to the oocyte via polarized microtubules (MTs), which in stage 2-6 oocytes emanate from microtubule organizing center (MTOC) at the posterior region and extend their plus ends through the ring canals into the nurse cells (Harrison and Huebner, 1997; Roth and Lynch, 2009; Weil, 2014). In stages 7–10a (mid-oogenesis), MTs reorganize, presumably in response to the signals from the posterior follicle cells, and emanate from the lateral and anterior cortex with a biased random polarity toward the posterior (Parton et al., 2011; Steinhauer and Kalderon, 2006; Zimyanin et al., 2008). Germ cell components then become localized to the pole plasm at the posterior region of the oocyte (Fig. 1.1 E). Different mechanisms have been reported for successful localization of germ cell determinants (Meignin and Davis, 2010): 1) degradation in the bulk of cytoplasm and stabilization at the posterior pole,

as observed for *nos* mRNA (Zaessinger et al., 2006), 2) diffusion and anchoring, for example, reported for *cyclin B*, *germ cell-less* (*gcl*) and *nos* (Forrest and Gavis, 2003), 3) active transport mediated by motor proteins and cis elements in 3′ UTR of mRNAs such as *oskar* (*osk*) (Clark et al., 2007; Zimyanin et al., 2008), and 4) localized translation of proteins such as Osk exclusively in the pole plasm (Kim-Ha et al., 1995). Following fertilization and during syncytial divisions, pole plasm complexes are recruited by posterior nuclei through a dynein-mediated transport on the astral microtubules emanated from centrosomes (Lerit and Gavis, 2011). By this mechanism germ plasm becomes segregated into the pole buds, which pinch off from posterior region at mitotic cycle 9 (stage 3; Campos-Ortega and Hartenstein, 1985) and undergo two more divisions before cellularization at the end of cycle 10 (early stage 4). Nuclear divisions at the posterior region are asynchronous and lag behind the somatic divisions. Germ cell precursors then enter a prolonged quiescence until arriving in the presumptive gonads in late embryogenesis (Su et al., 1998). The mitotic quiescence is partly due to translational repression of *cyclin B* by RNA-binding proteins Nos and Pum (Asaoka-Taguchi et al., 1999).

Transcriptional quiescence is a key feature of germ cell development during early stages, and often depends on inhibition of RNA polymerase II phosphorylation (Seydoux and Dunn, 1997). In *Drosophila polar granule component* (*pgc*), which localizes to the pole plasm and becomes translated as a small protein, is involved in suppression of mRNA transcription in germ cells (Hanyu-Nakamura et al., 2008; Timinszky et al., 2008; section 1.4.5). Transcriptional quiescence in germ cells continues until 3.5 hours after egg laying (stage 8), in contrast to the somatic cells that start their zygotic transcription after 1 hour (Van Doren et al., 1998; Zalokar, 1976). In addition, maintenance of germline lineage is governed by epigenetic regulation such as reduced methylation of histone H3 (H3meK4), which in germ cells unlike somatic cells persists until stage 9 of embryogenesis (Schaner et al., 2003). In the absence of transcription in germ cells translational control of pre-existing mRNAs becomes important. Supporting this notion several RNA-binding proteins, including Nos and Pum, which are known as translation inhibitors, and Vas, a translational activator, are present in the germ cells (Irish et al., 1989; Johnstone and Lasko, 2004; Wreden et al., 1997). Furthermore, identification of germ cell-specific isoforms of eIF4E and eIF4G suggests selective translation of mRNAs in the germline (Ghosh and Lasko, 2015; Hernandez et al., 2006).

During gastrulation germ cells start a passive migration mediated by posterior midgut invagination. At late stage 10, concomitant with germ band retraction, germ cells actively cross the midgut endoderm to find their way to the gonadal mesoderm, and by stage 13-14 germ cells condense with the somatic gonad cells on each side of embryo (Richardson and Lehmann, 2010). Transepithelial migration depends on transient intercellular gaps between epithelial cells, and activation of G-protein coupled receptor, *trapped in endoderm-1 (tre-1)*, on germ cells, which triggers cytoskeletal remodeling and mediates extension of pseudopodia (Jaglarz and Howard, 1995; Kunwar et al., 2003).

Germ cells exit G2 mitotic phase arrest by entering gonads, concurrent with an increase in their transcription (Su et al., 1998; Zalokar, 1976). At the same time germ cells start their sex-specific differentiation both as a result of internal cues, such as expression of Sex lethal (SXL) in females, and signals from surrounding somatic cells (Hashiyama et al., 2011). In *Drosophila* meiosis begins around the time of pupation; males produce sperms by the time of eclosion and females produce eggs a few hours after (Lesch and Page, 2012).

For the purpose of this thesis a summary of *Drosophila* oogenesis is presented. Details of spermatogenesis in *Drosophila* are reviewed by Fabian and Brill (2012). In adult females each ovary is composed of several ovarioles, which are considered as egg production units. At the anterior tip of each ovariole there is a morphologically distinct region called germarium, accommodating germline stem cells (GSCs). Asymmetric divisions of these GSCs are regulated by somatic cells present in the niche, and result in two daughter cells, one of which is another stem cell and the second differentiates into a cystoblast (CB). Each CB undergoes four mitotic divisions to produce a cyst of 16 interconnected germ cells, in which one cell differentiates into oocye and the others become nurse cells. Germline cysts that leave the germarium at this stage are called egg chambers. Meiosis in ovaries starts in the germline cysts, where two cells among 16 start meiosis but only one differentiates into an oocyte and remains in prophase I until late oogenesis. The oocyte then proceeds to metaphase I, where it is arrested again until fertilization (Ables, 2015; Lehmann, 2012). Egg chamber polarity, which includes posterior position of oocyte and differentiation of epithelial follicle cells, is established via reciprocal interactions between germline and somatic follicle cells (Assa-Kunik et al., 2007; Roth and Lynch, 2009; Torres et al., 2003). The oocyte itself is also polarized through a process that originally depends

on the cyst structure, and becomes more pronounced as the egg chambers develop. Oocyte polarity is internally linked to reposition of microtubule organizing center (Grieder et al., 2000; Steinhauer and Kalderon, 2006; Theurkauf et al., 1992). Anterior posterior (AP) axis establishment in the oocyte is associated with asymmetric localization of key mRNAs, such as *bicoid* (*bcd*) and *oskar* (*osk*) to the anterior and posterior poles, respectively (Berleth et al., 1988; Ephrussi et al., 1991). Similarly dorsal ventral (DV) axis is determined by nucleus position and, tightly associated with that, mRNA localization of a TGF-α-like protein, Gurken (Grk), in the future dorsal anterior corner (Neuman-Silberberg and Schupbach, 1993). Oocyte patterning determines polarization of the overlaying epithelial follicle cells, which is critical for proper development of egg shell structures such as respiratory appendages and the micropyle in the anterior-dorsal side of the egg (Berg, 2005). During stages 10B and 11 nurse cells transport their cytoplasmic content to the oocyte and undergo apoptosis (Cavaliere et al., 1998). Mature oocyte (stage 14) becomes activated and completes meiosis as it passes through the oviduct (Horner and Wolfner, 2008; Mahowald et al., 1983).

1.4 Structure and components of pole plasm in *Drosophila*

The critical role of pole plasm in *Drosophila* germ cell formation was first demonstrated by UV-irradiating the posterior pole of the fertilized eggs, which resulted in sterile flies (Geigy, 1931). Germ cell formation was restored in these embryos by transplantation of a new pole plasm (Okada et al., 1974). In addition, it is shown that ectopic transplantation of pole plasm results in ectopic formation of pole cells, suggesting that posterior cytoplasm contains all the necessary components for germ cell formation (Illmensee and Mahowald, 1974). Ultra-structural analyses show that the main components of germ plasm, the polar granules, first appearing in the posterior region at stage 9, are composed of a non-membrane bound fibrous meshwork. These structures become associated with the surface of mitochondria in the older oocytes, and the ribosomes in the activated egg (Mahowald, 1968; Mahowald et al., 1983).

The full composition of polar granules is not clear but several mRNA or proteins, including *osk*, *nos*, Tudor (Tud), Vas, *gcl* and *pgc* are found localized to the pole plasm, and play important

roles in pole plasm assembly, germ cell formation and/or posterior abdominal segmentation. Some of these components will be discussed in the next sections.

1.4.1 *oskar*

osk mRNA is synthesized in the nurse cells and is transported into the oocyte via a dynein-mediated movement toward the minus ends of microtubules (MTs), which emanate from nurse cells to the oocyte. This process depends on two proteins, Bicaudal D (BicD) and Egalitarian (Egl), which couple mRNA cargoes to dynein motor proteins (Bullock and Ish-Horowicz, 2001; Clark et al., 2007). Stable localization of *osk* to the posterior pole, which is observed from stage 8 onward, depends on reorganization of MTs in mid-oogenesis (Kim-Ha et al., 1991; St Johnston, 2005). Imaging *osk* mRNA particles in live oocytes indicates an active, but random, walk to all directions along a weakly polarized MT network, which nevertheless results in a robust localization to the posterior region (Parton et al., 2011; Zimyanin et al., 2008). Several trans-acting proteins have been implicated in *osk* localization. The double stranded (dsRNA)-binding protein, Staufen (Stau), colocalizes with *osk* and is required for its accumulation in the pole plasm by a proposed mechanism that depends on the interaction between Stau and Kinesin (Kim-Ha et al., 1991; St Johnston et al., 1991). Cytoplasmic localization of *osk* in oocyte depends on its mRNA splicing in the nurse cell nucleus (Hachet and Ephrussi, 2004), and its associations with core proteins of exon junction complex (EJC), such as Mago Nashi (Mago) and Barentsz (Btz) (Mohr et al., 2001; van Eeden et al., 2001).

Spatio-temporal restriction of Osk in oocyte not only requires mRNA localization, but also strictly depends on its translational control. Unlocalized *osk* mRNA is translationally repressed to avoid ectopic expression of Osk and its detrimental consequences for embryonic patterning. Mechanisms of translational regulation of Osk are discussed in section 1.6.2.

Osk is required for localization of all other pole plasm components, including Vas, Tud and *nos* (Mahowald, 2001). A hybrid RNA, in which the 3′ UTR of *osk* is replaced with that from becoid (*osk-bcd 3′ UTR*), localizes to the anterior region of oocyte, and induces ectopic germ cell formation, indicating that *osk* is the most upstream components in pole plasm assembly (Ephrussi and Lehmann, 1992). However, when *osk-bcd 3′ UTR* was expressed in mutant backgrounds of *vas* or *tud* it failed to induce ectopic germ cells, showing that *vas* and *tud* are also

required for pole plasm assembly. A recent crystal structure analysis of Osk together with genetic and biochemical assays provide evidence that Osk contains RNA binding domains, and point to the possible roles of Osk in mRNA localization, stability and translation (Yang et al., 2015b). *osk* mRNA is translated as two isoforms, short and long Osk (Markussen et al., 1995). While short Osk recruits components of the pole plasm, including Vas, the long isoform is required for anchoring of these factors, including Osk itself, to the posterior cortex (Vanzo and Ephrussi, 2002). The latter depends on reorganization of actin cytoskeleton by long Osk via recruitment of endosomal proteins, and thereby stimulating endocytosis (Babu et al., 2004; Jankovics et al., 2002; Tanaka et al., 2011).

1.4.2 *nanos*

nanos (*nos*)-related genes have been identified in the germline of many vertebrate and invertebrate organisms (Hashimoto et al., 2008; Jaruzelska et al., 2003; Koprunner et al., 2001; Mochizuki et al., 2000; Mosquera et al., 1993; Subramaniam and Seydoux, 1999). In *Drosophila* ovaries *nos* is expressed in GSCs and cystoblasts (CBs), and is also localized to the posterior region of the oocyte, where it is translated and incorporated into the germ cells. Continued expression of *nos* is observed throughout embryogenesis (Moore et al., 1998; Wang et al., 1994). Nos, together with another RNA-binding protein Pum, blocks translation of the maternal mRNA, *hunchback* (*hb*), thereby creating a gradient of Hb protein across the egg, which opposes the steep posterior-anterior gradient of Nos (Murata and Wharton, 1995; Tautz, 1988). Hb acts as a transcriptional activator of the genes essential for development of anterior structures; thus, its repression in the posterior region is critical for abdominal segmentation (Lehmann and Nüsslein-Volhard, 1987). In addition to their role in abdominal patterning, Nos and Pum play roles in germ cell maintenance and migration (Deshpande et al., 1999; Hayashi et al., 2004; Wang and Lin, 2004). Germ cells cease mitosis shortly after their formation, being released from this quiescence only after their arrival in presumptive gonads. This exit from mitosis is governed, at least partly, by Pum- and Nos-dependent repression of *CycB* mRNA (Asaoka-Taguchi et al., 1999). Unlike regulation of *hb* translation, where Nos plays its role by recruiting NHL-domain protein, Brain tumor (Brat), translational repression of *CycB* is independent of Brat and requires CCR4-NOT deadenylase complex, recruited by Nos (Kadyrova et al., 2007). Nos is also required for GSC self-renewal in a cell autonomous manner (Wang and Lin, 2004). As a result in *nos*

mutants daughter cells from a GSC division both differentiate to germline cysts. This function of Nos is likely through translational repression of unknown differentiation factors.

In the oocyte *nos* mRNA localization and translation are coupled to confine accumulation of Nos protein to the posterior (Kugler and Lasko, 2009). *nos* mRNA is synthesized in the nurse cells and is transferred to the oocyte starting in early oogenesis (stages 4-5); its posterior localization however is detectable from stage 10 (Wang et al., 1994). *In vivo* studies show that posterior localization of *nos* is independent of MTs or cytoplasmic streaming, and is accomplished through a diffusion/entrapment mechanism requiring localization of other pole plasm components such as Osk and Vas (Forrest and Gavis, 2003). Actin-dependent anchoring then maintains *nos* RNA complexes in the posterior pole. Furthermore, degradation of *nos* mRNA in the bulk cytoplasm is promoted by CCR4-NOT complex recruited to the 3′ UTR of *nos* by Smaug, an RNA binding protein, which is also involved in translational repression of *nos* (Zaessinger et al., 2006; section 1.6.3). Degradation of maternal mRNAs such as *nos* during maternal to zygotic transition through CCR4-mediated deadenylation depends on the activity of Piwi-interacting RNA (piRNA) pathway, otherwise better known for its function in transposon silencing (Rouget et al., 2010). *nos* mRNA in its 3′ UTR contains complementary sequences to specific piRNAs, and its stabilization in mutants of *aubergine* (*aub*) and *spindle-E* (*spn-E*), two piRNA pathway components, results in head development defects. With all these mechanisms only 4% of *nos* mRNA is localized to the posterior; the vast majority is still diffused in the cytoplasm (Bergsten and Gavis, 1999), suggesting a significant contribution from translational regulation in spatial restriction of Nos. The experimental evidences for regulation of *nos* translation are discussed in section 1.6.3.

1.4.3 Tudor

tudor (*tud*) gene was first discovered in *Drosophila* in a screen for maternal factors that affect embryonic viability or fertility (Boswell and Mahowald, 1985). Tud is mostly dispensable for oogenesis and posterior patterning of embryos, but the progeny from females homozygous for a null allele of *tud* do not produce germ cells. Thus *tud*-null mutants are categorized as "grandchildless" (Thomson and Lasko, 2004). Tud is a large protein containing 11 repeats of a domain called Tudor domain (Ponting, 1997). Tudor domain binds methylated arginines or lysines in target proteins, thereby mediating assembly of macromolecular complexes (Chen et

al., 2011). Some Tudor domain proteins merely act as adaptors to connect effector proteins to their substrates, but the majority also contain catalytic domains and could play the effector role. Many Tudor domain proteins are involved in RNA metabolism pathways, including RNA splicing or small RNA processing, through RNA binding domains of themselves or their arginine methylated partners. In addition, a number of Tudor domain proteins regulate histone modifications or DNA damage responses (Pek et al., 2012a).

In *tud*-null oocytes initial localization of *osk* and *nos* mRNAs and Vas protein to the pole plasm is not affected although this localization diminishes in later stages (Thomson and Lasko, 2004). On the other hand localization of *gcl* and *pgc*, the two genes that are specifically involved in germ cell development (section 1.4.5), is undetectable in these mutants, further supporting the role of Tud in germ cell formation pathway. Electron micrographs indicate that in *tud* mutants polar granules are significantly reduced in size and number, consistent with the failure to form germ cells (Thomson and Lasko, 2004). It remains unknown if Tud function involves direct binding to RNA, although sequence comparisons do not suggest a well-described RNA binding motif in Tud (Thomson and Lasko, 2004). Tud binding to Aub, the piRNA component, is required for localization of Aub to the pole plasm (Kirino et al., 2010b). Tud specifically targets symmetrically dimethylated arginines (sDMAs) of Aub, in an interaction that is regulated by *Drosophila* arginine methyltransferase 5 (dRMT5/Capsuleen) and its co-factor Valois (Vls), both required for germ cell formation (Anne and Mechler, 2005; Anne et al., 2007; Kirino et al., 2010b).

Tud and several other Tudor domain proteins including Tejas (Tej), with a dsRNA-binding domain; Spn-E, a DExH-box helicase; and Kumo/Qin, an E3 ligase, localize to the perinuclear region of the nurse cells, the nuage (Anand and Kai, 2012; Findley et al., 2003; Patil and Kai, 2010). The perinuclear nuage is found in the germline of numerous animals, and is known as the site for piRNA processing (Lim and Kai, 2007; see section 1.7.1).

1.4.4 Vasa

The first *vas* mutant alleles were found in screens for maternal-effect mutations that impair *embryonic* patterning or germ cell formation (Schupbach and Wieschaus, 1986a, b). The *vas* gene was subsequently cloned, and its sequence was found similar to the DEAD-box RNA

helicase, eukaryotic initiation factor-4A (eIF4A) (Hay et al., 1988b; Lasko and Ashburner, 1988). Immunostaining for Vas indicates that it localizes to the pole plasm and is present in the pole cells and their germ cell derivatives in all stages of development (Hay et al., 1990; Lasko and Ashburner, 1990). In the genetic hierarchy of pole plasm formation, *vas* is located downstream of *osk* and *stau*, as mutants of these two genes do not display a posterior localization of Vas. However, initial Vas localization is unaffected in mutants of most other pole plasm components, including *nos*, *tud*, *valois* (*vls)*, *gcl* and *pgc* (Jongens et al., 1992; Lasko and Ashburner, 1990; Nakamura et al., 1996). Homozygous females for *vas* hypomorphic alleles produce embryos which lack polar granules, fail to form germ cells and exhibit defects in abdominal segmentation (Schupbach and Wieschaus, 1986b). *vas*-null females have aberrant ovarioles with the majority of oocytes not reaching vitellogenic stage, indicating additional roles for *vas* in oogenesis (Styhler et al., 1998; see sections 1.6.5, 1.7.2, 1.8). On the contrary, *vas* mutant males are fertile, indicating that Vas is dispensable for spermatogenesis (Lasko and Ashburner, 1988; Renault, 2012). Extensive studies have provided significant insights into Vas functions; yet the exact role of Vas in pole plasm is not well understood, confounded by the fact that *vas* mutations abolish pole plasm assembly altogether, thus providing little information about specific Vas interactions. RNA binding is not required for Vas localization to the pole plasm (Liang et al., 1994); however, this activity might mediate mRNA anchoring of the other components to the posterior region. In *vas* mutants, consistent with defects in abdominal segmentation, posterior localization of *nos* mRNA is abolished (Gavis and Lehmann, 1992); however, there is no direct evidence that Vas binds specifically to *nos* mRNA. Vas is an essential component of polar granules, which are enriched for many germline-specific RNAs (Hay et al., 1988b). Although the functional nature of polar granules remains elusive, presence of RNA-binding proteins, such as Vas, in polar granule structure suggests a role for them in posttranscriptional regulation of RNA. In addition, strong evidence supports Vas function in translational activation (section 1.6.5, 1.12), outside of pole plasm. By a similar mechanism Vas may derepress translation of proteins, such as Nos or Osk, in the pole plasm, for the latter perhaps after its initial translation through another mechanism.

Vas localization is not limited to pole plasm, and is detected through all stages of oogenesis. Vas accumulates in the GSCs and CBs in the germarium, becomes localized to the perinuclear region of nurse cells soon after their specification in the cysts, and is transported from nurse cells

to the oocyte from stage 8 onward. Posterior localization of Vas becomes detectable at stage 10a (Lasko and Ashburner, 1990). Live imaging of GFP-tagged Vas and Aub indicates that nuage particles become displaced to the cytoplasm of nurse cells and then enter the oocyte through ring canals (Snee and Macdonald, 2004). This study argues against the previous model suggesting that pole plasm is directly formed from the nuage fragments, by providing evidence that Vas becomes dissociated from these precursors and is later entrapped in the pole plasm, after its *de novo* nucleation by Osk. The direct interaction between Vas and Osk is essential for pole plasm assembly, as long Osk, which has significantly lower affinity to Vas compared to short Osk, is not able to induce ectopic Vas localization and germ cell formation in the anterior region (Breitwieser et al., 1996). Furthermore, persistent trafficking of pole plasm components, such as Vas, on cortical microtubules, mediated by dynein, is required for their retention at the posterior cortex (Sinsimer et al., 2013).

Vas can be co-purified from *Drosophila* ovaries and embryos with the ubiquitin-specific protease Fat facets (Faf), which also localizes to the pole plasm (Huang et al., 1995; Liu et al., 2003). Embryos from *faf* mutant females form fewer pole cells than wild-type, consistent with decreased levels of Vas and its posterior localization in these mutants. In addition, Vas purified from *faf* ovaries is highly ubiquitinated, suggesting that Vas is stabilized in pole plasm through its Faf-dependent deubiquitination. Vas ubiqutination in pole plasm is also modulated by the two ubiquitin Cullin-RING E3 ligase specificity receptors, Fsn and Gustavus (Gus), which both bind to the same motif (DINNN) on Vas (Kugler et al., 2010; Styhler et al., 2002). *fsn* mutants show a precocious accumulation of Vas in pole plasm, whereas *gus* mutants additionally lacking one copy of *cullin-5,* exhibit delay in posterior localization of Vas. While Fsn clearly acts as a negative regulator of Vas stability, the function of Gus is more complex, as its over-expression also reduces levels of Vas in the ovaries.

1.4.5 *germ-cell less* and *polar granule component*

germ-cell less (*gcl*) was first identified in a screen for grandchildless mutants (Jongens et al., 1992). Both *gcl* transcript and protein were found localized to the posterior pole depending on localization of other germ cell determinants, such as *osk*, Vas and Tud. Gcl protein primarily localizes to the nuclear envelope in germ cell precursors, and is often associated with nuclear pore complex (NPC) (Jongens et al., 1994). Analysis of a null allele (gcl^{Δ}) indicates that

maternal *gcl* is required for germ cell formation but not for embryonic patterning or viability (Robertson et al., 1999). Germ cell defects start to appear at the time syncytial nuclei enter the germ plasm to form the pole buds, and continue to increase until stage 14, the time pole cells reach the embryonic gonad, resulting in only 32% of gcl^{Δ} adults being fertile. Staining with H5 antibody, which recognizes a phosphorylated form of RNA polymerase II associated with active transcription, indicates that most pole cells in gcl^{Δ} fail to establish and maintain transcriptional quiescence (Leatherman et al., 2002). Transcriptional repression by Gcl, which could be also induced ectopically, is specific to a subset of genes rather than being global. Furthermore, Gcl is implicated in a spindle-independent cleavage pathway that results in separation of pole buds from embryo (Cinalli and Lehmann, 2013). Mis-expression of Gcl together with the contractile ring components, Anillin, but not alone in the anterior region promotes ectopic PGC-like cell formation, suggesting that the presence of these two proteins is the minimum requirement for inducing the pole bud furrow.

polar granule component (*pgc*), which was first studied based on its specific localization to polar granules, was originally considered a noncoding RNA (Nakamura et al., 1996), but was later shown to encode a small protein (Hanyu-Nakamura et al., 2008). Through expression of an antisense *pgc* RNA it was shown that *pgc* activity is required for germ cells to migrate and colonize the gonads. In mutants with complete loss of *pgc* expression, germ cells ectopically transcribe genes, such as *zerknullt* (*zen*) and *tailless* (*tll*), that in *wt* embryos are only expressed in the somatic cells at the posterior region (Martinho et al., 2004). Consistent with this, RNA pol II CTD phosphorylation on Ser2, and methylation of K4 in histone H3, which are both associated with active transcription, are increased in *pgc* mutant germ cells (Kouzarides, 2002; Martinho et al., 2004; Seydoux and Dunn, 1997). The 71-amino-acid Pgc polypeptide, like its RNA, is specifically detected in pole cells (Hanyu-Nakamura et al., 2008). Supporting the previously identified role of *pgc* in transcriptional quiescence of germ cells, Pgc protein was found to directly interact with CDK9, a component of Positive transcription elongation factor-b (P-TEFb), which is responsible for phosphorylation of CTD Ser2 (Hanyu-Nakamura et al., 2008; Peterlin and Price, 2006). Pgc did not have any effect on CDK9 kinase activity when tested *in vitro*, however when ectopically expressed in salivary glands it prevented recruitment of P-TEFb to the polytene chromosomes, which was associated with ~40% reduction in the level of

phosphorylated Ser 2. Together these results suggest that in germ cells Pgc sequesters P-TEFb to prevent its recruitment to active promoters (Hanyu-Nakamura et al., 2008).

1.5 DEAD-box family of RNA helicases

The DEAD-box protein family represents the largest group of helicases (Linder and Fuller-Pace, 2013). These proteins unwind dsRNAs in an unprocessive manner, an activity requiring ATP-binding and -hydrolysis.

1.5.1 The conserved helicase core

The highly conserved core is composed of two recombinase A (RecA)-like domains and contains 12 characteristic motifs, which are involved in RNA binding, ATP binding and inter-domain interactions (Fig. 1.2 A). The name of this family originates from motif II, the Asp-Glu-Ala-Asp (D-E-A-D) motif. A similar mechanism of duplex separation is utilized by all DEAD-box proteins, whereby concomitant with ATP-binding the helicase core undergoes a conformation change, closing the cleft between the two RecA domains (Chen et al., 2008; Linder and Fuller-Pace, 2013; Rudolph and Klostermeier, 2015). The close state imposes a sharp bend in the RNA strand, bound opposite to the ATP-binding site, and causes local duplex destabilization and RNA unwinding. ATP hydrolysis occurs subsequently to restore open conformation and release RNA before the helicase could start another catalytic cycle (Fig. 1.2 B).

Motifs Q, I (also known as the Walker A motif), II (also known as the Walker B motif) and VI are involved in ATP binding and hydrolysis; motifs 1a, 1b, 1c, IV, IVa and V mainly contain residues, which interact with the sugar phosphate backbone of RNA; and motifs III and Va mediate inter-domain interactions which couple ATP binding with RNA unwinding (Cordin et al., 2006; Linder and Jankowsky, 2011; Sengoku et al., 2006). These conserved motifs, however, are often not restricted to one function and contain residues participating in different interactions; for example, motif Ia, in addition to its well characterized role in RNA binding, is involved in structural rearrangements upon ATP-binding (Schwer and Meszaros, 2000; Sengoku et al., 2006).

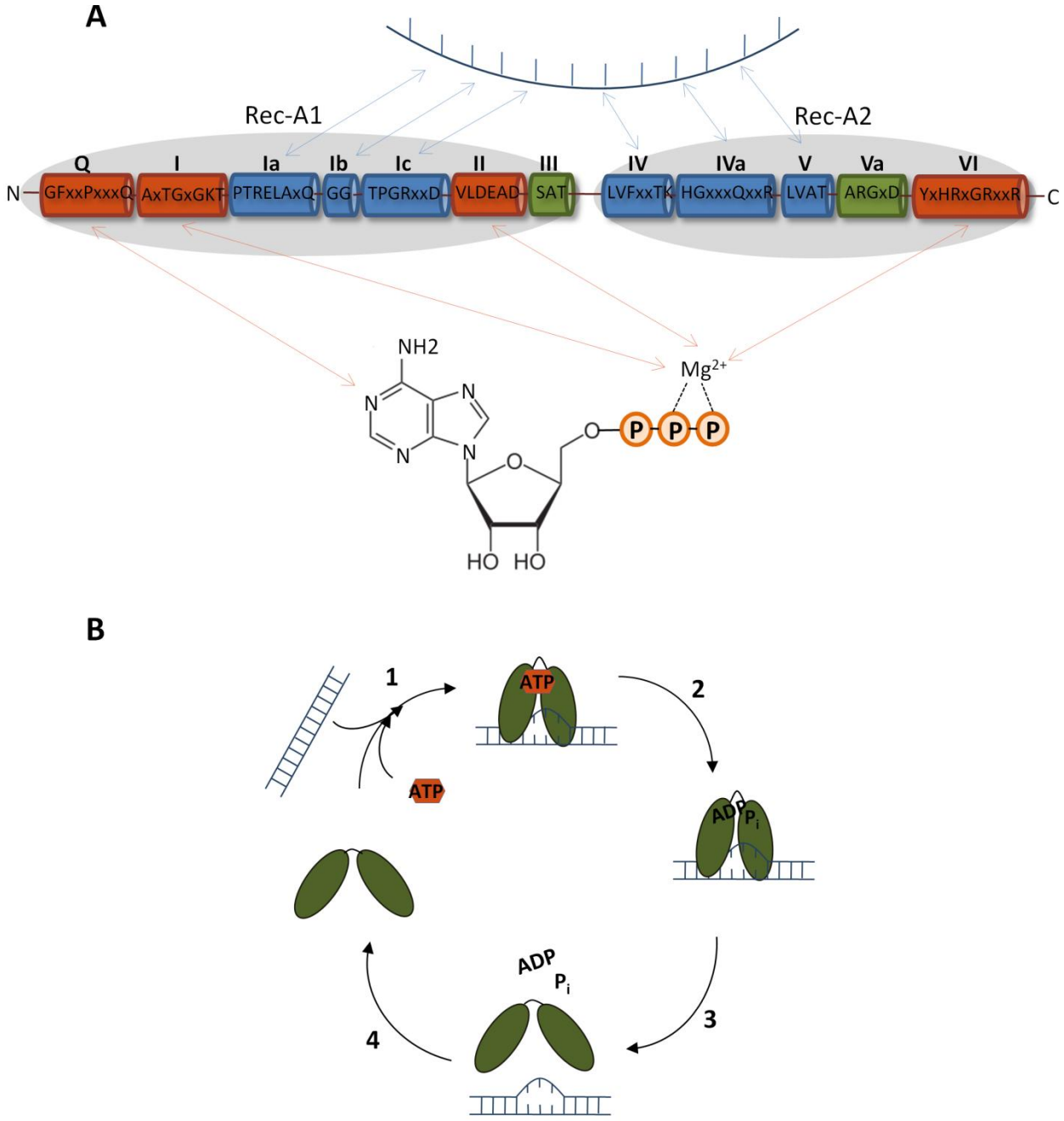


Fig. 1.2 Conserved structure and mechanism of function of DEAD-box proteins. **A**. DEAD-box proteins share twelve conserved motifs in their core region, which are involved in RNA-binding (blue), ATP-binding (red) and inter-domain communications (green). These motifs are arranged within two Rec-A domains (Linder and Jankowsky, 2011). **B**. In DEAD-box proteins ATP-binding, which occurs concomitant with RNA-binding, induces a closed conformation in the protein, and results in RNA duplex separation (step 1). ATP hydrolysis and product (ADP and P_i) release (steps 2, 3) are subsequently required for dissociation of unwound RNA and to reset the enzyme for further catalytic cycles (steps 4) (Chen et al., 2008).

1.5.2 Diverse flanking regions

eIF4A, the prototypic member of DEAD-box protein family, is one of the few examples that are only composed of a minimum helicase core domain (Linder et al., 1989). However, for a vast majority of DEAD-box proteins the conserved core is flanked by N- and C-terminal sequences (Hilbert et al., 2009; Rudolph and Klostermeier, 2015). Our knowledge of the core domain and the mechanism of RNA unwinding has substantially increased by high resolution structure analyses. Yet, little is known about the role of N- and C-terminal flanking sequences in substrate recognition and catalytic activity, as very few studies have structurally analyzed full length proteins. Some insight is gained through models constructed from the crystal structure of overlapping fragments in conjunction with predicted homology structures. The *Bacillus subtilis* YxiN and its *Escherichia coli* homolog, DbpA, which are implicated in ribosome biogenesis, contain a C-terminal RNA-binding domain (RBD) (Diges and Uhlenbeck, 2001; Karginov et al., 2005). This auxiliary domain has a high affinity for hairpin 92 of ribosomal 23S rRNA, specifically through a GYYX stretch (Y = pyrimidine, X = purine/pyrimidine) (Hardin et al., 2010). The YxiN RBD presumably acts as an anchor to direct unwinding activity of the helicase core to the nearby duplexes and to tether the helicase on the RNA molecule during multiple cycles of unwinding. Another example is *Thermus thermophilus* DEAD-box protein Hera, which contains a dimerization domain (DD) and an RBD domain in its C-terminal region (Klostermeier and Rudolph, 2009; Morlang et al., 1999). The RBD domain specifically mediates binding to ribosomal RNA and RNAse P RNA through a GGXY stretch (Linden et al., 2008; Steimer et al., 2013). In addition, the C-terminal region contains a flexible short basic tail, which interacts with RNA duplexes, possibly mediating recruitment of the helicase core to these regions (Steimer et al., 2013). Similar C-terminal tails, which loosely target structured RNAs for a subsequent strong binding of the helicase core, have been found in two helicases involved in RNA splicing, Cyt-19 and Mss116 from *Neurospora crassa* and *Saccharomyces cerevisiae*, respectively (Grohman et al., 2007; Mallam et al., 2011; Mohr et al., 2008). Mss166 is an interesting example as it also contains a C-terminal extension (CTE) that forms a compact structural unit with RecA domain. In this format the unwound ssRNA emanating from helicase core faces a steric hindrance by the CTE that causes a second bend in the RNA, thereby increasing the unwinding activity (Del Campo and Lambowitz, 2009).

In addition to cis regulatory flanking regions, the catalytic activity of DEAD-box proteins can be modulated by their interacting partners in trans. A prominent example is regulation of eIF4A by eIF4G, which binds to both Rec-A domains of eIF4A and fixes it in a half-open conformation (Hilbert et al., 2011; Schuetz et al., 2008). The RNA-binding site is not properly aligned in this conformation, thus stimulating release of RNA. Another example is the interaction between eIF4A-III, and Btz/MLN51, two components of the exon junction complex (EJC) (Andersen et al., 2006). This interaction does not affect ATP hydrolysis but inhibits product dissociation, a necessary step for releasing RNA (Nielsen et al., 2009).

1.5.3 Cellular functions of DEAD-box proteins

DEAD-box proteins have been implicated in a wide range of RNA-related processes (Linder, 2010). Some of these cellular functions will be discussed in the next paragraphs.

mRNA splicing: Pre-mRNA splicing, which removes introns, entails two sequential transesterification reactions. The splicing events, in most cases, are accomplished by a spliceosome complex, composed of several small nuclear ribonucleoprotein (snRNP) particles (Clancy, 2008). The assembly of snRNP particles and the conformational changes in their protein and RNA components as well as in the mRNA molecule, coordinate chemical steps during splicing. This process utilizes ATP and is mediated by several members of the DEXD/H-box protein family (Schwer and Meszaros, 2000). UAP56 is a ubiquitous DEAD-box protein found in a wide range of organisms from yeast to human, and is involved in a series of pre-mRNA splicing steps including U2 snRNP-branchpoint interaction (Shen, 2009). The yeast DEAD-box protein, Prp28p, is required to destabilize the duplex between U1 snRNA and pre-mRNA 5′ splice site, thereby promoting the U1/U6 exchange in the first step of splicing (Chen et al., 2001). Furthermore, mRNA splicing is accompanied by the assembly of EJC upstream of spliced junctions, mediated by spliceosomes. EJC remains on mRNA during its transport to the cytoplasm, and recruits other transacting factors involved in post-transcriptional regulation (Le Hir et al., 2016). For example, EJC acts as a molecular signature of splicing, whose presence downstream of a premature stop codon marked by a terminating ribosome, would trigger nonsense-mediated mRNA decay (NMD). The DEAD-box protein, eIF4A-III, is a component of EJC and associates with the two other components, Y14 and Mago (Chan et al., 2004; Dostie and Dreyfuss, 2002). The exact function of eIF4A-III in EJC is not known; one possibility is that

consistent with its reported activity as a translational inhibitor, eIF4A-III prevents premature translation of mRNA in the nucleus or cytoplasm (Chan et al., 2004; Li et al., 1999).

mRNA export: Following post-transcriptional processes in the nucleus, the mature mRNA needs to be exported through the nuclear pore complex (NPC) into the cytoplasm. The conserved DEAD-box protein, Dbp5, binds the cytoplasmic NPC face, and through its function in RNP remodeling, displaces mRNA from Nab2, which shuttles mRNPs from nucleus to cytoplasm (Tran et al., 2007; Tseng et al., 1998). In addition, experiments in different organisms show that mutations in UAP56 result in retention of polyadenylated RNA in the nucleus and provide evidence that UAP56 couples mRNA splicing with nuclear export (Shen, 2009).

Translation initiation: eIF4A, the founding member of DEAD-box protein family, is a component of the cap-binding complex and is involved in translational regulation (Gingras et al., 1999; Rogers et al., 2002). The role of eIF4A in translation activation could be through unwinding secondary structures in the 5′ UTR, or displacing translation inhibitors, which otherwise prevent 40S subunit binding to mRNA (Poulin and Sonenberg, 2003). The *S. cerevisiae* protein Ded1, ortholog of *Drosophila* and human proteins Bel and DDX3, is another DEAD-box protein implicated in translation (Chuang et al., 1997; Johnstone et al., 2005; Lee et al., 2008). Ded1 further interacts with eIF4G to form a Ded1-mRNA-eIF4F complex, which accumulates in stress granules. Subsequent ATP hydrolysis by Ded1, acts as a switch to release mRNA for completing translation (Hilliker et al., 2011). The germline specific DEAD-box protein, Vas, interacts with eIF5B, a translation factor required for joining ribosomal subunits, and promotes translation of the EGFR ligand, Grk, in *Drosophila* ovaries (Carrera et al., 2000; Pestova et al., 2000; section 1.6.5)

mRNA decay: mRNA degradation is a well established pathway, which allows rapid response to changes in transcription. The DEAD-box RNA helicase, RhlB, is a component of RNA degradosome in *E. coli*, which through unwinding structured RNA fragments facilitates their degradation by polynucleotide phosphorylase (PNPase; Khemici et al., 2005). Another example is the yeast protein Dhh1, which accumulates in distinct cytoplasmic foci named processing bodies (P bodies) involved in mRNA decay and translational control (Jain and Parker, 2013). Dhh1 interacts with both decapping and deadenylation complexes, thereby it functions in a major pathway of mRNA turnover, in which initial shortening of poly-adenosine (poly-A) tail

is followed by decapping and 5′-3′ degradation (Coller et al., 2001; Teixeira and Parker, 2007). Dhh1 orthologs, Xp54 in *Xenopus*, Me31B in *Drosophila* and CGH-1 in *C. elegans*, are all components of the cytoplasmic particles containing maternal mRNAs in the oocyte, and are required for translational silencing of the mRNAs (Ladomery et al., 1997; Nakamura et al., 2001; Navarro et al., 2001; Weston and Sommerville, 2006).

Ribosome biogenesis: Three out of four rRNAs are produced as a long primary transcript. The rRNA processing includes sequential cleavage of spacers from nascent transcript and maturation of the rRNA 3′ ends (Henras et al., 2015). These events are catalyzed by numerous ribosomal proteins and small nucleolar RNAs (snoRNAs), which co-transcriptionally associate with the primary rRNA. Many DEAD-box proteins, including DbpA/YxiN in *E. coli*, Dbp8 in yeast and DDX5 (p68) in human have been implicated in ribosome assembly (Martin et al., 2013). It is proposed that duplex dissociation by DEAD-box helicases promotes rearrangements between rRNAs and snoRNAs, or displaces proteins that hamper rRNA processing.

Transcription: Several DEAD-box helicases have been more recently shown to play roles in transcription (Fuller-Pace and Ali, 2008). In human cell lines p68 and its analog, p72, directly interact with tumor suppressor, p53, and activate transcription of target genes in response to DNA damage (Bates et al., 2005). Similarly p68 is shown to synergize with the non-coding RNA, *SRA*, for activating transcription of muscle differentiation factor, MyoD (Caretti et al., 2006). In *Drosophila* p68 ortholog, Rm62, is involved in gene deactivation by clearing pre-existing transcripts from transcription sites and resetting chromatin to a fully inactive state (Buszczak and Spradling, 2006).

1.6 Translational regulation in *Drosophila* oogenesis

Post-transcriptional regulation of maternally expressed mRNAs is essential for many developmental events during *Drosophila* oogenesis and early embryogenesis (Lasko, 2012). Translational control coupled to mRNA localization governs spatial and temporal expression of several genes in the *Drosophila* oocyte. In this section I will first overview translation initiation in eukaryotes, as a step that is often subjected to regulatory mechanisms, and then discuss some of the best studied examples of translation regulation in *Drosophila* oogenesis.

1.6.1 Eukaryotic translation initiation

The goal of translation initiation is to assemble ribosomes loaded with the initiator methionyl-tRNA (Met-tRNAi) on the start (AUG) codon of the mRNA, which is generally identified by a scanning mechanism. Most translation in eukaryotes depends on the assembly of eukaryotic initiation factors (eIFs) on the cap structure, a guanine nucleotide methylated on the 7 position (m7G) at the 5′ end of mRNA (Merrick, 2004). A preassembled complex of 40S ribosomal subunit, Met-tRNAi and eIFs 1, 1A, 2, 3, and 5, named preinitiation complex (PIC) is recruited to the capped mRNA through eIF4F complex, which is composed of cap-binding factor, eIF4E, the scaffolding protein, eIF4G, and the RNA helicase, eIF4A. PIC scans 5′ UTR for an AUG codon; thus secondary structures that prevent ribosomes from accessing 5′ UTR often act as the initiation inhibitors (Aitken and Lorsch, 2012; Sonenberg and Hinnebusch, 2009). Once the start codon is recognized a conformational change in PIC, following the release of eIF1 and conversion of eIF2 to its GDP-bound state, stabilizes PIC on the mRNA. At this step eIF2 and eIF5 are ejected while eIF5B is recruited to mediate joining of the 60S ribosomal subunit. GTPase activity of eIF5B causes a conformational rearrangement in the ribosome, which releases eIFs and starts the elongation phase.

In addition to the 5′ UTR, the specific sequences in the 3′ UTR of mRNA could affect translational efficiency. According to the "closed loop" model the 3′ poly(A)-binding protein (PABP) interacts with eIF4G, thereby circularizing mRNA and mediating the regulatory impact of 3′ UTR associated proteins on translation initiation (Sonenberg and Hinnebusch, 2009; Wells et al., 1998). The PABP-interacting proteins, Paipl and Paip2, could compete with eIF4G and thus modulate translation (Craig et al., 1998; Derry et al., 2006). Several advantages have been envisaged for the cap-to-tail circularization of mRNA. For example, the interaction with PABP could increase eIF4F affinity to 5′cap (Borman et al., 2000). The closed loop mechanism could also facilitate re-initiation of translation by polysomal ribosomes that terminate at the same mRNA (Kopeina et al., 2008), and could further ensure that only full length mRNA with intact cap and poly(A) tail are robustly translated (Sachs, 2000).

1.6.2 Oskar translation

In *Drosophila* Osk translation is restricted to the pole plasm, while it is repressed in the bulk of oocyte or in the nurse cells. Several mechanisms have been described for regulating Osk translation at different stages (Wilhelm and Smibert, 2005). An ovarian RNA-binding protein, Bruno (Bru) targets the 3′ UTR of *osk* through repeated Bru response elements, BREs (Kim-Ha et al., 1995). Bru interacts with Cup, another ovarian protein, which competes with eIF4G for its interaction with the 5′-cap binding protein eIF4E, therefore blocking initiation of Osk translation (Nakamura et al., 2004). Furthermore, loss of Me31B, a DEAD-box protein, leads to premature translation of Osk in the nurse cells and oocyte (Nakamura et al., 2001). Me31B associates with RNP complexes that contain Exuperantia (Exu), a protein involved in several steps of mRNA localization, suggesting that transcripts are translationally silenced during their transport to the oocyte (Nakamura et al., 2001; Wang and Hazelrigg, 1994; Wilhelm et al., 2000). Consistent with this notion, the RNA-binding protein, Hrp48, binds to both 5′ and 3′ UTRs of *osk* mRNA, without affecting Bru binding, and bifunctionally mediates localization and translational repression (Gunkel et al., 1998; Yano et al., 2004). Premature translation of Osk is also observed in *Bicaudal-C* (*Bic-C*) mutant oocytes (Saffman et al., 1998). Bic-C is a maternally inherited RNA-binding protein, which associates with CCR4-NOT deadenylation complex. Through this interaction Bic-C represses its mRNA targets, which are mostly involved in embryonic patterning, by controlling their poly(A)-tail length (Chicoine et al., 2007).

Little is known about how Osk translation is derepressed or activated in the pole plasm. The *Drosophila* homolog of *Xenopus* cytoplasmic polyadenylation element binding (CPEB), Orb, binds to the 3' UTR of *osk* mRNA and is shown to activate its translation (Chang et al., 1999). Undetectable levels of Osk protein in *orb* mutants correspond to decreased length of poly(A) tail in *osk* mRNA. Orb is thought to function in the same manner as CPEB, which binds to the cytoplasmic polyadenylation element (CPE) and “masks” its mRNA targets, until CPEB becomes activated during egg maturation to promote polyadenylation-dependent translation (Hake and Richter, 1994). In addition, a recent study shows that Bru phosphorylation by protein kinase A (PKA) prevents its dimerization required for binding to Cup, and thus positively regulates Osk translation (Kim et al., 2015). Stau, which associates with *osk* mRNA in the nurse cells and is required for its localization to the posterior of the oocyte, is also involved in Osk

translation (Micklem et al., 2000). However these two functions might be through different mechanisms as they depend on distinct motifs in the Stau sequence. Furthermore, levels of Osk protein are reduced in *aub* mutants, with no significant decrease in mRNA or protein stability, suggesting that *aub* also enhances Osk translation with a mechanism which remains unknown (Malone et al., 2009; Wilson et al., 1996). There is also evidence that Bru interacts with Vas, raising the possibility of Vas being involved in translational activation of Osk in the pole plasm (Webster et al., 1997).

1.6.3 Nanos translation

Similar to Osk, translation of Nos is repressed outside of the oocyte posterior region through elaborate mechanisms. The RNA-binding protein, Smg, which is evenly diffused in the oocyte, binds two cis acting elements in the 3′ UTR of *nos*, known as Smaug recognition elements (SREs; Smibert et al., 1996). In a similar manner to Bru, Smg interacts with Cup and represses translation of Nos (Nelson et al., 2004). The translational barrier is removed in the pole plasm by factors, such as Osk, which prevent Smg binding to SREs. The translational control element (TCE) in 3′ UTR of *nos* has a bipartite structure, which not only contains SREs in the loop of stem-loop II but also other elements in the stem of stem-loop III. The latter interact with Glorund (Glo), which belongs to the heterogeneous nuclear ribonucleoprotein (hnRNP) family (Kalifa et al., 2006). *glo* mRNA and protein are present in the nucleus and cytoplasm of nurse cells and epithelial follicle cells, and also at a lower level in the oocyte. Analyses of a null mutant indicate that Glo acts to repress translation of *nos* outside of pole plasm in late oocytes, similar to the Smg-mediated suppression in early embryogenesis. Furthermore, regulation of Nos stability and translation through CCR4-mediated deadenylation depends on the activity of piRNA components such as *aub* and specific sequences in the 3′ UTR of *nos* (Rouget et al. 2010; section 1.4.2).

1.6.4 Translation of Caudal and Hunchback

Translational repression of *caudal* (*cad*) and *hunchback* (*hb*) in the anterior and posterior regions, respectively, are crucial for embryonic patterning. Cad is a homeodomain transcription factor and a key regulator of the *Hox* genes in *Drosophila* (Hoey et al., 1986). Mutations that alter posterior-to-anterior gradient of Cad result in abnormal zygotic expression of several

segmentation genes including the pair rule gene, *fushi tarazu* (*ftz*), and thus are lethal (Dearolf et al., 1989; Macdonald and Struhl, 1986). The Cad gradient is established from uniformly distributed maternal mRNA through translational repression by Bicoid (Bcd), whose gradient in the oocyte is a mirror image of Cad gradient (RiveraPomar et al., 1996). Bcd, as originally described, acts as a transcriptional activator of segmentation genes, and in addition to that, binds a specific element in the 3′ UTR of *cad* mRNA, thereby inhibiting its translation in a cap-dependent manner (Chan and Struhl, 1997; Driever and Nüsslein-Volhard, 1989; Niessing et al., 1999; Struhl et al., 1989). Anterior localization of *bcd* mRNA depends on a cis regulatory element in its 3′ UTR, and trans-acting factors such as Exu and the dsRNA-binding proteins, Stau and Swallow (Swa) (Macdonald and Struhl, 1988; St Johnston, 2005; Weil et al., 2010). Bcd interacts with *Drosophila* eIF4E-homologous protein (d4EHP), which directly binds to 5′-cap of *cad* mRNA and blocks formation of the cap-binding complex (Cho et al., 2006).

An anterior posterior gradient of the transcription activator, Hb, is established through translational repression, exerted by cooperation between Nos, Pum and Brat (section 1.4.2). A recent study shows that Brat could also directly bind *hb* mRNA and regulate its translation in a Pum-independent manner (Loedige et al., 2014). In addition to mRNA deadenylation, mediated through Pum activity, a cap-dependent mechanism has been identified for inactivation of HB translation (Chagnovich and Lehmann, 2001; Cho et al., 2006). This mechanism involves d4EHP, which simultaneously binds Brat and the 5' cap of *hb* mRNA, to regulate HB translation through a similar mechanism establishing Cad gradient.

1.6.5 Gurken translation: role of Vas as a translational activator

The EGFR ligand, Grk, is secreted by the oocyte and induces dorsal fate in the overlying epithelial follicle cells, through its interaction with the EGFR receptor, Torpedo (Top) (González-Reyes et al., 1995). During oogenesis *grk* mRNA is expressed by the nurse cells and transported to the oocyte via a mechanism depending on dynein and its cofactors, BicD and Egl (Mach and Lehmann, 1997; NeumanSilberberg and Schupbach, 1996). Throughout oogenesis *grk* mRNA remains tightly associated with the oocyte nucleus, which in early stages is positioned at the posterior. Starting at stage 8, growing microtubules push the nucleus to the anterodorsal corner of the oocyte, and accordingly *grk* mRNA forms a crescent between nucleus and cell cortex in this region (Steinhauer and Kalderon, 2006; Zhao et al., 2012). The exact

mechanisms of *grk* localization are unknown, but this process is mediated at different stages by dispersed elements in the 5′ UTR, coding region and 3′ UTR of *grk* transcript (Thio et al., 2000).

Genetic interactions suggest that translation of unlocalized *grk* mRNA is inhibited through a complex including Cup and Squid (Sqd), and that another component of this complex, PABP55B, through its interaction with Encore (Enc), facilitates translation of localized *grk* by a yet unknown mechanism (Clouse et al., 2008). *grk* mRNA is not efficiently translated in *orb* mutant ovaries, suggesting that polyadenylation mediates translation of localized *grk*, a mechanism through which PABP55B is also proposed to exert its regulatory effect on Grk translation (Chang et al., 2001; Clouse et al., 2008). Furthermore, Grk translation is severely reduced in *vas*-null mutants, whereas mRNA localization remains normal (Styhler et al., 1998). A direct interaction between Vas and the translation factor, eIF5B, required for 60S ribosomal subunit joining at the initiation step, was found through a yeast two-hybrid screen against *Drosophila* ovarian cDNAs (Pestova et al., 2000). A mutant form of Vas, $Vas^{\Delta 617}$, does not interact with eIF5B and once expressed in the ovaries does not support Grk translation, suggesting that Vas directly activates Grk expression by recruiting translational machinery to *grk* transcripts (Johnstone and Lasko, 2004).

Grk translation is blocked as a response to meiotic checkpoints that in turn are activated by unrepaired dsDNA breaks (DSBs) (Ghabrial and Schupbach, 1999). Thus mutations in genes such as *okra* and *spindle-B*, which are involved in DSB repair, cause defects in dorsoventral patterning. *vas* mutants produce similar nuclear-morphology phenotypes as spindle-class-mutants, suggesting that meiotic checkpoints act through down-regulation of Vas (Styhler et al., 1998; Tomancak et al., 1998). Vas expression is not affected in spindle-class-mutants; however, band shift assays indicate that the mobility of Vas protein from ovarian lysates prepared from these mutants is decreased compared to wild-type, and is restored by an additional mutation in *mei-41*, a gene involved in checkpoint pathway. These observations support a model, whereby meiotic checkpoints suppress Vas activity in Grk translation through post-translation modifications (Abdu et al., 2002; Ghabrial and Schupbach, 1999). Grk translation is restored in *spnB* mutants by lowering the dosage of eIF1A, correlated with an increase in polysome-associated *grk* transcripts (Li et al., 2014). eIF1A facilitates recruitment of Met-tRNAi-eIF2-GTP ternary complex (TC) to the 40S ribosomal subunit and thus promotes translation initiation

(Hinnebusch, 2014). The unexpected inhibitory effect of eIF1A on Grk translation, is thought to be associated with three upstream open reading frames (uORFs) in *grk* 5′ UTR (Li et al., 2014). uORFs often reduce translation from the main ORF, by consuming pre-initiation complex, and creating a block on the mRNA as the ribosome stall on premature stop codons (Barbosa et al., 2013). eIF1A mutations could result in a larger fraction of ribosomes to pass the uORFs before TC recruitment and thus to start translation from main ORF (Fekete et al., 2005; Li et al., 2014).

In addition to its role in Grk translation, Vas has been also implicated in translation of *mei-P26*, a gene that promotes GSC differentiation through its function in the microRNA (miRNA) pathway (Liu et al., 2009; Neumuller et al., 2008). Mei-P26 translation is significantly decreased in $vas^{\Delta 617}$, suggesting that Vas interaction with eIF5B is required for translational activation of Mei-P26 in GSCs. Details of the mechanism by which Vas activates translation remains elusive; but structure analyses and biochemical data suggest that through its unwinding activity Vas resolves secondary structures in 5′ or 3′ UTRs of specific mRNA targets, and thus removes translation barriers (Lasko and Ashburner, 1988; Liang et al., 1994; Sengoku et al., 2006).

1.7 Vas role in piRNA pathway

Maintaining genome integrity is particularly important in the germline, because deleterious mutations in these cells could be transmitted to the next generation. Active transposable elements (TEs) impose a constant threat to the host cells through their mobility and insertion into essential genes. Therefore, the germline has evolved defence mechanisms to protect its genome against these internal enemies. Most notably the Piwi-interacting RNA (piRNA) pathway, which is highly conserved amongst animals, targets TE-derived transcripts for degradation, and thus blocks TE replication (Grimson et al., 2008). Increasing evidences show that Vas is a critical component of piRNA biogenesis consistent with its functional conservation in germline of different animals.

1.7.1 piRNA pathway protects genome integrity in the germline

Transposable elements, which were first identified about 50 years ago, are DNA sequences that can move from one location in the genome to another (Pray, 2008). TEs comprise a large

portion of all eukaryotic genomes; for example, roughly 50% of the human genome is occupied by transposons and transposon-like repetitive elements (Mills et al., 2007). Transposons are classified based on their mode of propagation. Retrotransposons employ a "copy and paste" mechanism by which the original insertion remains in the genome and a new copy is generated by reverse transcription, resulting in amplification of the element. In contrast DNA transposons move through a "cut and paste" mechanism, which includes excision of the original copy and subsequent repair of the resulting gap in the DNA (Pray, 2008). Transposon mobility could have deleterious effects by altering expression of the neighboring genes, although in rare cases it could be also evolutionarily beneficial. Germline is a battlefield for the race between transposons and the host, as TE-mediated changes in the genome could be transmitted to the new generation when they affect germ cells. Transposons become particularly active in the germ cells and their embryonic precursors as a result of epigenetic reprogramming (Zamudio and Bourc'his, 2010). TEs have also evolved mechanisms to ensure their vertical propagation across generations. For example, expression of the active P element transposase in *Drosophila*, or the proteins required for mobility of L1, the major non-LTR transposon in human, is restricted to the germline (Branciforte and Martin, 1994; Seleme et al., 1999). In parallel germ cells have evolved special strategies to tame transposons and protect their genome integrity. One highly conserved mechanism to do this is mediated by piRNA, a large class of small non-coding RNAs, which target and cleave TE-derived RNA molecules (reviewed by Haase, 2016).

Small RNA pathways include small-interfering RNA (siRNA), micro-RNA (miRNA) and piRNA. siRNA and miRNA, which are present in both plants and animals, originate from long dsRNA precursors with perfect or imperfect duplexes, respectively. dsRNA is processed to 22-23-nucleotide small interfering RNAs (siRNAs) by the RNase III ribonuclease Dicer. siRNA/miRNA is then loaded onto Argonaute (AGO) family proteins, such as Ago1, to form the RNA induced silencing complex (RISC), which subsequently targets transcripts for destruction in the case of siRNA or translational inhibition by miRNA (Carthew and Sontheimer, 2009). In contrast piRNAs, which are primarily found in the animal germline, derive from ssRNA precursors and associate with members of the Piwi clade, named after *Drosophila* Piwi (P-element induced wimpy testis), including Piwi, Aub and Ago3 (Meister, 2013). Over 90% of piRNAs derive from discrete genomic loci, called piRNA clusters (Fig. 1.3; Brennecke et al., 2007). These loci are mainly located in pericentromeric and telomeric heterochromatin, and act

as a registry, which contains inactive copies or truncated fragments of all transposons hosted in the genome. Numerous genetic studies have provided a functional link between piRNA clusters and transposon silencing. For example, mutations in *flamenco/COM* locus lead to upregulation of retrotransposons, *gypsy, Idefix* and *ZAM* (Desset et al., 2003; Pelisson et al., 1994). Little is known about piRNA cluster transcription from these, otherwise silenced, heterochromatic regions. The Heterochromatin Protein 1 (HP1) variant, Rhino (Rhi), is required for production of piRNA precursors from 42AB dual-strand cluster (Klattenhoff et al., 2009). A recent study shows that Rhi interaction with the transcription termination factor Cutoff (Cuff), via the adaptor protein Deadlock, is required for piRNA production. This study suggests a model for non-canonical transcription from piRNA clusters that involves preventing RNA Polymerase II from termination, and protecting 5′ end of the nascent transcripts (Mohn et al., 2014). The long single stranded transcripts are then exported to cytoplasm for processing into small RNAs, through a mechanism that involves the DEAD-box protein, UAP56 (Zhang et al., 2012). On piRNA clusters, close to the inner surface of the nuclear membrane UAP56 colocalizes with Rhino (Fig. 1.3). UAP56 foci are directly across Vas-containing foci on cytoplasmic side of nuclear pores. This together with biochemical and genetic evidence suggests that UAP56 and Vas interact to transfer piRNA transcripts across the nuclear envelope.

Once transcripts reach perinuclear region of the cytoplasm they undergo primary processing, which requires activity of the ssRNA-specific endonuclease, Zucchini (Zuc), to generate the monophosphorylated 5′ end (Nishimasu et al., 2012). Subsequently, the pre-piRNA is loaded on Aub or Piwi and the 3′ end is trimmed, by an unknown mechanism, to produce mature piRNAs (Haase, 2016; Meister, 2013). Piwi-piRNA complexes are translocated to nucleus to bind complementary sequences in the genome, and by recruiting epigenetic factors such as HP1, silence gene expression (Huang et al., 2013). piRNA guided epigenetic silencing by Piwi is also an alternative mode for transposon regulation (Saito, 2013).

The primary piRNA loaded on Aub targets complementary sequences in transposon transcripts, for their cleavage by Aub slicer activity. Simultaneously collaboration between Aub and Ago3 produces secondary piRNAs through an amplification cycle also referred to as the "ping-pong loop". This mechanism is considered as an adaptive strategy for robust elimination of

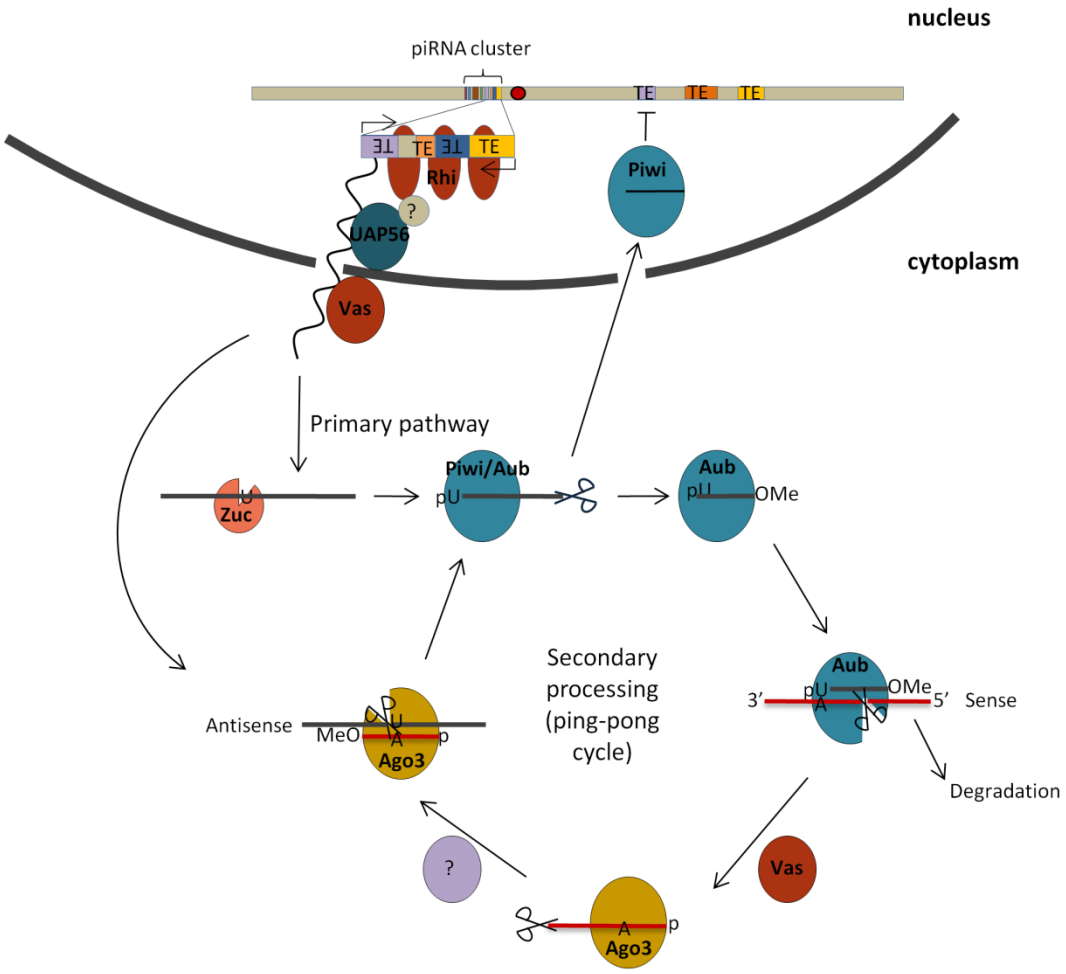


Fig. 1.3 piRNA transcripts are derived from transposon-rich heterochromatin, called piRNA clusters, in the pericentromeric or subtelomeric regions (Brennecke et al., 2007). The *Drosophila* HP1 variant, Rhino (Rhi), binds to piRNA cluster sequences and colocalizes with the DEAD-box protein UAP56 adjacent to the nuclear pore (Klattenhoff et al., 2009; Zhang et al., 2012). The other DEAD-box protein, Vas, is localized to the cytoplasmic face of the nuclear envelope and functions with UAP56 to transfer cluster transcripts to the cytoplasm. During primary processing the long piRNA transcripts (antisense) are cleaved upstream of the 5′ uridine (U) bias by the nuclease Zucchini (Zuc) and loaded on Aubergine (Aub) or Piwi (Nishimasu et al., 2012). The 3′ end of primary piRNA is trimmed by an unknown nuclease and is 2′-O- methylated (Ome). The Piwi-piRNA complex is imported to the nucleus to exert transposon silencing (Huang et al., 2013), whereas Aub-piRNA complex targets trasnposon transcripts (sense) for cleavage. The 3′ cleavage product is transferred to Ago3 with the help of Vas and undergoes further 3′ end trimming (Xiol et al., 2014). Ago3-piRNA complex then targets antisense transcripts to produce piRNA intermediate fragments, which are delivered to Aub for starting a new cycle (Hirakata and Siomi, 2016). TE: transposable element

active transposons as opposed to immobile ones or the host genes which might have been accidently targeted (Brennecke et al., 2007; Weick and Miska, 2014). Sequences of small RNAs associated with Aub or Piwi indicate that they often correspond to the antisense strand of retrotransposons, and have a strong bias for uracil (U) at their 5′ ends, whereas Ago3-bound RNAs are predominantly derived from sense strands, do not have a preference for 5′ U, but rather they are biased for adenine (A) at the 10th nucleotide position from 5′ terminus (Gunawardane et al., 2007). These observations led to a model, whereby Aub-associated piRNA-induced RISC complex (Aub-piRISC) targets transposon transcripts to be cleaved 10 nucleotides upstream to the biased A (Fig. 1.3). The 5′ cleavage product is subsequently released for degradation, and the 3′ product is delivered to Ago3, where its 5′ end is phosphorylated and the 3′ end is trimmed and methylated. The resulting Ago3-piRISC will then target sense strands, which are cleaved at the 5′ side of the biased U. Again the 5′ cleavage product is degraded and the 3′ product, which is sometimes considered as the tertiary piRNA, is loaded on Aub for subsequent 3′ trimming and to initiate a new round of the feed-forward loop (Hirakata and Siomi, 2016).

1.7.2 Vas function in piRNA pathway

The role of Vas in piRNA biogenesis was first discovered by studying Mvh (Mouse Vas homolog)-deficient mice (Kuramochi-Miyagawa et al., 2010). Defective spermatogenesis in these mutants was associated with elevated levels of several retrotransposons, and impaired *de novo* DNA methylation of TE regulatory regions, resembling phenotypes observed in Mili- or Miwi2 (mouse homologs of Piwi)-deficient mice (Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004). Further analyses indicated that Miwi2-bound piRNAs are significantly decreased in Mvh mutants, and suggested a crucial role for Mvh in the ping-pong cycle (Kuramochi-Miyagawa et al., 2010).

Additional insights to the role of Vas in piRNA pathway come from studies in *Drosophila* and in *Bombyx mori* (Bm; silkworm) ovary-derived BmN4 cell line (Nishida et al., 2015; Xiol et al., 2014; Zhang et al., 2012). Vas is localized to, and is required for the assembly of the nuage, a specialized perinuclear structure found in the germline of various organisms (Hay et al., 1990; Lasko and Ashburner, 1990; Liang et al., 1994). Many piRNA components including Aub, Ago3 and Tud domain proteins, Tej and Kumo, also localize to the nuage, supporting the idea that

nauge functions as a processing site for the amplification cycle of piRNA pathway (Anand and Kai, 2012; Harris and Macdonald, 2001; Patil and Kai, 2010; Webster et al., 2015). As mentioned previously a significant portion of Vas-containing loci in the nuage are localized on the nuclear pores, adjacent to and dependent on localization of Rhi and UAP56 foci on the nuclear side (Zhang et al., 2012). Consistent with this colocalization and similar to *rhi* and *uap56* mutants, several transposon families are over-expressed in *vas* null ovaries. These observations, together with co-immunoprecipitation of piRNA cluster transcripts with Vas and the reduced levels of germline-specific piRNAs in *vas* null ovaries, suggest that Vas and UAP56 cooperate on opposite sides of nuclear pores for the export of cluster transcripts to cytoplasm.

The role of Vas in piRNA amplification cycle and its molecular mechanisms were unknown for a long time partly due to the dynamics of Vas interactions, similar to many other RNA helicases (Linder and Jankowsky, 2011). ATP hydrolysis by DEAD-box proteins and product dissociation are required for progression of repeated RNA unwinding events (section 1.5.1); thus, mutations in conserved residues, for example in motif II, that are involved in ATPase activity, are expected to create an RNA clamp and stabilize transient complexes. Based on this mechanism of function, a mutant allele of *vas* (vas^{DQAD}) was tested in BmN4 cell line, which is the only cell culture model with active piRNA biogenesis pathways (Kawaoka et al., 2009; Xiol et al., 2014). This study provided several important results that clarified the mechanism of Vas function in ping-pong cycle. Fluorescence recovery after photobleaching (FRAP) experiments indicated that in contrast to eGFP-BmVas (*B. mori* Vas) in nuage granules, which is rapidly replaced from surrounding cytoplasm, eGFP-$BmVas^{DQAD}$ fails to recover after photobleaching (Xiol et al., 2014). Furthermore, DQAD mutation stabilized Vas in a complex containing a number of proteins involved in piRNA amplification cycle such as Siwi (the *Bombyx* ortholog of Aub), Ago3 and Qin/Kumo. Deep sequencing of small RNA associated with Vas^{DQAD} shows that these have characteristics of piRNAs and, in addition, a majority of them map to the Siwi-bound antisense piRNAs. The Vas-Siwi-Ago3-Qin amplifier complex also contained poly(A) piRNA precursors corresponding to transposon transcripts, suggesting that Vas cooperates with Siwi for processing sense strands either by regulating the slicer activity of Siwi, or by transferring the sense piRNA intermediates from Siwi to Ago3. Further experiments, using a construct expressing an artificial secondary piRNA precursor (sense), supported the latter function by indicating that 100% of reporter-derived fragments associated with Vas^{DQAD} have 5′

monophosphate ends already generated by Siwi slicer cleavage activity. Together these observations suggest a model, whereby Vas in its RNA-bound close conformation provides a platform for the assembly of a piRNA amplifier complex and promotes transfer of the sliced precursor fragments from Siwi/Aub to Ago3 (Xiol et al., 2012).

Consistent with Xiol et al. (2012), co-IP experiments in a subsequent study isolated Siwi from a wild type BmVas complex, albeit at much lower levels than from $BmVas^{DQAD}$ complex (Nishida et al., 2015). This study also shows that a FLAG-BmVas purified from BmN4 cells could liberate RNA from Siwi-piRISC *in vitro*. The Siwi/BmVas complex, which is required for production of BmAgo3-secondary piRNAs, is found distinct from Siwi/BmSpn-E/BmQin complex, involved in primary piRNA production, thus excluding Vas from the primary phase of piRNA processing.

1.8 Vas function in regulating mitosis and meiosis

In *vas* null (vas^{PH165}) ovaries germarium atrophy is associated with a decrease in GSCs, CBs and dividing cystocyte clusters, which are recognized based on their specific structures, spectrosomes and fusomes (Styhler et al., 1998). This phenotype resembles, although is less severe than, that observed in null *nos* mutants. Nos is required for GSC proliferation (Forbes and Lehmann, 1998), however evidence to support that Vas function in regulating GSC divisions is mediated by *nos* is still lacking. More recently it was shown that concomitant with a delay in mitotic progression, a higher percentage of GSCs and CBs in *vas* null, compared to *wt*, do not progress to metaphase and remain in prometaphase (Pek and Kai, 2011a). Consistent with this Vas becomes localized at pericentromeric piRNA clusters in GSCs and CBs during prometaphase/metaphase, dependent on its interaction with two piRNA pathway components Aub and Spn-E. Co-IP from GSC/CB-enriched *bag-of-marbles* (*bam*) mutant ovaries indicates that Vas directly interacts with the condensin I components, Barren (Barr) and CAP-D2. Localization of Barr to chromosomes, which is required for chromosome condensation and segregation, is abolished in *vas* null GSCs and CBs, and is restored by transgenically expressed Vas (Bhat et al., 1996; Pek and Kai, 2011a; Somma et al., 2003). Interestingly $vas^{\Delta 617}$, which does not interact with eIF5B, could still rescue Barr recruitment to the chromosomes in *vas* null,

suggesting that *vas* role in mitotic progression of germline stem cells is independent of its function in translation regulation.

Work in other species also supports Vas function in cell cycle progression of embryonic stem cells. In sea urchin embryos at fifth cell division Vas protein becomes restricted to the small micromeres, which are early stem cells in larva giving rise to the adult rudiment (Voronina et al., 2008; Yajima and Wessel, 2011a). In early embryos however, Vas is expressed in all blastomeres, where it exhibits a dynamic localization during cell cycle (Yajima and Wessel, 2011b). Vas is localized perinuclearly during prophase, but becomes associated with chromosomes concurrent with their condensation starting at prometaphase, and continued at metaphase and anaphase. Vas-depleted cells have a prolonged mitotic (M) phase, with chromosomes that replicate but fail to segregate. Further experiments indicate that Vas is required for translation of Cyclin B, which colocalizes with Vas on the spindle complex. This study together with the evidence from other species for association between Vas and mitotic spindles, suggests that an ancient function of Vas, prior to its restriction to germ cell lineage, has been to promote mitosis in rapidly dividing cells through local activation of cell cycle factors (Carre et al., 2002; Oyama and Shimizu, 2007; Yajima and Wessel, 2011b). A recent study in spider *Parasteatoda tepidariorum*, also confirms Vas role in maintaining mitosis integrity (Schwager et al., 2015). In this species conserved germ cell markers for the first time appear after germ band formation, suggesting that germ cells are specified independent of a pole plasm. Yet injection of *Pt-vasa* or *Pt-piwi* RNAi to the hemolymph of females significantly reduces their number of eggs and results in a high rate of lethality associated with aberrant mitotic morphologies in early embryos. These mitotic defects included fewer cells entering metaphase, and impaired chromosome segregation.

Bel, the *Drosophila* ortholog of mammalian DDX3, has the closest sequence homology to Vas, and is expressed both in somatic cells and in germ cells (Johnstone et al., 2005). Bel, which is required for larval growth and adult fertility, can functionally substitute for yeast Ded1p, a protein implicated in translation (Chuang et al., 1997), suggesting that the essential role of Bel in development is through regulation of translation. In addition, Bel has been shown to function in gene silencing through RNA interference (RNAi) (Ulvila et al., 2006; Zhou et al., 2008). Similar to Vas, which regulates mitotic progression in GSCs via its interaction with piRNA components,

Bel functions, through endogenous (endo)-siRNA pathway, to promote chromosome segregation in somatic cells (Pek and Kai, 2011a; Pek and Kai, 2011b). This function is through a physical interaction between Bel and the condensin I components, Barr and CAP-D2, while localization of Bel to the pericentric region of chromosomes is mediated by endo-siRNA pathway components, Ago2 and Dicer-2.

Furthermore, ectopic expression of Vas in human embryonic stem cells (hESCs) or induced pluripotent stem cells (iPSCs) is shown to promote germ cell differentiation and progression through meiosis (Medrano et al., 2012). In these experiments over-expression of both Vas or/and another RNA-binding protein, Dazl, resulted in a significant increase in the number of monosomic cells and the cells expressing a post-meiotic marker named Acrosin. Interestingly Vas-induced differentiations were associated with a significant decrease in DNA methylation consistent with the fact that epigenetic marks at imprinted loci in mammals are specifically removed in germ cells (Lee et al., 2002).

1.9 Vas orthologs in other species

1.9.1 *C. elegans*

Unlike most other species that have only one Vas protein, in *C. elegans* there are four Vas orthologs: the germline helicases, GLH-1, GLH-2, GLH-3, and GLH-4. All these proteins are associated with P granules and segregate into germline P blastomeres (Gruidl et al., 1996; Kuznicki et al., 2000). In addition to the conserved helicase domain and similar to *Drosophila* Vas, GLH-1, -2 and -4 carry an N-terminal Gly-rich domain, which however unlike most other Vas orthologs does not contain RGG repeats. Instead, these proteins contain FGG motifs commonly found in nuclear pore proteins (Suntharalingam and Wente, 2003) and carry multiple CCHC zinc fingers. Among four GLH proteins, only deletion of GLH-1 results in complete sterility at elevated temperatures while GLH-4 functions redundantly with GLH-1, and the animals lacking both are sterile at all temperatures (Spike et al., 2008). Detailed phenotypic analyses of *glh-1* mutants suggest impaired stem cell divisions together with dissociation of several P granule components, most dramatically PGL-1 and PGL-2, into the cytoplasm.

P granules in *C. elegans*, similar to nuclear pore complex (NPC), provide a size-exclusion barrier depending on hydrophobic interactions between FGG motifs found in GLH proteins (Patel et al., 2007; Ribbeck and Gorlich, 2002; Updike et al., 2011). GLH proteins, alone, cannot form P granules when ectopically expressed in intestinal cells, and require co-expression of PGL proteins with intrinsic self-aggregating properties. P granules normally localize around the nuclear periphery depending on the presence of GLH in their structure (Updike et al., 2011). P granules were first thought to be required for germ cell determination in *C. elegans* based on their asymmetric segregation to the germline blastomeres (P cells), and sterility of mutants lacking P granule components (Updike and Strome, 2010). Surprisingly, it was found that germ cells are still specified if P granules are destabilized through mutations resulting in symmetrical participating of germ cell components (Gallo et al., 2010). A recent study however highlights the significance of P granules for maintaining totipotency in the germ cells after their specification, by showing that in the absence of P granules germ cells are reprogrammed to a somatic fate and start expressing neuronal and muscle markers (Updike et al., 2014).

Consistent with the interaction between Vas and Dicer observed in *Drosophila* and mouse, the *C. elegans* ortholog, GLH-1, also binds Dcr-1 (Beshore et al., 2011; Kotaja et al., 2006; Megosh et al., 2006). Immunofluorescent images indicate interdependent localization of GLH-1 and Dcr-1, which overlap transiently, on nuclear pores of germ cells in the meiotic pachytene region of ovaries. Furthermore, both GLH-1 and Dcr-1 relocalize to RNP granules and regulate their assembly in arrested oocytes. These observations resulted in a model in which GLH-1 associates with Dcr-1 to regulate maternal mRNAs transferred from nucleus to cytoplasm and later stored in the cytoplasmic and cortical RNP granules of arrested oocyte (Beshore et al., 2011).

1.9.2 Sea urchin

In sea urchins, the close relatives of chordates, four small micromeres (smms) are formed by the fifth embryonic division from micromeres (mms), and solely contribute to the adult rudiments in the larvae (Tanaka and Dan, 1990). *vas* mRNA, which is uniformly distributed in the early embryo, later exclusively accumulates and becomes translated in smms (Juliano et al., 2006). The large micromeres, the other group of cells produced from mms, do not exhibit Vas upregulation and have a single fate to produce larval skeletal system. In the first attempt to

investigate if smms are the definitive primordial germ cell lineage, their parental cells mms were removed from embryos (Ransick et al., 1996). Surprisingly mm-depleted embryos could still develop into adults with functional gonads, leading to the conclusion that germ cells are not exclusively produced from smms. Indeed after removing mms the remaining cells in the embryo respond by a significant upregulation of Vas, indicating a compensatory fate transition (Voronina et al., 2008). Further studies show that removing smms after fifth division has no effect on normal development, but the resulting adults have poorly developed gonads and produce no gametes (Yajima and Wessel, 2011a). Accordingly, removal of smms does not induce Vas upregulation in the remaining cells. These results indicate that smms are required for germ cell development in sea urchin, and could be considered as PGCs although their exact lineage fate has not been determined.

The exact mechanism by which Vas protein is enriched in small micromeres is not known, but injection of the reporter RNAs in which GFP ORF is flanked by *vas* UTRs, or is fused to Vas ORF indicates that Vas coding region is required and sufficient for accumulation of GFP in smms (Gustafson et al., 2011). Additional experiments show that this enrichment at least partly depends on proteasome-mediated degradation throughout the embryos, which is negatively regulated in the germ cells via direct interaction between Vas and Gus. This together with a similar function identified for Gus in *Drosophila* to increase Vas stability in pole plasm suggest a conserved regulatory mechanism for localization of Vas in PGCs (Kugler et al., 2010). A recent study implicates Seawi, the sea urchin homolog of Piwi, in PGC enrichment of Vas by showing that *seawi* knockdown results in abnormally high levels of Vas throughout the embryo (Yajima et al., 2014).

In sea urchin Vas also accumulates outside of the germline in the rapidly proliferating cells in the coelomic pouches, which form the adult rudiments. Furthermore, Vas is expressed in physically damaged tissues, coincidently with an increase in mitosis, and is required for normal wound healing (Yajima and Wessel, 2015). Consistent with this function of Vas in proliferating cells, Vas-depleted embryos show a significant decrease in general protein synthesis. Immunoprecipitation of RNA associated with Vas shows that Vas binds to a broad population of mRNAs at a similar level to eIF4E, suggesting that Vas acts as a general translational regulator during specific developmental stages.

1.9.3 Zebrafish

In zebrafish maternal *vas* RNA, which is initially distributed uniformly throughout oocyte, becomes embedded in germ plasm adjacent to the indented furrow starting at 2-cell stage (Olsen et al., 1997; Yoon et al., 1997). Vas protein however is not a component of germ plasm, is perinuclearly localized only during early oogenesis, and its translation from maternal mRNA starts in PGCs at cell cycle 13 (Knaut et al., 2000). Vas mRNA localization in zebrafish, similar to ostariophysan species, depends on conserved elements in the 3′ UTR, which are commonly missing in nonlocalized *vas* mRNA of the euteleost clade including medaka (Knaut et al., 2002; Shinomiya et al., 2000). Interestingly reporter constructs containing zebrafish *vas* 3′ UTR, when injected into a *Xenopus* oocyte, could localize despite no clear sequence similarity to *DEADSouth*, an RNA encoding a DEAD-box protein in *Xenopus*, which localizes to the vegetal pole (Knaut et al., 2002; MacArthur et al., 2000). This indicates that the localization machinery utilized by zebrafish *vas* and *DEADSouth* is conserved. Analysis of zebrafish *vas* locus reveals presence of several large repeats in the 5′ end, including smaller tandem repeats with multiple RGG boxes (Bartfai and Orban, 2003). In addition numerous isoforms were identified which result from alternative splicing and polyadenylation.

In larvae homozygous for a loss of function allele of *vas*, maternal Vas protein is detectable until 10-12 days after fertilization consistent with normal PGC specification and migration in these animals (Hartung et al., 2014). However *vas* mutant adults exclusively develop to sterile males similar to the mutants for other germ cell-essential genes such as zebrafish *piwi* (z*iwi*; Houwing et al., 2007). Closer examination of gonads in these mutants shows that germ cells are present in bipotential gonads of juveniles, but meiosis does not progress beyond the pachytene stage. Thus none of these animals develop ovaries, and the immature testis that is formed is devoid from germ cells. Loss of germ cells in *vas* mutants, which is likely due to apoptosis, is however independent of Caspase-3 and is not suppressed in *tumor suppressor protein p53 (tp53)* mutants, lacking one of the main activators of apoptosis (Berghmans et al., 2005).

Since *vas* mutants do not develop into females, they cannot be used to study germ cell formation in the absence of maternal *vas* mRNA. However, injecting a *vas* morpholino to early zebrafish embryos, which resulted in undetectable Vas levels until 24 hour after fertilization, did

not affect PGC number (Braat et al., 2001). Authors argue that this might be due to failure of the morpholino to efficiently inhibit translation of zygotically transcribed *vas* mRNA.

1.9.4 *Xenopus*

Xenopus vasa-like gene 1 (Xvlg1), first appears in presumptive PGCs (pPGCs) at late gastrula (Ikenishi et al., 1996). Significance of Xvlg1 in germ cell formation was shown through an experiment, where individual vegetal blastomeres were injected with a specific antibody, to perturb Xvlg1 function (Ikenishi and Tanaka, 1997). This study indicated that the injected blastomeres, which were detected using a lineage tracer, although produced a normal number of somatic cells did not contribute to the PGCs.

In situ hybridization for *Xvlg1* in oocyte, embryos and tadpoles reveals that *Xvlg1* mRNA is uniformly present throughout stage I oocytes except in the mitochondrial cloud, from which germ plasm originates (Heasman et al., 1984; Ikenishi and Tanaka, 2000); signal progressively fades in stage II and III oocytes and is generally absent from the vegetal blastomeres including germ plasm-bearing cells (GPBC) in the embryos until gastrula. Thereafter, and concurrent with maternal to zygotic transition, *Xvlg1* RNA becomes expressed again and remains present in PGCs throughout most stages of life. The RNA is also detected in certain somatic cells until embryonic stage 46, suggesting involvement of Xvlg1 in differentiation of somatic tissues (Ikenishi and Tanaka, 2000).

In contract to *Xvlg1*, *DEADSouth*, which belongs to a different family of DEAD-box helicases, localizes to mitochondrial cloud and subsequently segregates with germ plasm (MacArthur et al., 2000). Using a reporter construct it was shown that the 3′ UTR of *DEADSouth* mRNA is sufficient to drive localization to the germ plasm (Kataoka et al., 2006). The 3′ UTR of *DEADSouth* also contains recognition elements for miRNA miR-427, which ensures degradation of *DEADSouth* mRNA in the somatic cells after mid-blastula transition (MBT), when maternal mRNA are generally cleared from embryos (Yamaguchi et al., 2014). Both over-expression and knock-down of DEADSouth in the embryos results in a reduction of PGC numbers (Yamaguchi et al., 2013). The decreased PGC number in DEADSouth-depleted embryos is caused by reduced PGC divisions and is associated with defects in perinuclear localization of germ plasm in PGCs

after MBT. Interestingly PGC number is restored in these embryos by a construct containing the ORF from *Xenopus Xvlg1* or mouse *mvh*, and the 3′ UTR of *DEADSouth*.

1.9.5 Mouse

Northern blot analyses of several adult tissues indicate that the transcript of mouse *vas* homolog, *mvh,* is exclusively expressed in testis (Fujiwara et al., 1994). Careful examination using immunofluorescence and electron microscopy in testis from mice, rat and guinea pig confirm that Mvh is present in all nuage-related structures in spermatocytes and spermatids, including small dense particles, intermitochondrial cement (IMC), loose strands and chromatoid body (CB) (Onohara et al., 2010; Russell and Frank, 1978). During early development in both male and female embryos Mvh is first detected in PGCs 10.5–11.5 dpc, coincident with colonization of the gonadal ridges (Fujiwara et al., 1994). This expression continues in germ cells throughout gonad development and in males is detected until post-meiotic spermatid stage. *mvh*-deficient males are infertile while mutant females do not exhibit reproductive defects (Tanaka et al., 2000). Histological examinations using several stage-specific markers show defects in premeiotic differentiation of spermatogenic cells, which results in a complete lack of postmeiotic cells in seminiferous tubules. A male-specific decrease in proliferation of PGCs, after reaching genital ridges, also contributes to the sterility of *mvh* male. Later analyses show that spermatogenesis defects are associated with over-expression of retrotransposons in *mvh* mutant testis, and provide evidence that Mvh plays a role in ping-pong cycle of piRNA pathway (Kuramochi-Miyagawa et al., 2010).

In a study examining arginine methylation of Vas orthologs from *Drosophila*, *Xenopus* and mouse, Mvh was found methylated at Arg62 and Arg105 (Kirino et al., 2010a). Based on their flanking amino acids these two Arg residues are expected to be targeted for symmetric and asymmetric argenine dimethylation (sDMA and aDMA), respectively (Bedford and Richard, 2005; Krause et al., 2007). Consistent with this prediction SYM11 antibody, recognizing sDMA-glycine, and ASYM24, reacting with aDMA-glycine, both detect Mvh (Boisvert et al., 2002; Boisvert et al., 2003). Arginine methylation of *Drosophila* Aub and mouse Mili, mediates their respective interactions with Tudor domain proteins, Tud and Tdrd6 (Kirino et al., 2010b; Vagin et al., 2009). Similarly Myc-tagged Mvh was found to interact with Tdrd1, Tdrd6, Mili or Miwi, all expressed in 293T cells. However, none of these interactions were disrupted through co-

expression of a methyltransferase inhibitor, which appeared to only reduce sDMA levels of Mvh (Kirino et al., 2010). Another study shows that Mvh is also acetylated at Lys405 through histone acetyl transferase 1 (Hat1) and its co-factor p46, which specifically localizes to chromatoid body (Nagamori et al., 2011). More than 800 distinct mRNAs were further identified by this study as Mvh targets in testis including eIF4B, whose product stimulates ATPase activity of eIF4A and thus activates translation (Altmann et al., 1993). Consistent with decreased RNA-binding of Mvh following its acetylation *in vitro*, Mvh association with its mRNA targets, such as *eIF4B*, could be developmentally modulated by changes in Mvh acetylation, which parallel temporally regulated localization of Hat1 and p46 in the CB (Nagamori et al., 2011).

1.9.6 Human

The human *vas* gene (also known as *DDX4*), is mapped to chromosome 5 and encodes a 79 kDa protein, which in addition to the conserved motifs in the core region shares with the other *vas* orthologs a Gly-rich N-terminus containing multiple RGG motifs (Castrillon et al., 2000). *vas* expression in human is confined to ovaries and testis and is undetectable in somatic tissues. The first appearance of Vas-positive cells is at gestational age of 7 weeks, when PGCs migrate to the mesenchyme of gonadal ridge. As sexually dimorphic gonads develop Vas becomes most abundantly expressed in mature oocytes and spermatocytes, while it is absent in spermatozoa.

In testis hypomethylation of CpG islands in DDX4 promoter is correlated with its expression in spermatogenic cells (Kitamura et al., 2007). Consistently histological assessment of patients with azoospermia or severe oligozoospermia reveals reduction of Vas expression and methylation of *vas* promoter, being potentially responsible for sperm maturation arrest in these patients (Sugimoto et al., 2009). Ectopic expression of *vas* in hESCs and iPSCs induces germ cell formation and results in meiosis completion in a subset of cells (Medrano et al., 2012); section 1.8). This study also shows that, unlike another RNA-binding protein, Dazl, which similarly induces germ cell formation *in vitro*, ectopic expression of *vas* specifically results in erasure of epigenetic imprints, a critical process in germ cell development (Lee et al., 2002).

To isolate rare female germline or oogonial stem cells (OSCs) from ovaries of adult mice and human a fluorescence-activated cell sorting (FACS)-based protocol has been established that relies on immunological detection of cell-surface variants of DDX4 or Mvh (White et al., 2012).

The purified OSCs by this method could be expanded and produce functional oocytes *in vitro* or after transplantation into adult mouse ovaries. A recent study however raises concerns whether isolated OSCs by this protocol are true DDX4-expressing cells, and functional germline stem cells (Zhang et al., 2015).

1.10 Vas function in totipotent stem cells

The role of *vas* in stem cell biology has been mostly supported by studies of planarians, a model system with remarkable regenerative capabilities (Shibata et al., 1999). The regeneration process depends on a population of totipotent somatic stem cells called neoblasts. These cells are most notably characterized by the presence of large cytoplasmic ribonucleoprotein granules called chromatoid bodies, which are very similar to germline granules in other organisms (Coward, 1974; Hori, 1982). Two *vas*-related genes, *DjvlgA* and *DjvlgB,* have been identified in *Dugesia japonica*, which are specifically expressed in the germline of sexual planarians and have close sequence homology to PL10 and An3 in mouse (Shibata et al., 1999). Interestingly DjvlgA, which like Vas contains several RGG motifs in its N-terminal region, is expressed in growing blastema differentiated from neoblasts at the wound site. Phylogenetic analyses in lower metazoans including sponge, *Hydra* and planaria shows that *vas*-related genes, such as *DjVas-1*, evolved by duplication of *PL10*-related genes (Mochizuki et al., 2001). DjVas-1, together with several other RNA-binding factors, is exclusively expressed in neoblasts of asexual planaria, and its depletion through RNAi treatment results in the failure of animals to regenerate after amputation, without affecting the population of neoblasts and mitotic cells (Rouhana et al., 2010). Two *vas* homologs, *smed-vasa-1* and *smed-vasa-2*, in another planarian, *Schmidtea mediterranea*, are expressed in an irradiation-sensitive pattern indicating their specific expression in neoblasts (Wagner et al., 2012). *smed-vasa-1*(RNAi) animals are able to grow new tissues after amputation; however, these regenerated parts regress subsequently, suggesting a role for *vas* in maintenance of tissue integrity.

Several other germline-specific genes such as *tud, pum, piwi* and *bru* are expressed in totipotent stem cells and are required for tissue regeneration and in most cases for neoblast maintenance (Guo et al., 2006; Palakodeti et al., 2008; Salvetti et al., 2005; Solana et al., 2009).

This suggests a conserved multipotency program shared between germ cells in higher metazoans and somatic stem cells in the ancestral metazoans (Juliano et al., 2010). A new hypothesis challenges the classical division between preformation and epigenesis for germ cell determination (Solana, 2013). According to this model germ plasm flows from zygote to germ cells through primordial stem cells (PriSCs), such as neoblasts in planarians and totipotent cells of inner cell mass (ICM) in mammals. To obtain a comprehensive transcriptome of unlimited PriSCs (uPriSCs) a recent study compares transcriptomic profile of totipotent archeocytes in fresh water sponge, *Ephydatia fluviatilis*, and the available RNA sequencing data of uPRiSCs from two other early-diverging animal lineages, cnidarian *H. vulgaris* and flatworm *S. mediterranea* (Alie et al., 2015). These analyses strikingly indicated that transcription factors are poorly represented in the ancestral stem cells. In contrast, RNA-binding proteins (RBPs), particularly germline associated RNA regulators such as Piwi, Vas, Bru, Ddx6, Mago, PL10 and all Tudor domain proteins, are strongly over-expressed in uPRiSCs. Additional phylogenetic analyses reveal that contrary to DNA replication and DNA repair factors, which are shared between eukaryotes and bacteria, the vast majority of the identified RBPs, including the piRNA components, were acquired concomitant with the emergences of stem cells.

1.11 Vas implication in cancer

Increasing evidence shows that a soma-to-germline transition occurs in cancer, resulting in neoplastic characteristics such as rapid cell division and undifferentiated state (McFarlane et al., 2014). In *Drosophila* genome-wide gene expression profiling of mutants for the tumor suppressor gene, *lethal (3) malignant brain tumor* [*l(3)mbt*], shows that a quarter (26 of 102) of the up-regulated genes encode germline-required factors (Janic et al., 2010). These genes include *nos* and the piRNA components *piwi*, *aub*, *krimper* (*krimp*) and *tej*. Further analysis indicated that although *vas* mRNA is not increased in *l(3)mbt* brains, Vas protein is ectopically expressed. In addition, brain overgrowth was significantly reduced in double mutants for *l(3)mbt* and one of several germline genes including *vas*, confirming that these genes indeed contribute to malignancy. A subsequent study shows that 28 of 49 germline genes reported to be over-expressed in *l(3)mbt* tumors, at the mRNA or protein level, have human homologs corresponding to 48 genes due to human paralogs (Feichtinger et al., 2014). Among these human homologs

56% are associated with germ cell functions. Using microarray meta-analysis this study also shows that thirty-one of these genes are significantly over-expressed in 11 distinct cancer types. These findings are supported by a previous study showing that mutations in retinoblastoma pathway in *C. elegans* result in expression of the germline specific gene, *PGL-1*, in somatic tissues. Ectopically expressed PGL-1 localizes to the P granules in a perinuclear structure resembling what is normally found in the germline blastomeres (Wang et al., 2005).

A specific example of Vas implication in somatic cancer is the epithelial ovarian cancer (EOC) where *vas* was found over-expressed in 21 of 75 tissues tested (Hashimoto et al., 2008). This study also indicated that 14-3-3σ, a well-characterized cell cycle regulator which is involved in G2 checkpoint following DNA damage (Chan et al., 1999), is down-regulated in SKOV-3 cells stably expressing Vas. When these cells were exposed to epirubicin, a drug inducing DNA damage, they exhibited defects in G2 checkpoint function and entered mitosis, whereas vector-control cells were arrested in the G2 phase. This result suggests that abnormal regulation of Vas could disrupt G2 checkpoint and cause chromosomal instabilities. A later study analyzing expression pattern of DDX4 in 59 ovarian cancer patients confirms that DDX4 mostly colocalizes with CD133, an ovarian cancer stem cell (CSC) marker, and suggests the possibility of using DDX4 as another CSC marker in ovarian cancer (Kim et al., 2014). In addition, Vas expression has been detected in various types of germline-derived testicular tumors including seminoma, carcinoma *in situ*, dysgerminoma, and gonadoblastoma, again suggesting that Vas could serve as a highly specific marker for germ cell tumors (Zeeman et al., 2002).

1.12 Vas structural analyses; preferential binding to mRNA targets

Crystal structure of the helicase-core region of Vasa (residues 200–623) bound to a single-stranded RNA and AMP-PNP, a nonhydrolyzable analog of ATP, has been solved (Sengoku et al., 2006). These structural analyses indicate that the N-terminal domain (NTD; residues 233–454), and the C-terminal domain (CTD; residues 463–621) fold in essentially the same way as in the other DEAD-box proteins, although these two domains in Vas have a much more closed relative orientation possibly due to specific sequence of the 8-residue linker region (residues 455–462). As a unique feature of the DEAD-box proteins, in contrast to the other subfamilies of

SF2 proteins, the interdomain interactions in Vas ensure that the bound RNA is sharply bent, and thus duplex formation is disrupted. The interactions of the residues in twelve conserved motifs of Vas with RNA, ATP and the other parts of protein, are essentially the same as those in other DEAD-box proteins. This study also indicates that the core region only interacts with the phosphate-ribose backbone of the RNA and not the base moieties, suggesting that this region does not contribute to any potential sequence specificity of Vas toward its RNA targets.

Vas–mRNA complexes were purified from 0–2 h embryos through a tandem immunoprecipitation approach with anti-Vas antibody (Liu et al., 2009). In this work 221 mRNA were co-purified with Vas, 13.3% of which were enriched in the pole cells. One of the recovered mRNA was *mei-P26* with a known function in GSC differentiation (Neumuller et al., 2008; Page et al., 2000). Vas directly interacts with 3′ UTR of *mei-P26,* depending on a stretch of 10 consecutive U residues, and activates Mei-P26 translation (Liu et al., 2009). The polyuridine tracts are not overrepresented among the mRNAs purified with Vas compared to whole *Drosophila* mRNAs; however, germ cell enriched mRNA appear more frequently when the Vas-bound mRNA are filtered for the presence of (U)-rich motif, an approach which may identify other direct targets of Vas. It is further demonstrated that the preferential binding of Vas to the (U)-rich motif compared to the neighboring motifs depends on sequences in its N- or C-terminal regions, as the Vas core protein (VCP: residues 200–623) does not exhibit such discriminatory affinity.

1.13 Rationale for experiments

The Vas crystal structure provides detailed information about the conformation of the helicase domain in complex with RNA and ATP molecules. However, similar to most other DEAD-box proteins, much remains unknown about the role of N- and C-terminal flanking regions in Vas function. Nevertheless, data from *in vitro* assays suggest that these terminal sequences contribute to the sequence specificity of Vas for its mRNA targets (Sengoku et al., 2006; Liu et al., 2009). In the work presented in this thesis I undertook a systematic mutagenesis approach to functionally characterize Vas N- and C-terminal domains through different *in vivo* and biochemical assays. Furthermore, this work aimed to uncouple various Vas functions by

individual mutations or deletions throughout Vas sequence, including the conserved helicase core domain.

The Gly-rich N-terminus in most Vas orthologs contains multiple RGG motifs which are arranged within tandem repeats. The RGG box is proposed to function as an auxiliary motif to enhance RNA binding, and is shown in some cases to confer sequence specificity (Darnell et al., 2001; Thandapani et al., 2013). However, functional significance of RGG motifs in Vas orthologs has not been experimentally tested. Through different N-terminal truncations and point mutations, I show that the divergent N-terminal region, and in particular the RGG motifs, are critical for germline functions of Vas in *Drosophila*. These results are presented in chapter 2. A previous study indicates that Vas promotes germline stem cell divisions, independent of its role as a translational regulator. Extending this notion, through mutations in the conserved core motifs, I indicate that some functions of Vas in oogenesis, which contribute to female fecundity, are independent of its helicase activity. Furthermore, I show that two motifs in Vas C-terminus are essential: one for the efficient localization of Vas to pole plasm and the other for several functions in oogenesis and germ cell formation.

The C-terminal end of Vas orthologs is distinct among DEAD-box proteins, by the presence of multiple acidic residues and a highly conserved tryptophan (Trp660 in *Drosophila*). CRISPR-Cas9 induced homologous recombination was used to generate an endogenous allele of *vas*, replacing Trp660 with glutamic acid (Glu). Through comparative analyses with wild type allele, presented in chapter 3, I provide insights into the role of invariant Trp660 for different functions of Vas. This work also shows that, despite being highly conserved, Trp660 could be replaced with phenylalanine (Phe), another aromatic hydrophobic residue, without major effects on various Vas functions.

Multiple functions of Vas, which are described in this literature review, are likely linked to different RNP complexes that contain Vas. To separately study these functions, which are poorly understood at the molecular level, it is advantageous to test different alleles with mutations that only affect a subset of these RNPs. Among the nine *vas* alleles with point mutations, which have been previously characterized, eight affect the helicase core domain and one maps to R170 in the N-terminal region (Liang et al., 1994). In the majority of these mutants localization of Vas to

pole plasm is abolished, complicating analyses of their potential defects in specific germ cell formation pathways. Several of the transgenic alleles generated in my study exhibit normal localization, but they show defects associated with particular aspects of *vas* function. The next step is to explore protein or RNA interactions impaired by these mutations, to potentially provide further insights into mechanisms of Vas function. In this study I applied a yeast two hybrid (Y2H) assay to particularly investigate significance of the C-terminal acidic motif and the conserved Trp on the interactions between Vas and its known protein partners as well as putative interacting proteins identified in a new Y2H screen. These experiments are discussed in chapter 3 and appendixB.

Chapter 2: *In vivo* mapping of the functional regions of the DEAD-box helicase Vasa

Reproduced with minor modifications from

Dehghani and Lasko (2015). *In vivo* mapping of the functional regions of the DEAD-box helicase Vasa. Biol Open. 20; 4(4):450-62.

Abstract

The maternally expressed *Drosophila melanogaster* DEAD-box helicase Vasa (Vas) is necessary for many cellular and developmental processes, including specification of primordial germ cells (pole cells), posterior patterning of the embryo, piRNA-mediated repression of transposon-encoded mRNAs, translational activation of *gurken* (*grk*) mRNA, and completion of oogenesis itself. Vas protein accumulates in the perinuclear nuage in nurse cells soon after their specification, and then at stage 10 Vas translocates to the posterior pole plasm of the oocyte. We produced a series of transgenic constructs encoding eGFP-Vas proteins carrying mutations affecting different regions of the protein, and analyzed *in vivo* which Vas functions each could support. We identified novel domains in the N- and C-terminal regions of the protein that are essential for localization, transposon repression, posterior patterning, and pole cell specification. One such functional region, the most C-terminal seven amino acids, is specific to Vas orthologs and is thus critical to distinguishing Vas from other closely related DEAD-box helicases. Surprisingly, we also found that many eGFP-Vas proteins carrying mutations that would be expected to abrogate DEAD-box helicase function localized to the nuage and posterior pole, and retained the capacity to support oogenesis, although they did not function in embryonic patterning, pole cell specification, *grk* activation, or transposon repression. We conclude from these experiments that Vas, a multifunctional protein, uses different domains and different molecular associations to carry out its various cellular and developmental roles.

Introduction

The maternally-expressed *Drosophila melanogaster* gene *vasa* (*vas*) encodes a DEAD-box RNA binding protein that is required for posterior patterning and germ cell specification (Hay et al., 1988b; Lasko and Ashburner, 1990). Vas orthologs are expressed in the germ cell lineage, and linked to germ line development, in many other animals including mammals (Castrillon et al., 2000; Fujiwara et al., 1994; Komiya et al., 1994; Kuznicki et al., 2000; Lehmann and Nüsslein-Volhard, 1991; Olsen et al., 1997; Yoon et al., 1997). The first *vas* alleles that were studied in *Drosophila* have a recessive maternal-effect lethal phenotype, in that homozygous females produce non-viable embryos that lack primordial germ cells (pole cells) and most posterior segmentation (Schupbach and Wieschaus, 1986a). Subsequent study of null *vas* alleles such as vas^{PH165} revealed earlier functions in oogenesis; *vas*-null females produce very few mature eggs, and the eggs that are produced have dorsal appendage defects resulting from a failure to activate *gurken* (*grk*) translation in the oocyte (Styhler et al., 1998; Tomancak et al., 1998). Grk is an epidermal growth factor receptor (Egfr) ligand that in normal development is secreted from the anterodorsal corner of the oocyte, activating Egfr in the adjacent follicle cells and thus specifying dorsal fate (González-Reyes et al., 1995; Roth et al., 1995). Vas interacts with the translation initiation factor eIF5B, which is necessary for ribosomal subunit joining (Carrera et al., 2000). A mutant form of Vas, $Vas^{\Delta 617}$, that has a greatly reduced ability to interact with eIF5B does not activate translation of *grk* (Johnstone and Lasko, 2004) or of *mei-P26*, the product of a gene involved in differentiation of germ line stem cells whose expression is also dependent on Vas (Liu et al., 2009). These results suggest a model whereby Vas activates translation of target mRNAs through preferential recruitment of this initiation factor (Lasko, 2013).

Other studies, however, have implicated Vas in processes that do not appear to involve regulation of translation. In early oogenesis, Vas accumulates in nurse cells in a structure called the nuage, then at stage 10 it is transferred to the oocyte where it accumulates in the posterior pole plasm (Hay et al., 1988a; Hay et al., 1988b; Lasko and Ashburner, 1990). Vas is also required for the assembly of the nuage (Liang et al., 1994), which is the cytoplasmic site for processing of Piwi-associated RNA (piRNA) precursors (Pek et al., 2012b). Vas co-immunoprecipitates with RNA precursors produced from major piRNA clusters, and in

cooperation with a nuclear DEAD-box helicase, UAP56, facilitates the transport of piRNA precursors from the nucleus through nuclear pores to the cytoplasmic nuage (Zhang et al., 2012). piRNAs target transposon-encoded mRNAs and thus protect the genome against the deleterious effects of uncontrolled transposition events (Klattenhoff and Theurkauf, 2008; Pek et al., 2012b). Vas physically interacts with some components of the piRNA pathway thereby assembling an amplifier complex on the transposon-encoded transcripts and promoting the ping-pong cycle (Xiol et al., 2014). In *vas*^{PH165} ovaries, many transposon-encoded mRNAs are overexpressed, consistent with an essential role for Vas in piRNA biogenesis (Zhang et al., 2012). Another recent study links Vas to regulation of mitotic chromosome condensation in the *Drosophila* germ line, and shows that this function can be carried out by Vas$^{\Delta 617}$, and thus is independent of eIF5B association (Pek and Kai, 2011a).

These many Vas functions make it difficult to study each individually, especially in the case of later functions such as pole cell specification that depend upon earlier events such as posterior localization. Thus, mutants affecting only a single Vas-dependent process would be valuable. To identify such mutants, and to carry out a structure-function analysis of Vas, we generated a set of mutant *egfp-vas* transgenes, and examined their activity in both a maternal-effect lethal *vas* (*vas*^{1}) background and in a *vas*-null background. The Vas protein is organized into three domains: a central region of approximately 400 amino acids that is conserved in all DEAD-box helicases, a rapidly evolving N-terminal region of approximately 200 amino acids that contains numerous RGG motifs, and a short C-terminal region. In other RNA-binding proteins RGG motifs act as ancillary motifs, which increase affinity to RNA or confer sequence specificity for RNA binding (Burd and Dreyfuss, 1994; Darnell et al., 2001; Lamm et al., 1996). The C-terminal region contains some amino acids that are conserved in Vas orthologs in other *Drosophila* species and terminates with a stretch of highly acidic residues, conserved in Vas orthologs in *Drosophila* and beyond, but not in other DEAD-box helicases.

In this study we generated several *egfp-vas* constructs with deletions in the N-terminal region and one that additionally abrogates remaining RGG motifs. By expressing these eGFP fusion proteins in a *vas*^{PH165} or a *vas*^{1} background, we tested the significance of the N-terminal sequence, in general, and the RGG motifs, in particular, for localization of Vas and for its different cellular and developmental functions. Similarly, we investigated the importance of a

conserved motif near the C-terminus as well as the highly acidic sequence at the C-terminal end. In addition, we examined the effects on Vas function *in vivo* of non-conservative missense mutations in canonical DEAD-box helicase motifs that would be expected to render the protein catalytically inactive.

Materials and Methods

Fly stocks and transgenic lines

vas^{PH165} is a null allele, lacking the entire coding region of *vas*, which was generated by imprecise P-element excision (Styhler et al., 1998). *vas*^{1} is a hypomorphic allele in which there is no amino acid substitution in the coding region but Vas expression is limited to the germarium (Lasko and Ashburner, 1990; Liang et al., 1994; Schupbach and Wieschaus, 1986b). To generate *egfp-vas* constructs, the open reading frame (ORF) of *vas* was first cloned into a pENTR vector (Life Technologies) with *XhoI* and *NotI* restriction sites added to the 5' and 3' ends, respectively. This construct was used to generate deletions or point mutations in the *vas* sequence by PCR-based site-directed mutagenesis. The *XhoI*/*NotI* digested *vas* ORF from this construct was then inserted into *P*[*w*+;*Pvas-egfp*] (Nakamura et al., 2001). *egfp-vas*$^{+}$ and *egfp-vas*$^{\Delta 617}$, which were previously generated (Johnstone and Lasko, 2004), and all the N-terminally deleted *egfp-vas* constructs were based on a cDNA clone that lacks one copy of a 39 nucleotide tandem repeat, encoding amino acids 141-153 (Lasko and Ashburner, 1988). The *egfp-vas*^{G294A}, *egfp-vas*^{E400A} and *egfp-vas*^{D554A} constructs also lack these 39 nucleotides as well as the sequence encoding amino acids 15-75. P element-mediated germ-line transformation was performed using a standard technique (Rubin and Spradling, 1982).

Western blots

To compare protein levels between different transgenic lines, ovary lysates from 2-5 day-old females were resolved on 10% SDS-PAGE gels. Proteins were then transferred onto nitrocellulose membranes for immunoblotting. Primary antibodies were anti-Vas (1:10000), anti-α-tubulin (1:15000, Sigma) or anti-GFP (1:2500, Life Technologies). To examine stability of eGFP-Vas in the pole plasm, protein lysates were prepared from 0-2 h old embryos and processed as for the ovary lysates.

Egg-laying assay

For each experiment four or five females, after eclosion, were paired with the same number of males and allowed to lay eggs on grape juice plates supplied with yeast for 3 days at 25˚C. Eggs were collected and counted once per day. The embryos were also scored for dorsal appendage morphology.

Hatching assay

For this assay, virgin *vas*^{1} females, expressing different *egfp-vas* transgenes, were paired with Oregon-R males and allowed to lay eggs on yeasted grape juice plates. Overnight embryo collections were incubated at 25˚C for 48 h, and then larvae and unhatched eggs were counted.

Immunostaining and *in situ* hybridization

Ovaries from 2-4 day old females were dissected in PBS and fixed for 20 min in 200 µl of PBS containing 4% paraformaldehyde and 0.5% NP-40, plus 600 µl heptane. To immunostain embryos they were collected for 5 h and then fixed in 3 ml PBS containing 4% paraformaldehyde, plus 3 ml heptane. The fixed embryos were stored in methanol at -20˚C. Immunostaining was performed according to the protocol described by (Liu et al., 2009). Anti-Vas (1:1000) and anti-Grk (1:500) were used as the primary antibodies. Fluorescent *in situ* hybridization on the embryos was carried out according to the protocol described by Lécuyer et al. (2007). *fushi tarazu* (*ftz*) cDNA was amplified by PCR from total cDNA synthesized from embryonic mRNA, using gene-specific primers with T3 or T7 promoter sequences added to their 5' ends. This PCR product was then verified by sequencing and used as a template to produce a digoxigenin (DIG)-labeled probe. Horseradish peroxidise-conjugated anti-DIG (Jackson Immunoresearch) and Alexa Fluor 555 tyramide (Life Technologies) were used to detect the probe after hybridization. Samples were examined under a fluorescent Leica DM6000B microscope. Confocal images were captured using a Zeiss LSM510 microscope.

Reverse transcription quantitative PCR (RT-qPCR)

Total RNA was extracted from ovaries of 1-2 day-old females, using TRI reagent (Sigma). Samples were then treated with Turbo DNase (Ambion) and the RNA was precipitated using acid phenol: chloroform (Ambion). cDNA was synthesized from 2 µg RNA using Maxima H

Minus First Strand cDNA synthesis kit (Thermo Scientific). Random hexamers were used for reverse transcription, and to enrich *rp49* cDNA a specific RT primer (TTGGAGGAGACGCCG) was added to the mixture. The resulting cDNA was used for RT-qPCR using *HeT-A-1*, *18S* rRNA and pre-*rp49* primers described by (Zhang et al., 2012). RT-qPCR was performed in a CFX96 Real-Time machine (Bio-Rad) using DyNAmo Flash SYBR Green (ThermoScientific). The expression level of *HeT-A* was quantified relative to *18S* rRNA and pre-*rp49*. A minimum of three biological replicates were tested for each genotype. The graph shows the average and the error bars indicate the standard error of the mean (SEM).

Live imaging

Embryos were collected for 30 min, then dechorionated using 5% bleach for 1 min and rinsed with water. Embryos were then aligned in a Fluorodish cell culture dish (World Precision Instruments) and covered with Halocarbon oil 400. Live imaging was performed using a WaveFX spinning disk confocal system (Quorum) and a DM6000B inverted microscope (Leica). Images were captured and processed using Volocity 3D image analysis software (PerkinElmer).

Statistical analyses

A two-tailed t-test was used to calculate the p-values in each assay. The difference between two constructs was considered significant if the p-value was less than 0.05. To assess the effect of different deletions or point mutations on Vas function, the mutated *egfp-vas* constructs were compared to the wild type control (*egfp-vas*$^{+}$), unless otherwise specified.

Results

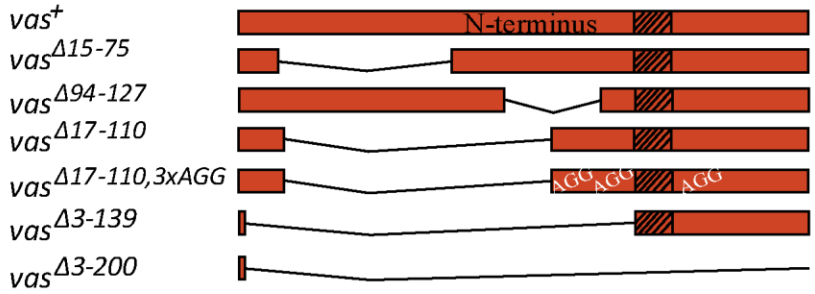
A series of eGFP-Vas proteins with mutations in the N-terminus, conserved helicase domains and the short C-terminus

The N-terminal region of Vas evolves rapidly; there are numerous sequence changes even between *D. melanogaster* and the closely related species *D. simulans* (Fig. 2.1 A). Nearly all Vas orthologues contain several RGG repeats within this region, although the number of such motifs and their spacing vary considerably. To examine the role of the variable N-terminal

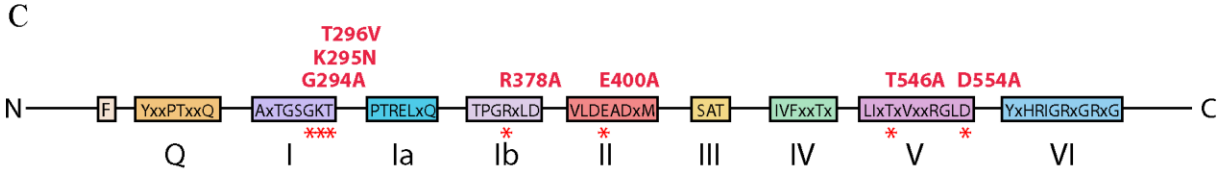
A

```
D.simulans      MS-DWEDEPFVDTNGARGGDWSDDE--------KGKSFSGGAEDGGSGGGG---AGGGYQ 48
D.sechellia     MS-DWEDEPFVDTNGARGGDWSDGEN-------KGKSFSGGAQDGGAGGGGD-GGGGGYQ 51
D.melanogaster  MSDDWDDEPIVDTRGARGGDWSDDED-------TAKSFSGEAEGDGVGGSG--GEGGGYQ 51
D.erecta        MS-DWDDEPSAQTSRG-GGDWDGAASN-----KPGEEVSEWSDDE------------HK 40
D.pseudoobscura MS-DWSEEEVDSNQ-----------------GTKAKEVDGWSDEETP------------ 29
D.virilis       MS-DWDDEPDCTNQPIHKPHIDDKWDSKGEASSNKKTVSEWSDEEVAGDGDGDGIAKGYR 59
                ** **.:*    .                      : ..  ::
D.simulans      GRNPDGFGRGFKRGDAGRGGGAGGYRGGNRDGDDDAGGFREGEGDFRGGR------REEK 102
D.sechellia     GRNPDGFGRGFKRGDAGRGGGAGGYRGGNRDGDDDAGGFREGEGDFRGGRGFHGGRREEE 111
D.melanogaster  GGNRDVFGR---IG-GGRGGGAGGYRGGNRDGGGFHGGRREGERDFRGGEG---GFRGGQ 104
D.erecta        GKS---FGG------GGRGGDGAGYRGGNRDGDRDGFGFRGDGGGFRGGR--------HE 83
D.pseudoobscura ---QTTASR----GGGPHGVDRGSYRGG-RA-GGDGG--YGRRYDAEDG----------- 67
D.virilis       GVGQGAGYR----GRRDDGQGGGGYRGR-RD-DGEGGGYRGRRDDGEGGG---------- 103
                          * . ..***  *  .          . ..*
D.simulans      SASR-----WRSR----------------------FRGGRREE---TR-----LR--E 123
D.sechellia     GGFRGEEGGFRGEEGGFRGGQGGFRGGEGGFRGGEGGFHGGRREEEGGSRSGEGGFRGDE 171
D.melanogaster  GGSRGGQGGSRGGQGGFRGG--------------EGGFRGRLYENEDGDER-RGRLDREE 149
D.erecta        GGGG---------------------------------FRGRRDDNE------------D 97
D.pseudoobscura ------------------------------------EGRR------------------- 71
D.virilis       ------------------------------------YRGRRDDGE-------------- 112
                                                    .*
D.simulans      RRGRREG-FNQEPRG-ERNE---RGEAGFERRRRNE--DDINNNND-IVEDVQKK-RE 172
D.sechellia     RHGRREG-FNKEPRG-ERNE---RGEAGFERRRRNE--DDINNNND-IVEDVQKK-RE 220
D.melanogaster  RGGERRGRLDREERGGERGE---RGDGGFARRRRNE--DDINNNNN-IVEDVERK-RE 200
D.erecta        RGGRGGG-FNRE-RG-ERGE---KGEGGFERRRRND--DDVNNNND-IIEDGEKK-RE 145
D.pseudoobscura -GGFNRDRGDRG-DGEERGE---RK----ERRRKDDDEDDLNNNKD--GEDAEKK-RE 117
D.virilis       -GGGYRGRRDND-DGEERRGGFVRKSGGFERRRRDD--GDMNNNQEDVGEDSEKKARE 166
                  *   .  :.   * **     :      ***:::  .*:***::   ** ::* **
```

B



C



D

```
H.sapiens       DVPAWLEEIA------FSTYIP-GFSG-STRG-NVFASVDTRK---GKSTLNTAG-----FSSSQAPNPVDDESW-
M.musculus      DVPAWLEEIA------FSTYVPPSFSS-STRGGAVFASVDTRKNYQGKHTLNTAG-----ISSSQAPNPVDDESWD
D.rerio         VVPKWLEEVA------FSAHGTTGFN---PRG-KVFASTDSRKGGSFKSDEPPPSQ----ISAPSAAAAADDEEWE
C.elegans       IVPDWMQGAA------GGNYGASGFG---------------------------------SSVPTQVPQDEEGW-
X.laevis        EVPAWLEEIA------FGGHGA--LN--------SFYAADSMG-------EQAGGN----AVTTPSFAQEEEASWD
D.simulans      IAADLIKILD------GAGQTVPEFLR-NLG---ACGGGGYSSQLFGGVDVRGRGN----YVNDATNVE-ADEDWE
D.sechellia     IAADLIKILD------GAGQTVPEFLR-NLG---ACGGGGYSTELFGGVDVRGRGN----YVNDATNVE-ADEDWE
D.melanogaster  IAADLVKILE------GSGQTVPDFLR-TCG---AGGDGGYSNQNFGGVDVRGRGN----YVGDATNVE-EEEQWD
D.erecta        LAGDLIKILE------GSGQTVPEFLR-DLGGGGGGGGGGYSNKKFGGVDVRGRGN----AVAEATYAE-ADEEWD
D.pseudoobscura LASDLVKILE------GSDQVVPGFLR-VLS--GGGGHGGYGSQ-FGGYDVRGSGN----VIQEASAVE-DQQEWD
D.virilis       LAADLIKILE------GSGQEVPEFLK-EIG----GGSSSYFGSKFGGIDVRG-GA----KNGHIENLE-DEEEWN
S.purpuratus    EVPEFLEEAADSAIGTYHGNAGGSFGGRDTRKFGGRGGGRGGGSAVGGGDEWGSGGGGPGGGGGGGGGAADDESWD
```

vasΔ636-646 vasΔ655-661

Fig. 2.1 Summary of the deletions and point mutations examined in this study. **A**. An alignment of the N-terminal ends of predicted Vas proteins from several *Drosophila* species. The N-terminal *vas* open reading frame is often incorrectly annotated in species that have not undergone extensive cDNA sequencing because of confounding factors such as poor sequence conservation, nested genes and the *solo* alternative splice form (Yan et al., 2010). Therefore, to produce this alignment, *vas* open reading frames were manually annotated from three-frame translations of genomic DNA, using the short highly-conserved amino-terminal end to identify the putative translational start site. Asterisks, colons and periods indicate full conservation, strong similarity and weak similarity, respectively. RGG motifs are shown in red. **B**. Schematic representation of the N-terminal deletions used in this study. The hash box marks amino acids 141-153, which are encoded by a copy of a 39-nucleotide tandem repeat that is absent from some *vas* cDNA clones (Lasko and Ashburner, 1988). This segment is absent in eGFP-Vas^{+} and all the N-terminally deleted proteins as the constructs were built from such a cDNA clone. Vas$^{\Delta 17\text{-}110,\ 3xAGG}$ contains a deletion of amino acids 17-110 and three mutations that convert RGG motifs to AGG. **C**. The amino acid substitution mutations in conserved DEAD-box helicase motifs that were produced for this study. Motifs are identified as previously defined (Rocak and Linder, 2004). **D**. Sequence alignment of the C-terminal region of Vas from *D. melanogaster* with orthologs from other species. The red box depicts amino acids 636-646, which are conserved among *Drosophila* species but not beyond. The purple letters show the conserved highly acidic residues found at the C-terminal ends of Vas orthologs from *Drosophila* and non-*Drosophila* species. A tryptophan residue (presented in blue) is also highly conserved.

domain in Vas function, we produced *egfp-vas* transgenic constructs deleted for different segments of the N-terminal region (*egfp-vas*$^{\Delta 15\text{-}75}$, *egfp-vas*$^{\Delta 17\text{-}110}$, *egfp-vas*$^{\Delta 94\text{-}127}$, *egfp-vas*$^{\Delta 3\text{-}139}$, *egfp-vas*$^{\Delta 3\text{-}200}$; Fig. 2.1 B). In addition, to specifically test the role of RGG motifs, we mutated arginine to alanine in the three RGG motifs that remain present in *egfp-vas*$^{\Delta 17\text{-}110}$ (*egfp-vas*$^{\Delta 17\text{-}110,\ 3xAGG}$); the three arginines mutated are R115, R122, and R163.

To examine the role of the conserved DEAD-box domains, we produced *egfp-vas* transgenic constructs with the following point mutations: G294A, K295N, T296V, R378A, E400A, T546A, and D554A (Fig. 2.1 C). Nine highly conserved motifs have been identified in DEAD-box helicases (Rocak and Linder, 2004). The first three of these mutations affect motif I, which is the Walker A motif, a well-established ATP binding domain conserved throughout evolution (Cordin et al., 2006; Walker et al., 1982). R378A affects motif Ib; E400A affects motif II, the DEAD-box; and the final two mutations affect motif V.

Near the C-terminal end, Vas contains a region (amino acids 636-646) that is conserved among closely related *Drosophila* species and that includes a potential arginine methylation site (R644). The C-terminal sequence also ends with a highly acidic region (amino acids 655-661), which is found in many Vas orthologs. To examine the functional significance of these sequences in the C-terminal regions we generated *egfp-vas*$^{\Delta 636\text{-}646}$, *egfp-vas*^{R644A} and *egfp-vas*$^{\Delta 655\text{-}661}$ transgenic lines (Fig. 2.1 D).

The *egfp-vas*$^{+}$ wild-type control and all the constructs with N-terminal deletions were generated from a *vas* cDNA that lacks one copy of a 39 nucleotide tandem repeat, encoding amino acids 141-153 (Lasko and Ashburner, 1988). The G294A, E400A and D554A constructs also carried the Δ15-75 deletion. Neither the deletion of amino acids 15-75 nor 141-153 has any measurable effect on Vas localization or function (Lasko and Ashburner, 1990; Styhler et al., 1998; this work). All the other constructs were generated from a full-length *vas* cDNA.

Most non-null mutations of *vas*, including some affecting residues highly conserved in DEAD-box helicases, can support egg production.

Females homozygous for *vas* deletions produce very few mature eggs (Styhler et al., 1998). To examine which domains of Vas are essential for oogenesis we expressed eGFP-Vas proteins

with various mutations in a *vas*^{PH165} mutant background, and counted the number of eggs produced (Fig. 2.2 A). For these and subsequent experiments, we selected individual transgenic lines that gave comparable to or higher levels of eGFP-Vas expression than the *egfp-vas*$^{+}$ wild-type control (Fig. 2.2 B, Fig. 2.S1).

In this manner, we found that deletion of the first 200 N-terminal amino acids, or of smaller segments within the N-terminal region, had less than a two-fold effect on female fecundity (Fig. 2.2 A). This indicates that sequences in the N-terminal region of Vas are largely dispensable for completion of oogenesis. Quite surprisingly, we also only observed modest effects (2-3 fold reduced from the *egfp-vas*$^{+}$ wild-type control) on egg production from females that expressed eGFP-Vas bearing one of three non-conservative mutations in the Walker A motif (G294A, K295N, T296V) or either of two mutations (R378A and T546A) in other highly conserved DEAD-box helicase motifs (Fig. 2.2 A). A slightly larger effect (around 5-fold) was observed from E400A. All these mutant forms of eGFP-Vas rescued egg production to *vas*^{PH165} mutant females at a statistically significant level (Table 2.S1). The Walker A and B motifs comprise a very well characterized and widespread ATP-binding domain (Cordin et al., 2006; Walker et al., 1982), and the four non-conservative substitutions in residues 294-296 and 400 would be expected to abrogate its function. R378A impacts DEAD-box helicase motif Ib, which in eIF4A is essential for RNA binding (Rogers et al., 2002), while T546A affects motif V, which is shown to couple the ATPase and helicase activities in other DEAD-box helicases (Caruthers et al., 2000; Sengoku et al., 2006). Taken together, these results indicate that mutant forms of Vas with non-conservative amino acid substitutions in motifs implicated in ATP binding, and thus also in its RNA helicase activity that depends on ATP hydrolysis, can nevertheless rescue oogenesis to *vas*-null females to a substantial degree. The sole point mutation that did not restore fecundity to *vas*^{PH165} females was D554A, which affects a highly conserved residue of motif V.

We also examined mutants affecting C-terminal residues well conserved among Vas orthologs but not among other DEAD-box helicases for their effects on oogenesis. Similar to the N-terminal deletions, these mutant forms of eGFP-Vas were able to restore fecundity to *vas*^{PH165} females 50-100% as effectively as a wild-type eGFP-Vas fusion (Fig. 2.2 A).

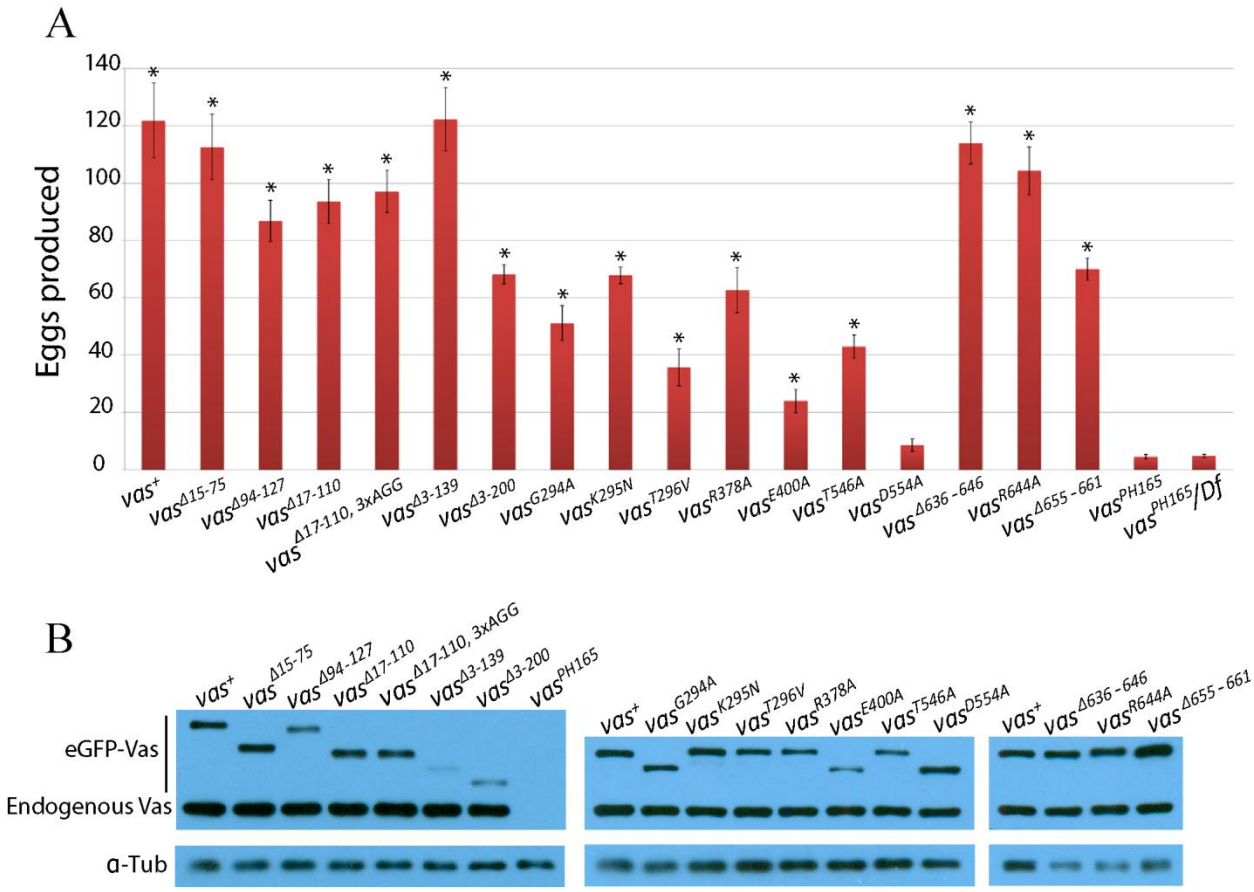


Fig. 2.2 Fecundity of *vas*^{PH165} females carrying different *egfp-vas* constructs, and expression levels from those constructs. **A**. Fecundity. The y-axis indicates the number of eggs laid by individual females in the first three days after eclosion, and the x-axis identifies each *egfp-vas* construct that was tested. Data from *vas*^{PH165} and *vas*^{PH165}*/Df(2L)A267* controls are presented at the far right. Asterisks indicate a significant increase compared to *vas*^{PH165} ($p<0.05$). Error bars indicate the standard error of the mean (SEM), n (number of females tested) >50 for all genotypes. **B**. Western blots (WB) comparing the expression level of eGFP-Vas in the ovaries from *vas*^{1}/+ flies carrying the different constructs, using an anti-Vas antibody. α-Tubulin (α-Tub) serves as a loading control. *egfp-vas*$^{+}$ was included in each blot for comparison. The eGFP-Vas bands migrate at different positions depending on the deletions that they carry. No Vasa protein was detected in the ovary lysate from *vas*^{PH165}. The Vasa antibody raised against the full-length protein reacts with some mutant forms of Vasa such as *vas*$^{\Delta 3\text{-}139}$ and *vas*$^{\Delta 3\text{-}200}$ less strongly than with *vas*$^{+}$ (see also Fig 2.S1).

Vas requires the conserved DEAD-box domains as well as the acidic C-terminus for the translational activation of Grk.

Females null for *vas* function fail to translate *grk* mRNA at wild-type levels (Styhler et al., 1998; Tomancak et al., 1998), and the consequent reduction of secreted Grk ligand available to adjacent follicle cells results in ventralization of their patterning, which can be observed through effects on the structure and position of the dorsal appendages of the eggshell. As patterning becomes more ventralized, the dorsal appendages move closer together, then fuse, and then become absent altogether. We compared the percentages of eggs produced from each line expressing a different eGFP-Vas mutant in a vas^{PH165} mutant background that had two dorsal appendages (separate or partly fused), a single fully fused appendage, or no appendages, as an index of Gurken (Grk) expression (Fig. 2.3 A). Using this assay, we found that most N-terminal deletions of eGFP-Vas had only modest, but nevertheless statistically significant, effects on dorsal appendage formation (Fig. 2.3 A, Table 2.S1). Larger N-terminal deletions had somewhat larger effects with greater statistical significance, and fewer than 20% of eggs produced from females expressing only eGFP-Vas deleted for amino acids 3-200 had two dorsal appendages (Fig. 2.3 A, Table 2.S1). In contrast to what we observed for oogenesis, all mutations in the conserved helicase domains severely compromised dorsal appendage formation (less than 10% of embryos with two dorsal appendages; Fig. 2.3A), and again D554A had the most extreme effect (no embryos with two appendages). The three C-terminal mutations we examined behaved divergently in this assay; deletion of amino acids 636-646 or a point mutation in a conserved residue within this segment (R644A) had only modest but statistically significant effects, while deletion of the seven most C-terminal amino acids (655-661) produced nearly as severe a phenotype as some of the mutations in conserved DEAD-box domains (Fig. 2.3 A, Table 2.S1).

We also examined Grk expression directly in these mutants by immunohistochemical staining of stage 8 egg chambers, where in wild-type a prominent crescent of Grk is apparent adjacent to the nucleus in the anterodorsal corner of the oocyte. We counted the percentage of stage 8 oocytes with a visible Grk crescent, and obtained results that were consistent with those we observed for dorsal appendage formation (Fig. 2.3 B). Again, mutations affecting conserved DEAD-box helicase residues, and deletion of amino acids 655-661, showed strong effects (>85%

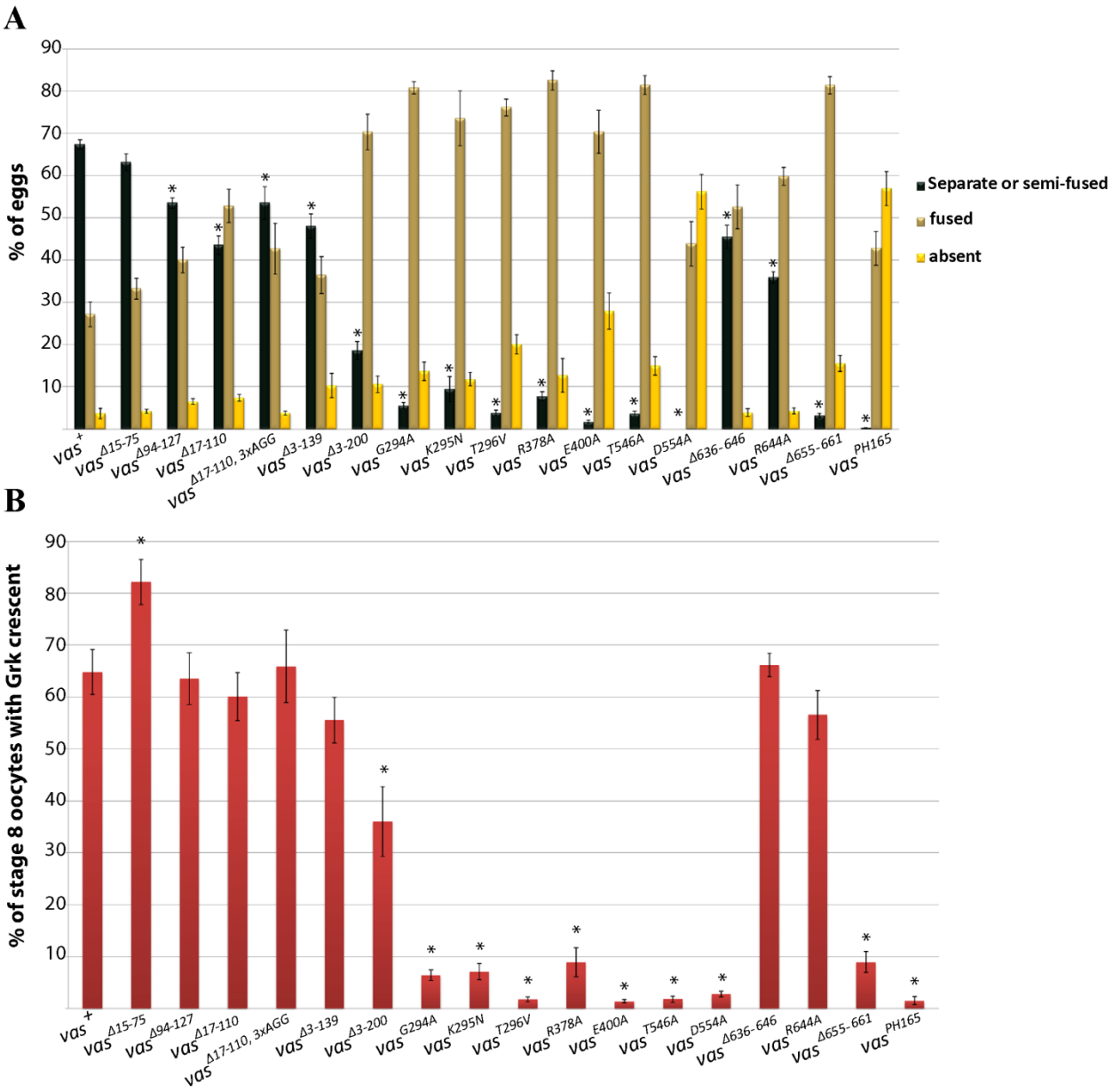


Fig. 2.3 Dorsal-ventral patterning in eggs produced by *vas*^{PH165} females carrying different *egfp-vas* constructs. **A**. Dorsal appendage formation. The green bars indicate the percentage of the embryos with two separate or partly fused dorsal appendages. The beige and yellow bars represent the portion of the embryos with fused or no dorsal appendages, respectively. Data from *vas*^{PH165} controls are presented at the far right. Error bars indicate SEM, n (number of females tested) >50 for all genotypes. Asterisks indicate a significant difference compared to *vas*$^{+}$ ($p<0.05$) **B**. Grk expression. Red bars indicate the percentage of stage 8 egg chambers positively stained for localized Grk. Error bars represent SEM from three different replicates. n (number of stage 8 egg chambers in each replicate) >50 for all genotypes.

decrease in the number of stage 8 egg chambers with a Grk crescent compared to *vas*$^{+}$). The other constructs could largely rescue Grk expression in *vas*^{PH165} oocytes with the exception of *egfp-vas*$^{\Delta 3\text{-}200}$ where a Grk crescent was visible in only 36 ± 6.6% of the oocytes ($p = 0.029$ when compared to *egfp-vas*$^{+}$). The Grk crescent was typically less pronounced in mutants where it was visible in a smaller proportion of oocytes (Fig. 2.S2). Both the indirect dorsal appendage assay and the direct immunohistochemical assay indicated that Vas requires its conserved DEAD-box helicase motifs, as well as its seven most extreme C-terminal amino acids, for translational activation of Grk. This is consistent with our earlier hypothesis that Vas actively promotes *grk* translation (Johnstone and Lasko, 2004).

The N-terminal and C-terminal domains as well as the conserved core region contain important elements for Vas localization.

Next, we analyzed our collection of *egfp-vas* mutants in a *vas*^{1} genetic background to assess phenotypes that manifest themselves at later developmental stages. Expression of Vas in the *vas*^{1} mutant is restricted to the germarium stage, and this mutant produces wild-type numbers of eggs, which however do not support germ cell specification or posterior patterning (Hay et al., 1990; Lasko and Ashburner, 1990; Schupbach and Wieschaus, 1986a). For simplicity we will refer to embryos produced by *vas*^{1} mothers as "*vas*^{1} embryos".

First, we examined the spatial distribution of each mutant form of eGFP-Vas. In early oogenesis, wild-type Vas accumulates in the perinuclear nuage of the nurse cells. Beginning at stage 10, Vas is then transported into the oocyte where it accumulates in the posterior pole plasm. This localization pattern is recapitulated by transgenically expressed wild-type eGFP-Vas, and by eGFP-Vas deleted for amino acids 3-139 or for several smaller segments within that interval (Fig. 2.4 A). However, a more extensive N-terminal deletion of amino acids 3-200 made the association of Vas with the perinuclear nuage weaker and reduced Vas localization to the pole plasm (Fig. 2.4 A).

Mutations affecting conserved DEAD-box residues had different effects on Vas localization. Surprisingly, most of the eGFP-Vas mutants we analyzed retained some ability to localize to the nuage in stage 5 nurse cells and to the posterior pole of the stage 10 oocyte (Fig. 2.4 B). The single exception was T546A, which presented an almost completely uniform distribution. For

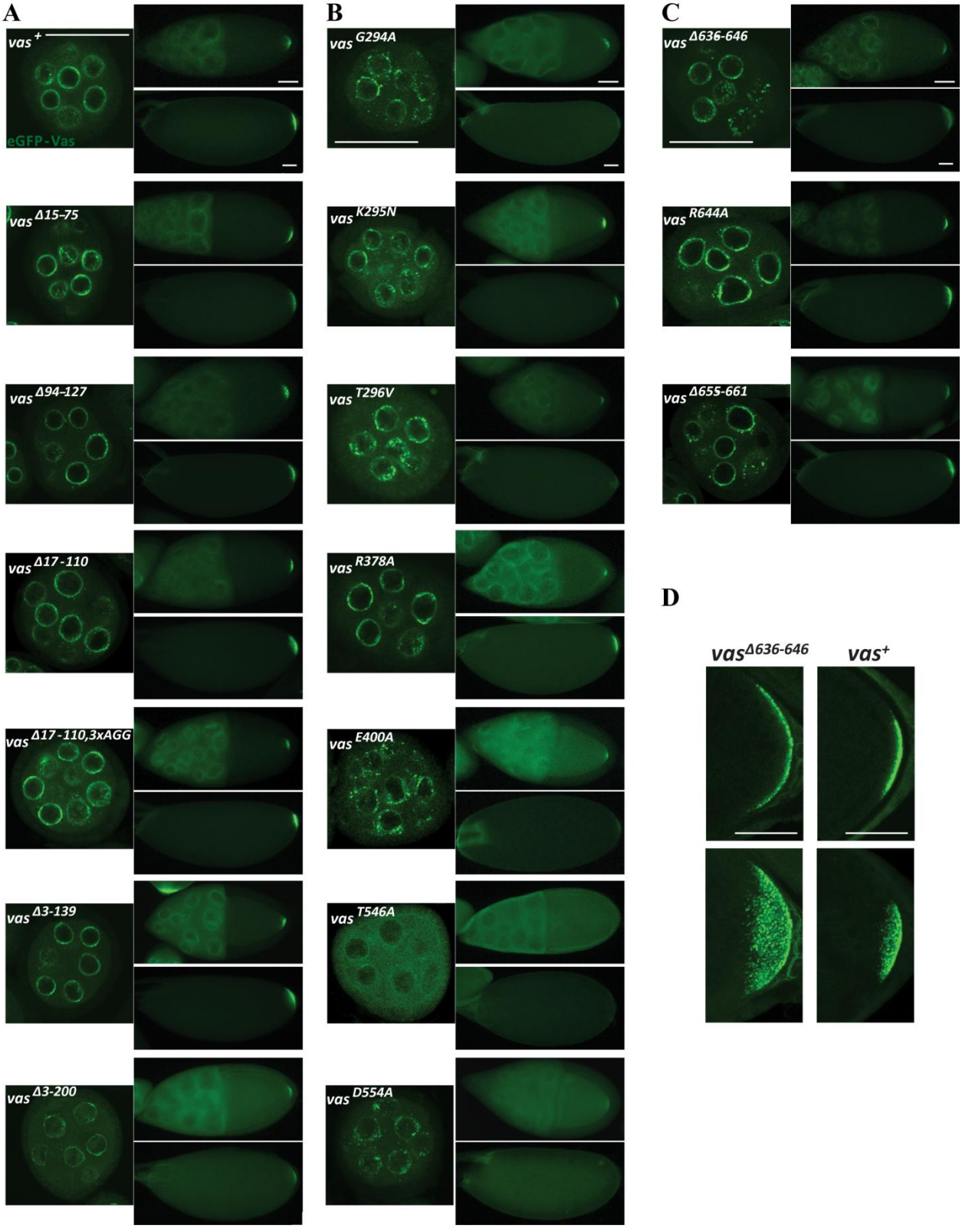

A
vas+
eGFP-Vas
vasΔ15-75
vasΔ94-127
vasΔ17-110
vasΔ17-110,3xAGG
vasΔ3-139
vasΔ3-200
B
vasG294A
vasK295N
vasT296V
vasR378A
vasE400A
vasT546A
vasD554A
C
vasΔ636-646
vasR644A
vasΔ655-661
D
vasΔ636-646
vas+

Fig. 2.4 Localization of eGFP-Vas proteins in *vas*^{1} ovaries. **A**. N-terminal deletions. **B**. Mutations in conserved DEAD-box helicase motifs. **C**. C-terminal mutations. For each genotype the left, the top right and the bottom right images show a stage 5 egg chamber (confocal image), an early stage 10 egg chamber, and a stage 14 oocyte. Scale bars (50 µm) are included on the top set of images; all corresponding images from other genotypes are at the same magnification. **D**. Higher magnification confocal images comparing distribution of eGFP-Vas at the posterior pole of *vas*^{1}*; egfp-vas*$^{+}$ and *vas*^{1}*;egfp-vas*$^{\Delta 636\text{-}646}$ stage 14 oocytes. eGFP-Vas$^{\Delta 636\text{-}646}$ distribution is more diffuse. For each genotype the top image shows the middle focal plane whereas the bottom image illustrates the maximum intensity projection of the z stack. Scale bars= 50 µm.

most other mutations affecting conserved DEAD-box residues, nuage-associated eGFP-Vas was distributed in a far more punctate pattern than in wild-type or in all N-terminal deletions except for Δ3-200 (compare Fig. 2.4 B with A). Posterior localization of most mutant eGFP-Vas proteins affecting conserved DEAD-box residues was transient and did not persist through the end of oogenesis. The R378A allele is an exception in that nuage localization appeared relatively normal and posterior localization was stable, persisting into stage 14 oocytes (Fig. 2.4 B). To a lesser extent, localized eGFP-Vas was also apparent in embryos produced from *vas*^{K295N} mothers (Fig. 2.4 B).

None of the C-terminal mutants we analyzed had drastic effects on localization; however, posterior accumulation of eGFP-Vas deleted for amino acids 636-646 was less robust and more diffuse than in wild-type (Fig. 2.4 C, D).

The role of Vas in repressing transposon-encoded mRNAs requires its helicase function and is also dependent on some N-terminal and C-terminal motifs.

Females lacking *vas* function overexpress mRNA from a subset of transposon families in their ovaries (Zhang et al., 2012). We examined levels of HeT-A, a retrotransposon that is highly expressed in *vas* deficient ovaries, in *vas*^{PH165} females carrying different *egfp-vas* constructs. Consistent with a previous report (Zhang et al. 2012), quantitative RT-PCR (RT-qPCR) analyses indicated that the expression level of HeT-A in the ovaries of homozygous *vas*^{PH165} or *vas*^{PH165} / *Df(2L)A267* females is more than 100-fold higher than in wild-type controls (Fig. 2.5). eGFP-Vas^{+}, as well as the N-terminally truncated proteins lacking the interval 17-110 or shorter, could suppress HeT-A expression in the *vas*^{PH165} ovaries to within an order of magnitude of the wild-type level. On the other hand, expression of HeT-A in ovaries expressing the largest N-terminal deletion (eGFP-Vas$^{\Delta 3\text{-}200}$) was about 50-fold higher than wild-type ($p = 0.0039$), indicating some requirement for N-terminal sequences between amino acids 111-200 in repression of transposon-encoded gene expression. HeT-A expression in *vas*-null ovaries expressing constructs that delete part of this region (*egfp-vas*$^{\Delta 3\text{-}139}$ or *egfp-vas*$^{\Delta 94\text{-}127)}$ was highly variable from experiment to experiment, ranging from near wild-type levels to levels similar to *vas*$^{\Delta 3\text{-}200}$, and because of this variability the difference from *egfp-vas*$^{+}$ was not statistically significant (Table 2.S1). We

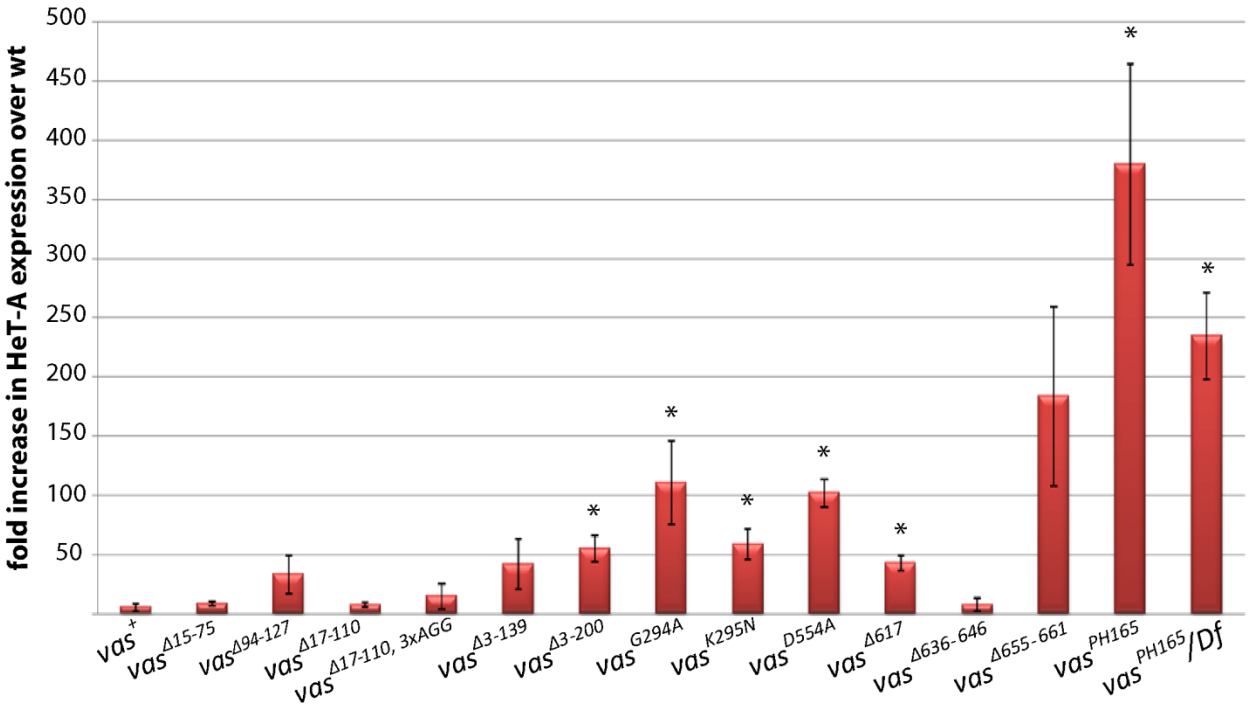


Fig. 2.5 HeT-A expression in ovaries of *vas*^{PH165} females carrying different *egfp-vas* constructs. Red bars indicate the expression level of HeT-A normalized to that of wild-type ovaries. Data from *vas*^{PH165} and *vas*^{PH165}*/Df(2L)A267* controls are presented at the far right. Asterisks indicate a significant increase compared to *vas*$^{+}$ ($p<0.05$). Each bar represents the average of at least three biological replicates, error bars indicate SEM.

observed that HeT-A expression in *vas*^{PH165}; *egfp-vas*$^{\Delta 17\text{-}110,\ 3xAGG}$ ovaries was also highly variable (Fig. 2.5).

Overexpression of HeT-A by 50- to 100-fold was found when ovaries expressing any of three mutant forms of eGFP-Vas affecting conserved DEAD-box domains (G294A, K295N, or D554A) were examined ($p < 0.05$ in all three cases). Surprisingly, Vas$^{\Delta 617}$, a mutant that reduces the interaction with eIF5B and thus affects the role of Vas in translational activation (Johnstone and Lasko, 2004; Liu et al., 2009), also could not rescue suppression of HeT-A expression ($p = 0.0024$). We further found that deletion of the interval 655-661, covering the highly conserved C-terminal acidic residues, reduces the ability of eGFP-Vas to support HeT-A silencing. While all individual data points for this construct showed elevated HeT-A expression (32-261 fold over wild-type), this high degree of variability rendered this difference not statistically significant (Table 2.S1).

In addition to its core helicase domain, Vas requires motifs in the N-terminus and C-terminus regions to support germ cell specification.

eGFP-Vas deleted for amino acids 15-75, 17-110, or 94-127 retained the ability to induce germ cell formation in *vas*^{1} embryos to a similar or even greater extent to eGFP-Vas^{+} (Fig. 2.6). However, deletion of amino acids 3-139 clearly impacted the ability of eGFP-Vas to restore pole cell formation to *vas*^{1} embryos ($p = 0.018$ when compared with eGFP-Vas^{+}), with only 7 ± 1.18% of embryos from females expressing this construct forming pole cells. This was not due to low expression level from this construct (Fig. 2.S1), but we did find that this protein is unstable in embryos (Fig. 2.S3). Next, we examined embryos produced from *vas*^{1} females expressing eGFP-Vas$^{\Delta 17\text{-}110,\ 3xAGG}$ (Fig. 2.6 and Fig. 2.S3). In this line the frequency of embryos forming germ cells was only 14 ± 4.9% ($p = 0.022$ when compared with eGFP-Vas^{+}), suggesting that the RGG motifs in the N-terminal region of Vas play a role in germ cell formation. Importantly, all of these forms of eGFP-Vas localize normally to the posterior pole plasm (Fig. 2.4 A). *vas*^{1} embryos expressing eGFP-Vas$^{\Delta 3\text{-}200}$ never produced germ cells, further implicating N-terminal motifs in pole cell specification.

Despite the ability of some to localize to the pole plasm, all mutant forms of eGFP-Vas that affect conserved DEAD-box residues completely failed to restore the ability to form pole cells to

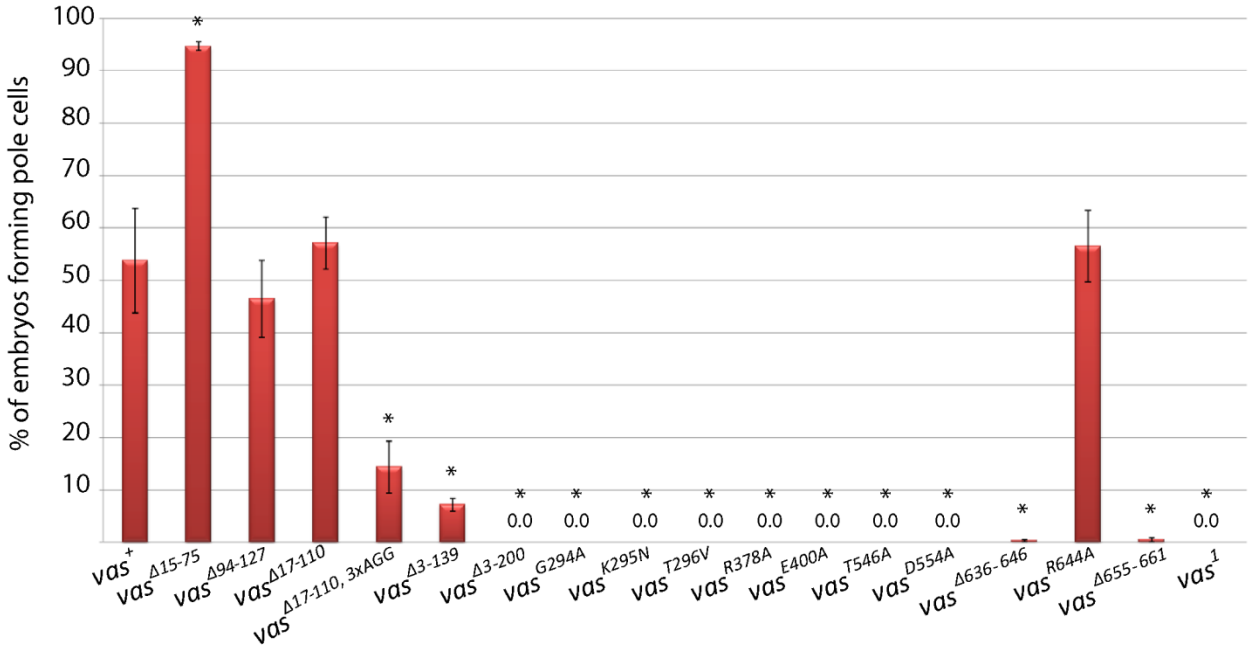


Fig. 2.6 Germ cell formation in embryos from *vas*^{1} females expressing different eGFP-Vas proteins. Red bars indicate the percentage of the embryos with germ cells. Data from the *vas*^{1} control is presented at the far right. Asterisks show a significant difference from *vas*$^{+}$ ($p<0.05$). Error bars represent SEM from at least three biological replicates each with more than 50 embryos.

vas^{1} embryos. This implies that Vas must be catalytically active to mediate germ cell specification.

One C-terminal mutant, eGFP-Vas$^{\Delta 636\text{-}646}$, when expressed at levels comparable to the wild-type control, restored almost no pole cell formation activity to *vas*^{1} embryos (Fig. 2.6). When examined in more detail, however, its phenotype is novel in that pole buds begin to form but then regress. Unlike in wild-type, mitotic divisions remain synchronous in the somatic and posterior pole region of embryos expressing eGFP-Vas$^{\Delta 636\text{-}646}$ (Fig. 2.7, Movie 2.S1 and Movie 2.S2). We hypothesize that the more diffuse localization of eGFP-Vas$^{\Delta 636\text{-}646}$ at the posterior region results in an insufficient concentration of Vas at the extreme posterior to catalyze pole cell formation, and as mitosis remains synchronous, the eGFP-Vas that is present gets divided into a larger number of foci than in wild-type. Supporting this hypothesis, higher expression of eGFP-Vas$^{\Delta 636\text{-}646}$ enables many more *vas*^{1} embryos to produce pole cells, however the number of pole cells each embryo produces is highly variable, and on average is much lower than that for example in *vas*^{1}*; egfp-vas*$^{\Delta 15\text{-}75}$, which is expressed at a comparable level (4.3 ± 0.3 versus 9.9 ± 0.2; Fig. 2.S4). We did not observe a similar phenotype for eGFP-Vas^{R644A}, a point mutation that affects a potential arginine methylation site within the 636-646 interval.

Lastly we found that the conserved sequence at the C-terminus of Vasa is required for germ cell formation ($p = 0.013$; Fig. 2.6). *vas*^{1}*; eGFP-vas*$^{\Delta 655\text{-}661}$ embryos produce almost no germ cells despite the normal localization of eGFP-Vas to the posterior pole.

Posterior patterning activity of Vas is abolished by removing the entire N-terminal domain or by point mutations in conserved DEAD-box helicase domains.

vas^{1} embryos fail to hatch because they lack posterior segments and thus die. We tested the ability of the various eGFP-Vas constructs to rescue embryonic viability to *vas*^{1} embryos. Proteins with N-terminal deletions within the interval 15-127 differ little from wild-type eGFP-Vas in their ability to restore viability to *vas*^{1} embryos (Fig. 2.8 A). However, embryonic viability decreased about 37% by removing amino acids 3-139 ($p = 0.044$) and was almost completely abolished by a more extensive N-terminal deletion (Δ3-200). We also found that all

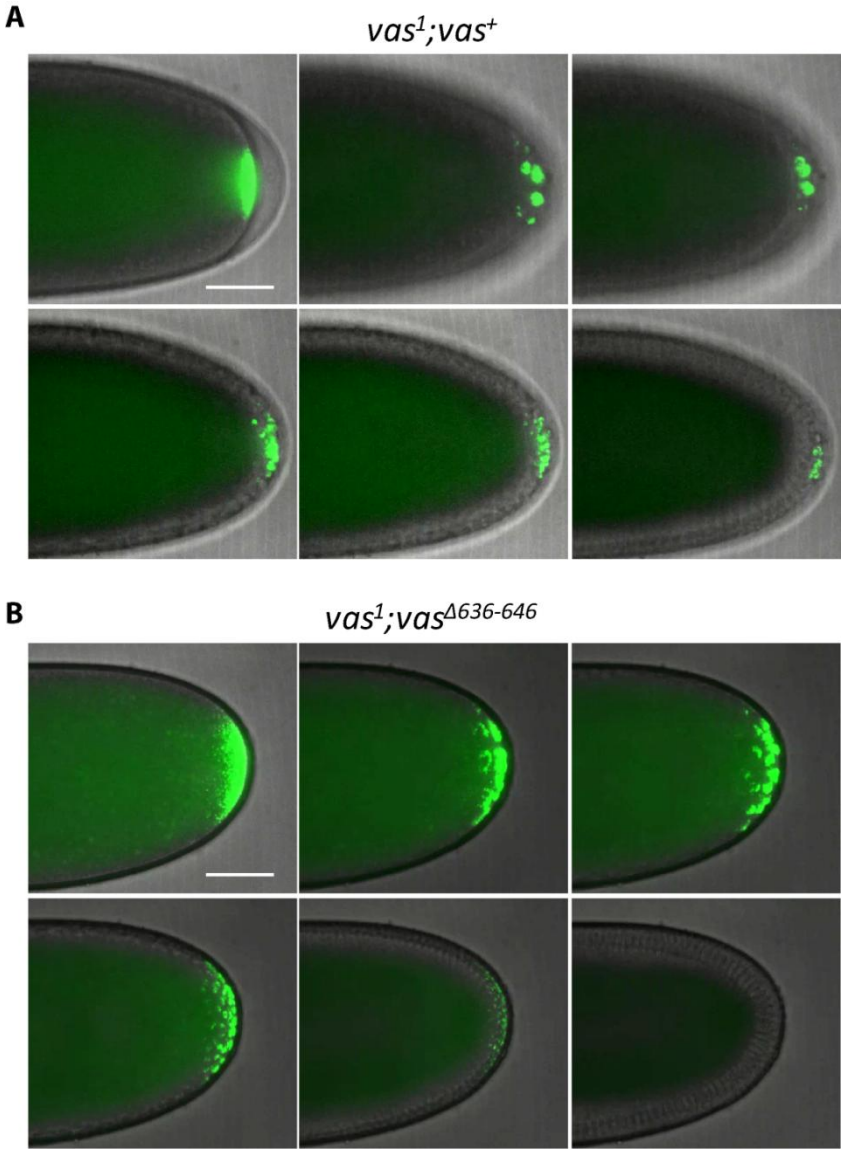


Fig. 2.7 Time course of pole cell development in eGFP-Vas^{+} and eGFP-Vas$^{\Delta 636\text{-}646}$ expressing embryos. **A**. A series of still shots from live imaging of pole cell formation in eGFP-Vas^{+} expressing embryos. eGFP-Vas is tightly localized at the posterior pole and then concentrates in foci within pole buds. Posterior nuclear divisions become asynchronous with somatic nuclear divisions, and eGFP-Vas positive pole cells completely form. **B**. A series of still shots from live imaging of pole cell formation in eGFP-Vas$^{\Delta 636\text{-}646}$ expressing embryos. eGFP-Vas is less tightly localized at the posterior pole and forms more foci than in wild type. Mitosis remains synchronous throughout the embryo. eGFP-Vas is then lost from the posterior region and pole cells fail to form. Scale bar= 50μm.

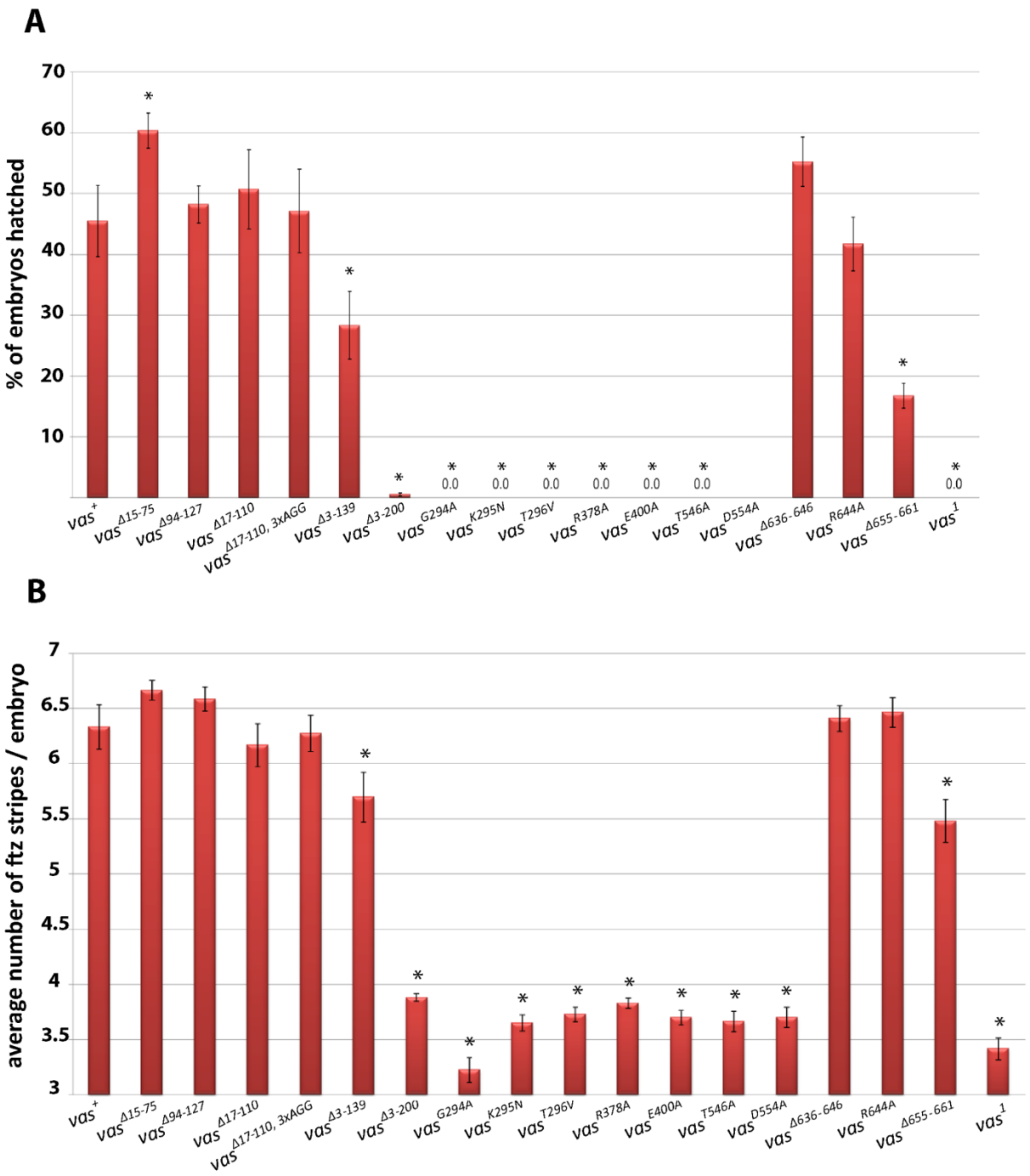


Fig 2.8 The ability of various *egfp-vas* constructs to restore abdominal segmentation in *vas*^{1} embryos. **A.** Hatching rates: The red bars indicate the percentage of embryos hatched after 48 hours, error bars indicate SEM from at least five plates. Between 500-1000 embryos in total were scored for each genotype. **B**. *ftz* expression in *vas*^{1} embryos containing various *egfp-vas* constructs: The y-axis indicates the average number of *ftz* stripes for each genoptype. Error bars show SEM calculated from more than 50 embryos examined for each genotype. In both A and B asterisks show a significant increase or decrease from *vas*$^{+}$ ($p<0.05$).

seven mutations in the conserved core motifs abrogated the function of Vas with regard to embryonic viability.

Interestingly, eGFP-Vas$^{\Delta 636\text{-}646}$, despite its inability to restore pole cell formation to *vas*^{1} embryos, was fully able to restore viability even through to adulthood; thus the *vas*^{1}*; egfp-vas*$^{\Delta 636\text{-}646}$ allelic combination produces a robust grandchildless phenotype, when the transgene is expressed at the level of the wild-type transgenic control. As in the other assays, eGFP-Vas^{R644A} was fully functional with respect to restoring viability to *vas*^{1} embryos, while eGFP-Vas$^{\Delta 655\text{-}661}$ rescued viability at a substantially lower rate than wild-type ($p = 0.0005$).

To determine whether these effects on viability correlate with defects in anterior-posterior patterning, we performed *in situ* hybridization for *ftz* mRNA on *vas*^{1} embryos expressing the different *egfp-vas* transgenes (Fig. 2.8 B). In wild-type blastoderm-stage embryos *ftz* mRNA is expressed in seven transverse stripes along the anterior-posterior axis, while in *vas*^{1} embryos *ftz* mRNA is usually expressed in only four stripes, with loss of stripes 4, 5, and 6 that correspond to future abdominal segments. For all N-terminal deletions within the interval 15-127, the average number of *ftz* stripes was above six and the embryos with seven stripes represented the most frequent class. However, the average number of *ftz* stripes in *vas*$^{\Delta 3\text{-}139}$ decreased to 5.69 ± 0.2 ($p = 0.037$) which was associated with about a 30% decrease in the number of embryos with seven stripes. For Δ3-200 and all mutations in conserved DEAD-box residues, similar to *vas*^{1} the average number of *ftz* stripes was less than four, and embryos with four *ftz* stripes were predominant, indicating a failure of these constructs to rescue posterior patterning and thus embryonic viability. Seven *ftz* stripes predominated in embryos expressing each of the constructs mutated at the C-terminal end, except for Δ655-661 where the average was 5.47 ± 0.19 ($p = 0.0028$) and in approximately 60% of these embryos fewer than seven *ftz* stripes were present.

Taken together, these results indicate that the conserved domains that Vas shares with other DEAD-box helicases are essential for posterior patterning, and that other residues in the intervals between amino acids 140-200 and 655-661 also contribute to this function.

Discussion

Mutant forms of Vas with abrogated DEAD-box motifs can support oogenesis and can localize to the posterior pole of the stage-10 oocyte, but RNA helicase activity of Vas is required for most of its cellular and developmental functions

This comprehensive analysis of the *in vivo* activities of a series of eGFP-Vas proteins allows us to draw several conclusions as to how its various functions are related to different motifs within the protein. Surprisingly, we found that most mutant eGFP-Vas proteins carrying non-conservative amino acid changes that would be expected to abrogate fundamental DEAD-box helicase functions such as ATP binding or RNA binding can nevertheless restore fecundity to a *vas*-null mutant. Furthermore, most of these same mutant proteins can localize to the posterior pole of the stage-10 vas^{1} oocyte. This argues that whatever role Vas performs to allow oogenesis to progress to completion, and also its posterior localization, does not require it to function as an RNA helicase. We speculate that these functions of Vas may be fulfilled through it operating as an inert scaffolding molecule. The exceptional mutations that do not localize or rescue oogenesis (T546A and D554A, respectively) both affect motif V. Consistent with this, an early study mapped vas^{D5}, an allele with a phenotype of similar severity as vas^{PH165} and that produces a protein that fails to localize, to a point mutation in motif V (Liang et al., 1994). Perhaps motif V is within an important region for a scaffolding function, or alternatively mutations in this domain produce alterations in protein folding that abrogate the ability of Vas to serve as a scaffold. Notwithstanding the above, however, most functions of Vas require it to be an active RNA helicase, as translational activation of *grk*, posterior patterning, pole cell specification, and repression of transposon-encoded mRNAs are all not supported by any mutant with alterations in the conserved DEAD-box motifs. We observe as well in many conserved DEAD-box residue mutants (and in Δ3-200) that the nuage appears punctate, and does not fully encircle the nurse cell nuclei. *vas* has previously been implicated in nuage formation in electron microscopy studies (Liang et al., 1994).

A highly acidic Vas-specific C-terminal motif is required for many Vas functions

At the extreme C-terminal end of Vas orthologues throughout the animal kingdom is found a 5-7 amino acid motif that contains several acidic amino acids and a tryptophan residue at the penultimate or ultimate position (Fig. 2.1 D). This motif is not conserved in other closely related DEAD-box proteins. The C-terminal seven amino acids of human DDX3 or *Drosophila* Belle, DEAD-box proteins that are most closely related to Vas, both include two consecutive tryptophan residues but include only one aspartate residue each, thus their C-terminal ends have substantially diverged from Vas and the human Vas orthologue DDX4. The C-terminal ends of *Drosophila* RM62 or human DDX5, the next most closely related DEAD-box proteins to Vas, include neither tryptophan residues nor acidic residues. In this study we show that this Vas-specific C-terminal motif is nearly as essential to its function *in vivo* as the conserved DEAD-box domains. The mutant form of eGFP-Vas deleted for this motif does not support translational activation of *grk*, pole cell specification, and its activity with respect to posterior patterning and repression of transposon-encoded mRNAs is clearly reduced. These results indicate that this C-terminal domain is a key feature of Vas that distinguishes it functionally from other closely related DEAD-box helicases.

We made two other noteworthy observations concerning the C-terminal region of Vas. First, we identified a motif (636-646) that is essential for fine-tuning the distribution of Vas at the posterior pole of the early embryo. When this motif is deleted, posterior localization of Vas is more diffuse, affecting pole cell specification. It is possible that Vas, depending on motif 636-646, plays a role in maintaining integrity of pole plasm, and thus ensuring sufficient concentration of germ cell components at the posterior region. Indeed our further analyses indicate that localization of $Vas^{\Delta 636\text{-}646}$ in a wild type background is comparable to Vas^{+}, suggesting that deletion of motif 636-646 primarily affects Vas function, which in a vas^{1} background will ultimately result in the protein mislocalization. Moreover, when $Vas^{\Delta 636\text{-}646}$ is expressed at high enough levels it can induce germ cell formation, indicating that it is not inherently defective for this function. Second, we found that deletion of amino acid 617, previously shown to affect the Vas-eIF5B interaction, also results in a failure to repress transposon-encoded mRNAs. This could suggest that the role of Vas in transposon silencing

involves translational regulation, but the accumulation of Vas in the nuage and the direct association of Vas with piRNAs support a more direct role (Xiol et al., 2014; Zhang et al., 2012). We speculate that deletion of amino acid 617 disrupts not only the eIF5B interaction but another interaction as well that is involved with piRNA processing.

The rapidly-evolving N-terminal region contributes to Vas function

After the first eight or nine amino acids at the extreme N-terminus, which are highly conserved among Vas orthologues in different Drosophilids, there is a rapidly evolving segment of the protein that extends for approximately the next 170 amino acids in *D. melanogaster* (Fig. 2.1 A). Strong sequence conservation among species begins again near the DINNN motif (amino acids 184-188) that binds to two Cullin-RING E3 ligase specificity receptors, Gustavus (Gus) and Fsn (Kugler et al., 2010). We found that eGFP-Vas proteins with deletions within the region 15-110 are able to support all functions of Vas that we examined. Analysis of more extensive deletions, however, reveals several roles for the N-terminal domain. Deletions affecting the region between amino acids 110-139 produce inconsistent effects on transposon mRNA silencing, suggesting that these forms of Vas are close to a threshold level of activity with respect to this function, and that this region of the protein has at least an accessory role in piRNA processing. Rapid evolution of some germ line-expressed genes must contribute to the reproductive isolation of species, and derepression of transposon activity has been observed in *D. melanogaster/D. simulans Hmr/+* hybrid females (Kelleher et al., 2012). It is interesting in this context that a segment of the rapidly evolving region of Vas might be implicated in transposon silencing, as it suggests a potential mechanism underlying hybrid sterility.

Complete deletion of the N-terminal region (Δ3-200) abrogates the ability of eGFP-Vas to support posterior patterning and pole cell specification, while eGFP-Vas$^{\Delta 3\text{-}139}$ supports these functions to some extent. The larger deletion includes the DINNN motif, which may contribute to these functions. However, as *gus* and *Fsn* mutations do not completely abrogate posterior patterning and pole cell specification (Kugler et al., 2010), other motifs probably contribute as well. Our analysis also indicates that the integrity of three RGG motifs in Vas$^{\Delta 17\text{-}110}$ is necessary for its function in pole cell specification.

Our results indicate that different functions of Vas require different domains of the protein, which in turn implies that its different functions involve a diversity of molecular associations. DEAD-box helicases often function as components of ribonucleoprotein particles (RNPs), and it is likely that each mutation in eGFP-Vas which affects particular functions might also affect its ability to assemble into some RNPs but not into others. The eGFP-Vas transgenes that we have generated open the door for future proteomic analysis that could address this issue and link specific molecular associations with cellular and developmental functions.

Acknowledgments

We are grateful to Ming-Der Lin and Chung-Che Chang for helpful discussions and for sharing unpublished data and materials. We also thank Sanjay Ghosh for technical advice and helpful discussions, and Beili Hu for microinjection into the embryos. Imaging was performed using the Cellular Imaging and Analysis Network (CIAN) facility. This work was supported by NSERC grant RGPIN 06340-14 and by a Tomlinson Science Award from the McGill University Faculty of Science to P. L.

Supplementary material

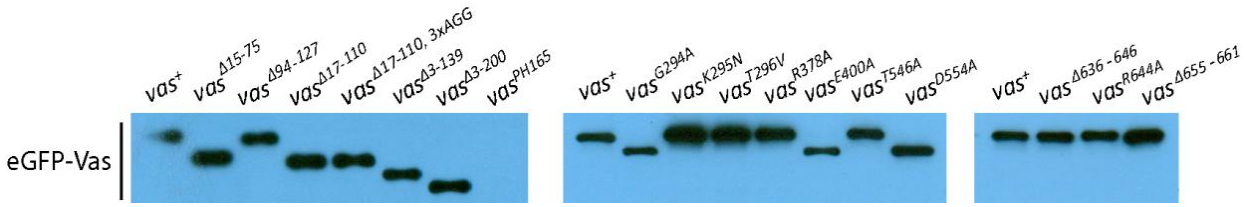


Fig. 2.S1 The same western blot as in Fig. 2.2B stained with anti-GFP to compare expression levels of eGFP-Vas proteins in ovaries from *vas*^{1}/+ females carrying the different constructs.

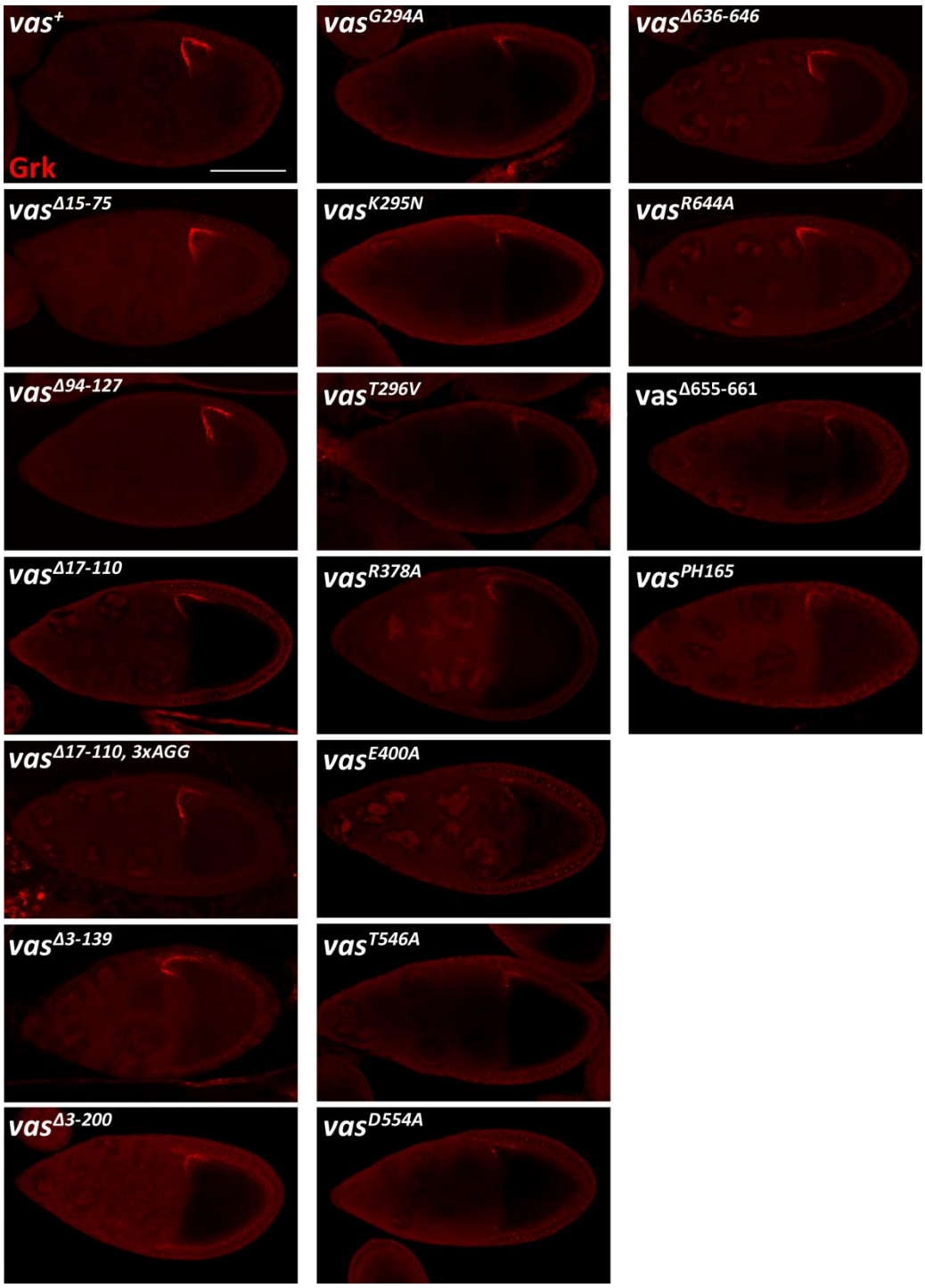


Fig. 2.S2 Representative confocal images of Grk immunostaining in the subset of stage 8 egg chambers that were positively stained in each genotype. Scale bar=50 μm

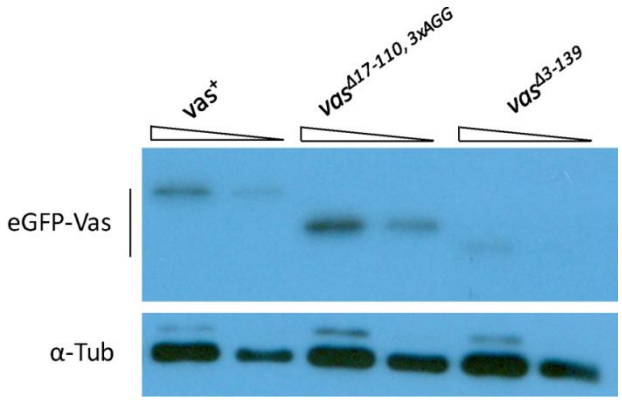


Fig. 2.S3 A comparison between eGFP-Vas^{+}, eGFP-Vas$^{\Delta 17\text{-}110,\ 3xAGG}$ and eGFP-Vas$^{\Delta 3\text{-}139}$ for their stability in 0-2 h embryos from *vas*^{1}/+ females. Top panel; western blot using a GFP antibody. Bottom panel: the same blot was stained for α-tubulin as a loading control. In the rightmost lane of each pair half as much lysate was loaded. The low level of eGFP-Vas$^{\Delta 3\text{-}139}$ in embryos suggests that this protein may be unstable.

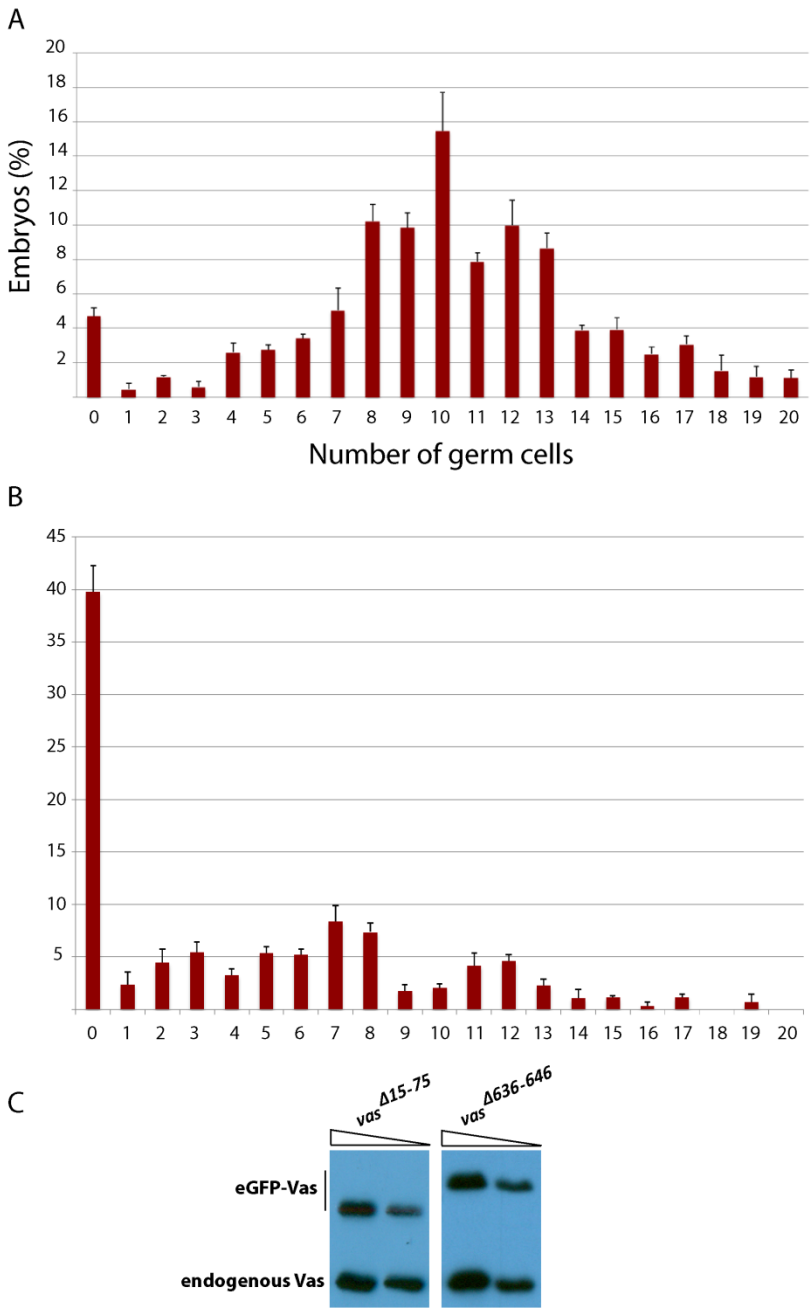


Fig. 2.S4 *vas*^{1}*; egfp-vas*$^{\Delta 15\text{-}75}$ and *vas*^{1}*; egfp-vas*$^{\Delta 636\text{-}646}$ are compared for their variability in the number of pole cells per stage 4-5 embryos. The numbers of pole cells have a normal distribution between *vas*^{1}*; egfp-vas*$^{\Delta 15\text{-}75}$ embryos (**A**), whereas in a *vas*^{1}*; egfp-vas*$^{\Delta 636\text{-}646}$ line (**B**) that has a similar expression level of eGFP-Vas, the number of pole cells in embryos is highly variable with a bias toward zero. Error bars represent SEM from three replicates with more than 50 embryos each. (**C**) Western blots from 0-2 h embryos compare expression levels of eGFP-Vas in *egfp-vas*$^{\Delta 15\text{-}75}$ and egfp-*vas*$^{\Delta 636\text{-}646}$ lines. In the rightmost lane of each pair half as much lysate was loaded. A Vas antibody was used to detect both the eGFP-Vas and the endogenous Vas.

Table 2.S1 Results of the statistical analyses: The p-values calculated in each assay are listed. Different genotypes were compared to *vas*^{PH165} for egg-laying assay, and to *egfp-vas*$^{+}$ for all the other assays.

	vas^{+}	*vas*$^{\Delta 15\text{-}75}$	*vas*$^{\Delta 94\text{-}127}$	*vas*$^{\Delta 17\text{-}110}$	*vas*$^{\Delta 17\text{-}110,\ 3xAGG}$	*vas*$^{\Delta 3\text{-}139}$	*vas*$^{\Delta 3\text{-}200}$	*vas*^{G294A}
Number of eggs produced/female[1]	1.05E-07	1.22E-05	1.13E-11	1.43E-07	1.64E-07	9.71E-09	8.46E-10	6.44E-08
Embryos with two dorsal appendages (%)[2]	—	0.0910	7.78E-08	5.91E-08	0.0047	0.0001	1.50E-12	1.07E-20
Stage 8 egg chambers with Grk crescent (%)[2]	—	0.0473	0.8568	0.4981	0.9018	0.2094	0.0291	0.0037
HeT-A overexpression[2]	—	0.3878	0.2320	0.5582	0.4822	0.1823	0.0039	0.0402
Embryos with germ cell (%)[2]	—	0.0262	0.5819	0.7785	0.0221	0.0181	0.0128	0.0128
Embryonic viability (%)[2]	—	0.0390	0.6891	0.5599	0.8610	0.0443	1.66E-05	1.52E-05
Average number of stripes/embryo[2]	—	0.1402	0.2807	0.5552	0.8248	0.0376	9.33E-18	2.29E-23

1: compared to *vas*^{PH165}

2: compared to *vas*$^{+}$

	vas^{K295N}	*vas*^{T296V}	*vas*^{R378A}	*vas*^{E400A}	*vas*^{T546A}	*vas*^{D554A}	*vas*$^{\Delta 617}$	*vas*$^{\Delta 636-646}$
Number of eggs produced/female[1]	1.64E-07	0.0002	1.86E-06	0.0001	6.32E-08	0.0846	—	9.28E-08
Embryos with two dorsal appendages (%)[2]	6.64E-10	3.24E-20	1.48E-21	1.71E-17	1.70E-20	4.29E-15	—	2.81E-06
Stage 8 egg chambers with Grk crescent (%)[2]	0.0025	0.0043	0.0009	0.0045	0.0041	0.0043	—	0.7973
HeT-A overexpression[2]	0.0237	—	—	—	—	0.0101	0.0024	0.7551
Embryos with germ cell (%)[2]	0.0128	0.0128	0.0128	0.0128	0.0128	0.0128	—	0.0130
Embryonic viability (%)[2]	1.52E-05	1.52E-05	1.52E-05	1.52E-05	1.52E-05	1.52E-05		0.1930
Average number of stripes/embryo[2]	7.01E-20	3.28E-19	3.20E-18	1.89E-19	1.05E-19	2.13E-19	—	0.7425

	vas^{R644A}	*vas*$^{\Delta 655-661}$	*vas*^{PH165} or *vas*^{1}	*vas*^{PH165}/*Df(2L)A267*
Number of eggs produced/female[1]	1.04E-09	3.95E-10	—	0.8269
Embryos with two dorsal appendages (%)[2]	4.13E-15	8.63E-19	3.07E-15	—
Stage 8 egg chambers with Grk crescent (%)[2]	0.2671	0.0018	0.0037	—
HeT-A overexpression[2]	—	0.1422	0.0216	0.0239
Embryos with germ cell (%)[2]	0.8308	0.0131	0.0128	—
Embryonic viability (%)[2]	0.6114	0.0005	1.52E-05	—
Average number of stripes/embryo[2]	0.2970	0.0028	7.62E-22	—

Appendix A

A.1 Mapping the structural domain in the N-terminus of Vas for sequence-specific binding to RNA

Despite the extensive knowledge of Vas, little is known about its mRNA targets and their recognition motifs. To approach this issue, mRNAs co-purified with Vasa from extracts of the early *Drosophila* embryos were screened (Liu et al., 2009). One of the identified mRNAs was *mei-P26*. Further analyses indicated that a (U)-rich motif in the *mei-P26* 3′ UTR is necessary for its translational activation by Vas. Although transcripts bearing a polyuridine motif were not over-represented among the recovered mRNAs, the number of pole cell-enriched transcripts among them was four to five-fold higher than that in a random sampling (Lecuyer et al., 2007; Tomancak et al., 2007), indicating that Vas may specifically recognize its targets by the presence of a (U)-rich motif in their 3′ UTR. This uncovers a novel RNA binding mechanism for Vasa, which was previously thought to be non-sequence specific in its interaction with RNA. Nonetheless, the Vas domain(s) responsible for its specific interactions with the mRNA still remains unknown. Previous studies suggest that the specificity arises from the part of the Vas sequence outside of the conserved motifs among DEAD-box proteins (Liu et al., 2009; Sengoku et al., 2006).

One hypothesis is that the sequence-specific binding is associated with one or more RGG motifs in the N-terminus of Vas. The RGG box, defined as closely spaced Arg-Gly-Gly repeats, often mediates RNA-protein or protein-protein interactions (reviewed by Thandapani et al., 2013). In addition, RGG motifs could contribute to sequence specificity of RNA binding proteins, the best example of which is the Fragile-X Mental Retardation Protein (FMRP) which binds a G-quartet in its mRNA targets via the RGG box (Darnell et al., 2001; Lamm et al., 1996).

To investigate this possibility, I generated glutathione S-transferase (GST)-Vas constructs with progressive deletions in their N-terminus that eliminate different numbers of RGG domains. Fusion proteins, produced from these constructs, were used in a cross-linking assay to study their interaction with fragments of the *mei-P26* 3′ UTR. I predicted that some of these constructs would lose their ability to bind the (U)-rich fragment from the *mei-P26* 3′ UTR, and thus uncover the region(s) responsible for the sequence-specific binding.

Plasmid construction and UV cross-linking assays

Different constructs of *vas* were generated encoding various lengths of the N-terminus (N) plus the Vas core protein (VCP: amino acids 201-623). These fragments were amplified from a *vas* coding sequence (CDS) clone, which lacks one copy of a 39 nucleotide tandem repeat, including one RGG motif, but which has been previously used successfully to rescue *vas*-null defects (Johnstone and Lasko, 2004). A GST-tag was also included at the N-terminus of each construct. As controls, GST-Vas^{FL} (containing the full-length Vas) and GST-VCP fusions were prepared from the previously generated constructs (Liu et al., 2009).

UV cross-linking assays were performed according to Liu et al. (2009). Each protein construct was tested against four different probes which were 30-nt fragments derived from the *mei-P26* 3′ UTR and radiolabelled by ^{32}P:
521-550; UUAACUACGAAUACGUGAAAAUGAAAAGGU,
551-580; CGAGC UCCUAUUUUUUUUUUCCCCGUUUAU,
581-610; CUGUAUAGUCCCCUAUCAACGUUCGAAUUC
and 611-640; UUAAAGCUGAUCAAAUCUGCGGCUGCACAG. Purified GST was tested as a negative control in the cross-linking assays.

Results: the *in vitro* assays map a sequence-specific RNA binding motif to amino acids 183-200 of Vas.

In the first round of experiments six GST-Vas fusion proteins including Vas^{FL} and VCP were bacterially expressed and purified (Fig. A.1 A, B). The UV cross-linking assays using these proteins show that N+VCP, $N^{\Delta 1-34}$+VCP , $N^{\Delta 1-88}$+VCP and $N^{\Delta 1-153}$+VCP fusions all have a stronger affinity to the (U)-rich oligonucleotide than the other three probes; this is similar to the pattern observed for Vas^{FL} (Fig. A.1 C). Consistent with the previously published data the VCP did not display a preferential binding to the (U)-rich fragment (Liu et al 2009). Furthermore, a basal affinity to all four RNA probes that is observed for $N^{\Delta 1-153}$ + VCP and VCP possibly reflects the non-sequence specific RNA binding of the core helicase domain, which becomes

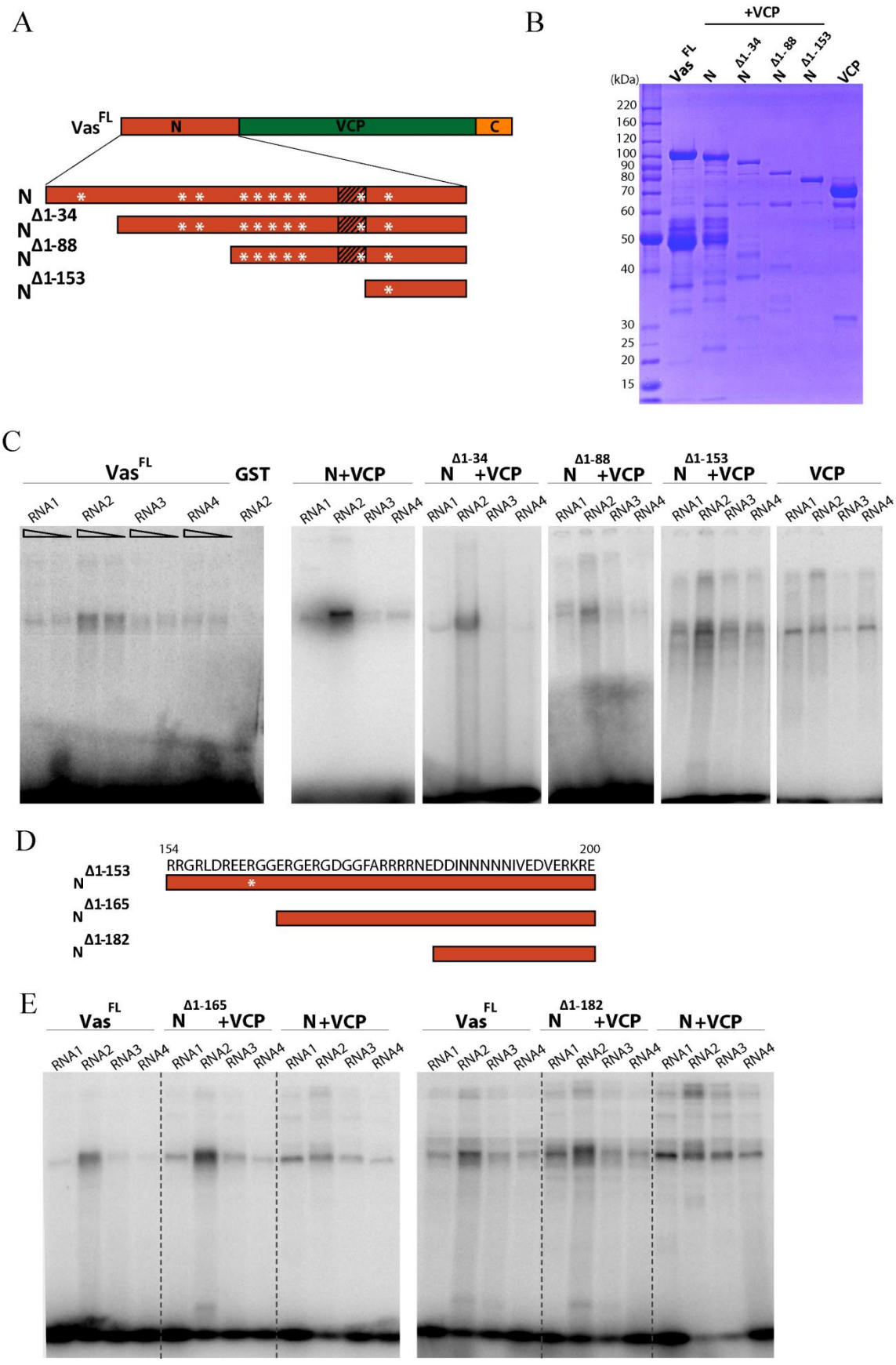

A
Vas FL
N
VCP
C
N
N Δ1-34
N Δ1-88
N Δ1-153
B
+VCP
Vas FL
N
N Δ1-34
N Δ1-88
N Δ1-153
VCP
(kDa)
220
160
120
100
90
80
70
60
50
40
30
25
20
15
C
Vas FL
GST
N+VCP
N Δ1-34 +VCP
N Δ1-88 +VCP
N Δ1-153 +VCP
VCP
RNA1
RNA2
RNA3
RNA4
D
154
200
RRGRLDREERGGERGERGDGGFARRRRNEDDINNNNNNIVEDVERKRE
N Δ1-153
N Δ1-165
N Δ1-182
E
Vas FL
N Δ1-165 +VCP
N+VCP
Vas FL
N Δ1-182 +VCP
N+VCP
RNA1
RNA2
RNA3
RNA4

Fig. A.1 Mapping sequence-specific RNA-binding domain in Vas through an *in vitro* cross-linking assay, **A**. The N-terminally tagged GST fusion proteins contained the full length Vas (Vas^{FL}), the minimum Vas core protein (VCP) or a fragment composed of VCP and different lengths of Vas N-terminus (N+VCP, $N^{\Delta 1-34}$+VCP, $N^{\Delta 1-88}$+VCP and $N^{\Delta 1-153}$+VCP). Asterisks show positions of the RGG motifs. The hashed box indicates a tandem repeat composed of 13 amino acids, including an RGG motif, which is missing in all GST-Vas constructs. **B**. Bacterially expressed GST-Vas fusions, after purification, were resolved by SDS PAGE and visualized by Coomassie staining. **C**. Each GST-Vas protein was tested through cross-linking assays against four different RNA probes derived from *mei-P26* 3′ UTR (Liu et al., 2009): RNA1; 521-550, RNA2; 551-580, RNA3; 581-610 and RNA4; 611-640. A stretch of nine uridines is located within nucleotides 551-580 of *mei-P26* 3′ UTR. The preferential binding of GST-Vas persists upon deleting a.a. 1-153, but disappears in GST-VCP. As a negative control GST was tested against RNA2. Non-specific binding of GST-Vas proteins to RNA1, RNA3 and RNA4 increases by deleting a.a. 1-153 ($N^{\Delta 1-153}$+VCP and VCP). Rectangles, each, correspond to two cross-linking reactions containing 3 or 1.5 μg of $GST\text{-}Vas^{FL}$. **D**. The schematics depict N-terminal amino acids in $N^{\Delta 1-153}$+VCP and the deletions in $N^{\Delta 1-165}$+VCP and $N^{\Delta 1-182}$+VCP. **E**. $N^{\Delta 1-165}$+VCP and $N^{\Delta 1-182}$+VCP, unlike GST-VCP, retain a preferential affinity to the (U)-rich RNA2.

progressively detectable *in vitro* by N-terminal truncations. In these assays one puzzling observation was that all the fusion proteins smaller than $N^{\Delta 1\text{-}34}$+VCP, once cross-linked, migrated similarly to the larger polypeptides, visualized on the SDS-PAGE by Coomassie staining. Similar assays showed that a histidine (His)-tagged $N^{\Delta 1\text{-}88}$+VCP also migrates higher than the expected size, and ruled out the possibility that GST-dimerization is responsible for this effect (data not shown).

To further map the sequence responsible for preferential binding of Vas to the (U)-rich motif, I generated two additional constructs: $N^{\Delta 1\text{-}165}$+VCP and $N^{\Delta 1\text{-}182}$+VCP (Fig. A.1 D). The first one lacks all the RGG motifs and, in addition to that, the second one lacks a potential RNA binding motif composed of four arginines. The UV cross-linking assays comparing these two proteins with Vas^{FL} and VCP indicate that the preferential affinity of Vas to the (U)-rich motif persists upon removing the first 182 amino acids including all of the RGG motifs (Fig. A.1 E). The N-terminal region of Vas proximal to VCP contains more conserved residues among different Vas orthologs, compared to the upstream sequence. For example, a five-amino-acid motif in this region (184-188; DINNN) is required for binding of Gustavus (Gus) and Fsn, two ubiquitin Cullin-RING E3 ligase specificity receptors, to Vas and for regulation of its stability (Kugler et al. 2010). Fine-mapping of the residues within a.a. 183-200, that are important for specific Vas-RNA binding, awaits analyses of additional constructs with deletions in this region.

A.2 *In vivo* mapping of functional domains in the N-terminus of Vas using a site-specific transgenesis system

To exclude variations in gene expression caused by the insertion position of the transgene, I used a PhiC31 mediated site-specific integration method and a unique landing site (*ZH-attP-86Fb*) for generating different *egfp-vas* alleles (Fig. A.2 A; Bischof et al., 2007). However, this system proved inefficient for the functional analyses that involve protein truncation. eGFP-$Vas^{\Delta 17\text{-}110}$ and other eGFP-Vas fusions, with the same size or smaller, were expressed at lower levels than eGFP-Vas^{+}, detected on western blots (Fig. A.2 B). This could be due to decreased

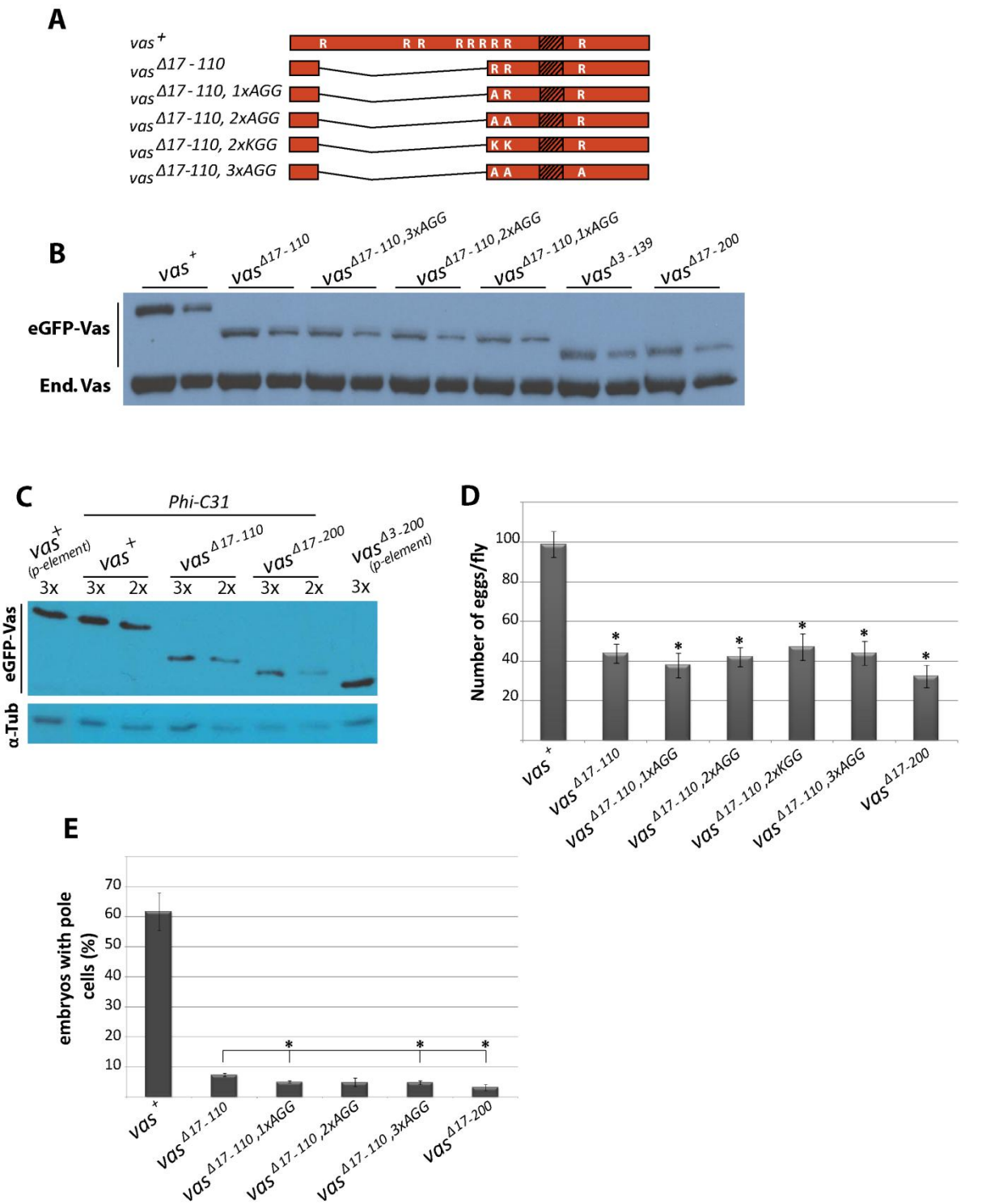

A
vas+
vasΔ17-110
vasΔ17-110, 1xAGG
vasΔ17-110, 2xAGG
vasΔ17-110, 2xKGG
vasΔ17-110, 3xAGG
R
R R
R R R R R
A R
A A
K K
A A
A
B
vas+
vasΔ17-110
vasΔ17-110,3xAGG
vasΔ17-110,2xAGG
vasΔ17-110,1xAGG
vasΔ3-139
vasΔ17-200
eGFP-Vas
End. Vas
C
Phi-C31
vas+ (p-element)
vas+
vasΔ17-110
vasΔ17-200
vasΔ3-200 (p-element)
3x
3x 2x
3x 2x
3x 2x
3x
eGFP-Vas
α-Tub
D
Number of eggs/fly
100
80
60
40
20
vas+
vasΔ17-110
vasΔ17-110,1xAGG
vasΔ17-110,2xAGG
vasΔ17-110,2xKGG
vasΔ17-110,3xAGG
vasΔ17-200
E
embryos with pole cells (%)
70
60
50
40
30
20
10
vas+
vasΔ17-110
vasΔ17-110,1xAGG
vasΔ17-110,2xAGG
vasΔ17-110,3xAGG
vasΔ17-200

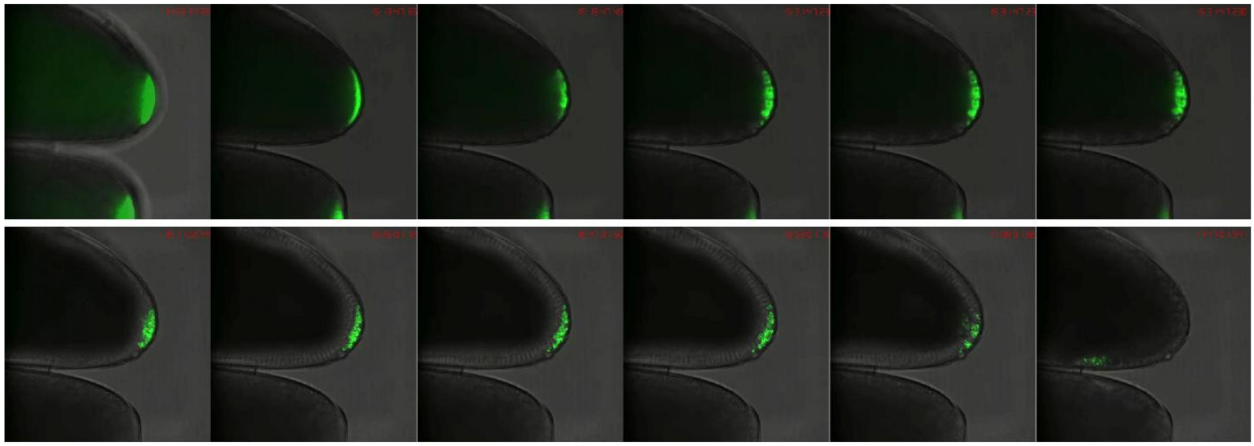
Fig. A.2 Functional comparisons between different *egfp-vas* alleles substituting RGG with AGG or KGG **A**. Schematics of the N-terminal region in vas^{+}, $vas^{\Delta 17\text{-}110}$, $vas^{\Delta 17\text{-}110,1xAGG}$, $vas^{\Delta 17\text{-}110,2xAGG}$, $vas^{\Delta 17\text{-}110,2xKGG}$ and $vas^{\Delta 17\text{-}110,3xAGG}$. Hashed box indicates a 13 amino acid tandem repeat that is absent in all constructs. R, A and K indicate RGG, AGG and KGG, respectively. **B**. Expression levels of eGFP-Vas proteins were compared between different transgenic alleles established using a site-specific integration system: *egfp .vas*$^{+}$, *.vas*$^{\Delta 17\text{-}110}$, *.vas*$^{\Delta 3\text{-}130}$, *.vas*$^{\Delta 17\text{-}200}$ and three alleles encoding eGFP-Vas$^{\Delta 17\text{-}110}$ proteins with AGG replacing RGG motifs. eGFP-Vas and endogenous (end.) Vas are both detected on the western blots by a Vas antibody. For all the alleles tested the ovarian lysates were prepared from vas^{1}/+ females carrying two copies of *egfp-vas* except for vas^{+} with only one copy. **C**. Western blot using a GFP antibody compares expression levels of eGFP-Vas from vas^{+}, $vas^{\Delta 17\text{-}110}$ and $vas^{\Delta 17\text{-}200}$ alleles generated by Phi-C31 integration, versus vas^{+} and $vas^{\Delta 3\text{-}200}$ alleles generated by P-element transgenesis methods. The fold differences (2x or 3x) in the total proteins loaded are shown at the top. All eGFP-Vas fusion proteins are tested from two copies of the transgene. **D**. Different *egfp-vas* constructs, integrated in the same genomic site, were compared for their ability to increase the number of eggs produced by vas^{PH165} females. Asterisks indicate a significant difference from $egfp.vas^{+}$. Error bars represent SEM. Number of females tested (n) ≥20. **E**. Some of the same *egfp-vas* alleles shown in D were compared for their ability to induce germ cell formation. All constructs were tested in a vas^{1} background carrying an additional copy of $egfp.vas^{+}$ inserted in a different region within genome. Each bar indicates the average from three replicates.

stability as a result of protein truncation. By using a *p-element* transgenesis, on the other hand, one could screen through multiple lines and select those that give an optimal expression of the transgene (Fig. A.2 C). For this reason, P-element transgenesis was used to generate the majority of transgenic alleles in this project, unless specified otherwise. Among the site-specifically integrated *egfp-* alleles $vas^{\Delta 17\text{-}110}$, $vas^{\Delta 17\text{-}110,1xAGG}$, $vas^{\Delta 17\text{-}110,2xAGG}$, $vas^{\Delta 17\text{-}110,2xKGG}$ and $vas^{\Delta 17\text{-}110,3xAGG}$ expressed similar levels of eGFP-Vas. Thus these alleles were tested, in a vas^{PH165} background, to specifically examine the role of RGG motifs for Vas function in female fecundity (Fig. A.2 D). Although removal of a.a. 17-110 significantly decreased the number of embryos (P<4.0E-06), later analyses with a different $vas^{\Delta 17\text{-}110}$ allele, which was comparable to vas^{+} for the level of eGFP-Vas, indicated that this effect is largely associated with the low expression of Vas (Fig. 2.2; Fig. A.2 A, B). This experiment also shows that replacing RGG motifs in $Vas^{\Delta 17\text{-}110}$, non-conservatively, with AGG or, conservatively, with KGG does not have a significant effect on female fecundity, consistent with the results from similar analyses using $vas^{\Delta 17\text{-}110}$ and $vas^{\Delta 17\text{-}110,\ 3AGG}$, generated by P-element transgenesis (Fig. 2.2). Further examination in a vas^{1} background indicated that *egfp-*$vas^{\Delta 17\text{-}110}$ and its counterparts bearing KGG or AGG, unlike *egfp-*vas^{+}, produce embryos that lack germ cells and are not viable. The insufficient levels of Vas, however, could account for this result; thus to follow up, I examined these constructs in a sensitized background where vas^{1} females carried an additional copy of *egfp.*vas^{+} in a different location of the genome. By this approach 7.3% of *egfp-*$vas^{\Delta 17\text{-}110}$ embryos produced germ cells, in contrast to 5% or less of the embryos carrying $vas^{\Delta 17\text{-}110}$ with AGG motifs (P<0.05 for $vas^{\Delta 17\text{-}110,1xAGG}$ and $vas^{\Delta 17\text{-}110,3xAGG}$; Fig. A.2 E). This experiment indicates that substitution of the first RGG motif in $Vas^{\Delta 17\text{-}110}$ has a similarly significant effect on germ cell formation as substitution of all three. To confirm these results additional experiments are required using transgenic lines with higher expression levels of eGFP-Vas.

A.3 Further analysis of $vas^{\Delta 636\text{-}646}$ for its defects in germ cell formation

$Vas^{\Delta 636\text{-}646}$, when examined in vas^{1} background, displays a less condensed localization in the posterior region compared to Vas^{+}. This is associated with the failure of pole buds in these mutants to develop into germ cells. To investigate if the abnormal localization of $Vas^{\Delta 636\text{-}646}$ is the primary reason for these phenotypes, versus inherent defects in protein function, I examined

A



B

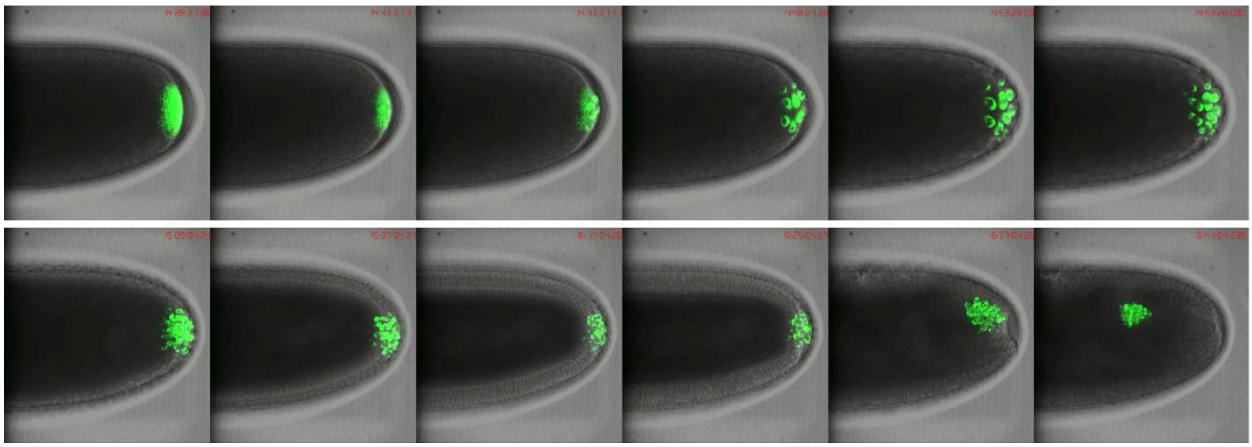


Fig. A.3 When examined in a *wt* background eGFP-Vas$^{\Delta 636\text{-}646}$ segregation into the germ cells and its stability is comparable to eGFP-Vas$^{\Delta 15\text{-}75}$ (*wt* control). Still images demonstrate germ cell formation at different time points in *vas*^{1}/+ embryos carrying *egfp.vas*$^{\Delta 15\text{-}75}$(**A**) or *egfp.vas*$^{\Delta 636\text{-}646}$ (**B**).

germ cell formation through live imaging of embryos expressing *egfp-vas*$^{+}$ or *egfp-vas*$^{\Delta 636\text{-}646}$ in a *wt* background (Fig. A.3). These analyses indicated that localization of Vas$^{\Delta 636\text{-}646}$ and its segregation into the germ cells, in the presence of a *wt* copy of *vas* to support pole plasm assembly, is similar to Vas$^{\Delta 15\text{-}75}$ (*wt* control). Thus diffusion of Vas$^{\Delta 636\text{-}646}$ in the posterior region of *vas*^{1} embryos is caused by its defects possibly in maintaining pole plasm integrity, which in turn is essential for sufficient concentration of germ cell components at the posterior region.

vas^{1} embryos expressing *egfp-vas*$^{\Delta 636\text{-}646}$ produce very few germ cells compared to *egfp-vas*$^{+}$ (Fig. 2.6, 2.7); however, abdominal segmentation and embryonic viability of these mutants, are comparable to the *wt* control (Fig. 2.8). To examine localization pattern of a few pole plasm components, that are required for germ cell specification but are dispensable for posterior patterning, I performed *in situ* hybridization for *polar granule component* (*pgc*) and *germ cell-less* (*gcl*) mRNA and immunostaining for Tudor (Tud) protein (Jongens et al., 1994; Nakamura et al., 1996; Thomson and Lasko, 2004). These experiments show that *pgc*, *gcl* and Tud localize to the pole plasm of *vas*$^{\Delta 636\text{-}646}$ embryos and their localization persists into the pole buds of *vas*$^{\Delta 636\text{-}646}$, as also observed in *wt* embryos (Fig. A.4 A, B, C). This experiment excludes the possibility that an initial failure of *gcl*, *pgc* or Tud localization is responsible for the inability of *vas*$^{\Delta 636\text{-}646}$ mutants to form germ cells. In the pole plasm *osk* mRNA localizes and its protein acts upstream of Vas (Breitwieser et al., 1996; Ephrussi and Lehmann, 1992). However, Vas is essential for stable accumulation of Osk protein in the posterior region (Ephrussi and Lehmann, 1992). Vas may play this role directly through translational regulation of Long Osk, which is required for anchoring both *osk* mRNA and Short Osk, or indirectly through its functions in pole plasm assembly (Vanzo and Ephrussi, 2002). Consistent with the role of Vas in stable localization of Osk, in *vas*$^{\Delta 636\text{-}646}$ embryos Osk and Vas colocalize in a similarly diffused domain at the posterior cortex, which is distinct from that in *vas*$^{+}$ (Fig. A.4 D).

Taken together these observations suggest that in *vas*$^{\Delta 636\text{-}646}$ embryos pole plasm integrity is not stably maintained. This causes germ plasm components, including Vas itself, to gradually diffuse along the posterior cortex. Diluted amounts of pole plasm that segregate into the posterior pole buds in these mutants are insufficient for germ cell determination, and thus mitotic divisions in the posterior remain synchronous with the soma. In contrast to germ cell formation, abdominal patterning, which depends on a posterior anterior gradient of Nos, appears not to be affected by a

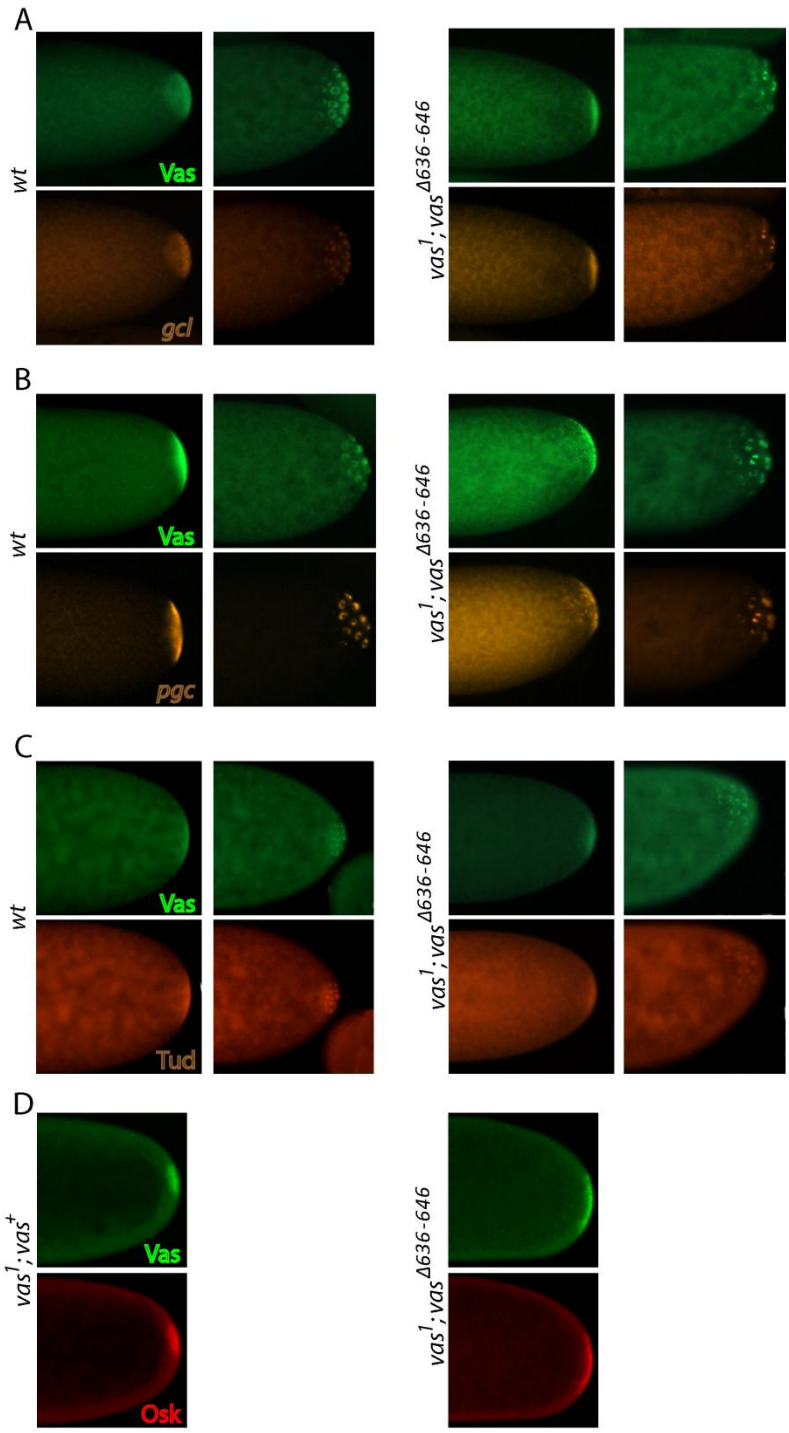


Fig. A.4 Localization of germ cell components in the posterior region of *vas*$^{\Delta 636\text{-}646}$ embryos. *gcl* (**A**) and *pgc* (**B**) mRNA and Tud protein (**C**) localize to the pole plasm and pole buds of *vas*^{1}*; egfp-vas*$^{\Delta 636\text{-}646}$ stage 2 and 3 embryos, respectively. *wt* (*OrR*) embryos are used as a control for *gcl* and *pgc in situ* hybridization and Tud immunostaining. **D**. Osk protein exhibits a diffuse localization pattern similar to Vas itself at the posterior region of *vas*^{1}*; egfp-vas*$^{\Delta 636\text{-}646}$ early embryos. *vas*^{1}*; egfp-vas*$^{+}$ serves as a positive control.

slightly more diffuse pole plasm. The germ cell specific defects of *vas*$^{\Delta 636\text{-}646}$ also suggest that the role of motif 636-646 is limited to particular interactions in the pole plasm, perhaps in association with the cortical microtubules, motor proteins and actin cytoskeleton, which are implicated in stable anchoring of germ plasm to the posterior cortex (Sinsimer et al., 2013).

Connecting text

Data presented in chapter 2, provides a comprehensive analysis examining numerous motifs and residues in Vas sequence with respect to the protein functions in oogenesis, germ cell development and embryonic posterior patterning. Among these elements there is a seven amino acids long motif at the end of Vas sequence containing five acidic residues, which in different numbers are often found at the end of Vas orthologs. This motif, covering amino acids 655-661, also includes a highly conserved tryptophan, Trp660, in the penultimate position. Our analyses showed that deletion of motif 655-661 abolishes germ cell formation and has severe effects on Grk translation and posterior patterning. To specifically study the role of Trp660 in Vas functions, I generated different endogenous and transgenic alleles of *vas* encoding non-conservative or conservative substitutions for this residue, and studied their functions *in vivo*. Furthermore, I extended our analyses on $vas^{\Delta 655\text{-}661}$ and vas^{W660E} to examining the protein interactions between Vas and its different partners. Results of this study are presented in chapter 3 and appendix B.

Chapter 3: Studying the functional role of an invariant tryptophan residue in the C-terminus of Vas

Manuscript to be submitted as:

Mehrnoush Dehghani and Paul Lasko (2016)

Abstract

The DEAD-box RNA helicase Vasa (Vas) has been implicated in germ cell development in many species. In *Drosophila*, females homozygous for a hypomorphic allele, *vas*^{1}, produce embryos which do not form germ cells and are not viable due to the defects in abdominal segmentation. Furthermore, *vas*-null females, which rarely complete oogenesis, exhibit defects in mitotic progression of germline stem cells, transposon silencing and Gurken (Grk) translation. In addition to the conserved helicase core, Vas contains N- and C-terminal sequences, which vary among species and are not shared with the other DEAD-box proteins; however they play important roles for Vas functions *in vivo*. Vas orthologs from many species carry an invariant tryptophan in the penultimate or ultimate position, Trp660 in *Drosophila melanogaster*, whose functional significance has not been studied. By modifying the endogenous allele of *vas* we found that Trp660 has a critical role in Vas function and its substitution with glutamic acid abolishes germ cell formation and embryonic patterning. In addition, this single amino acid substitution significantly reduces Vas activity in transposon silencing and Grk translation without decreasing the number of embryos produced by females. On the contrary, a conservative substitution to phenylalanine has only minor effects on germ cell formation without impairing anterior-posterior or dorsal-ventral patterning. Our study suggests that the conserved Trp in the C-terminal end of Vas plays a critical role in protein structure or in the interactions between Vas and other molecules, and that this function could be mostly supported by another hydrophobic aromatic amino acid.

Introduction

In many different species germ cells are set aside from somatic cells in very early stages of embryogenesis. Animals are divided into two groups according to their mode of germ cell specification which could be from maternally inherited determinants or as a result of inductive signals; nonetheless, in all of these species germ cells share many qualities that distinguish them from the soma. Work in a variety of animals reveals that the DEAD-box helicase Vasa (Vas) has a conserved role in specification of primordial germ cells (Castrillon et al., 2000; Ikenishi and Tanaka, 2000; Knaut et al., 2000; Kuznicki et al., 2000; Tanaka et al., 2000; Voronina et al., 2008), and is also involved in the function of somatic multipotent stem cells (Alie et al., 2011; Shibata et al., 1999; Yajima and Wessel, 2015). *vas* has been most extensively studied in *Drosophila*, where it is required for the assembly of pole plasm, a specialized cytoplasmic region at the posterior pole of oocyte containing all the germ cell determinants (Lasko and Ashburner, 1990). In the posterior pole, Vas also mediates abdominal segmentation through establishing a *nanos* (*nos*) gradient (Hay et al., 1988a; Lasko and Ashburner, 1990; Wang et al., 1994). Thus many alleles of *vas* cause females to produce embryos which lack germ cells and are not viable due to the defects in abdominal formation (Liang et al., 1994). In addition, in *Drosophila* ovaries Vas regulates translation of Grk, an Egfr ligand involved in dorsoventral patterning of the oocyte. As a result, embryos produced by females homozygous for vas^{PH165}, a null allele, have severe defects in dorsal appendage formation (Tomancak et al., 1998). vas^{PH165} females also produce a small number of embryos, which is mostly associated with the role that *vas* plays in progression of germline mitotic divisions and in differentiation of germline stem cells (Liu et al., 2009; Pek and Kai, 2011a). Furthermore, germ cells in many animals contain a unique structure in the perinuclear cytoplasm, termed the nuage, which is a compartmentalized site for piRNA processing to safeguard genome integrity against deleterious effects of transposable elements (Pek et al., 2012b). It has been shown that *vas* plays an essential role both for the nuage assembly and for the piRNA biogenesis (Liang et al., 1994; Xiol et al., 2014; Zhang et al., 2012). Two recent studies present evidence that *vas* is involved in the transfer of piRNA precursors across the nuclear envelope, and that together with some other components of piRNA pathway, such as Ago3 and Aub, form an amplifier complex participating in the ping-pong cycle (Xiol et al., 2014; Zhang et al., 2012).

The DEAD-box family of RNA helicases, which Vas belongs to, has crucial roles in a wide range of cellular processes including ribosome biogenesis, translation initiation, pre-mRNA splicing and mRNA decay (Jankowsky, 2011). Vas has been shown to interact with eIF5B, a translation initiation factor required for ribosome subunit joining (Carrera et al., 2000). This role of Vas, as a translation activator, is also required for the expression of *mei-P26*, whose product is involved in stem cell differentiation in *Drosophila* ovaries (Liu et al., 2009). Through coupling ATP binding and hydrolysis with RNA unwinding, DEAD-box proteins catalyze melting of the secondary structures in the RNA, and promote strand separation (Cordin et al., 2006; Jarmoskaite and Russell, 2011; Luking et al., 1998). In the helicase core these proteins are composed of two tandemly repeated RecA-like domains which share a set of highly conserved motifs. Motif II, in the N-terminal RecA-like domain, contains the D-E-A-D sequence which gives the name to this group. Aside from their conserved helicase domain DEAD-box proteins also have flanking N- and C-terminal sequences which vary significantly in their length and content. Many crystal analyses have excluded these, often disordered, terminal sequences (Carmel and Matthews, 2004; Caruthers et al., 2000; Kim et al., 1998; Korolev et al., 1997; Sengoku et al., 2006; Subramanya et al., 1996; Zhao et al., 2004); therefore, little is known about their contribution to the protein structure. A fragment composed of residues 200-623 was used to determine the crystal structure of Vas (Sengoku et al., 2006). This truncated form of Vas, which is essentially composed of the helicase core, exhibits RNA unwinding *in vitro*; however, *in vivo* analyses indicate that the conserved core domain is not sufficient to restore all Vas functions in a *va*s-null or -hypomorph background (Dehghani and Lasko, 2015).

The C-terminal region of Vas contains some conserved motifs including sequence 636-646 which is shared between closely related *Drosophila* species and is required for fine-tuning Vas localization to the pole plasm (Dehghani and Lasko, 2015). In addition, the most C-terminal amino acids from even distant Vas orthologs are highly acidic. In *Drosophila* deletion of the last seven amino acids, including five acidic residues, results in severe defects in different functions of Vas (Dehghani and Lasko, 2015). In the second last position of this sequence there is a highly conserved tryptophan, which is sometimes found as the last amino acid in other Vas orthologs. This conserved Trp is absent from the sequences of most other DEAD-box proteins in *Drosophila* and beyond. Thus it seems unlikely that this residue contributes to the general helicase activity of Vas orthologs; rather it might be involved in specific interaction between Vas

and the other proteins or mRNA targets. To gain insight into those Vas functions that are dependent on the conserved Trp660, we mutated this residue in the endogenous or transgenic alleles of *vas* and studied the effects *in vivo*. Results of our analyses show that a non-conservative substitution of Trp660 abolishes germ cell formation, posterior patterning, transposon silencing and dorsal-ventral axis formation but it does not affect female fecundity. Interestingly, although Trp has remained invariable through evolution we found that a transgenic allele encoding eGFP-Vas with phenylalanine, as a conservative substitution to Trp660, could effectively restore Vas functions in vas^{PH165} ovaries or vas^{1} embryos. The severe effects of W660E, but not W660F, on the majority of Vas functions suggest that Trp660 plays a fundamental role in the protein structure. Yet it is also possible that the conserved Trp mediates critical interactions between Vas and other molecules, which are efficiently supported by Phe in this position.

Materials and Methods

Endogenous *vas* alleles

vas^{PH165} is a null allele in which the entire coding sequence of *vas* has been deleted (Styhler et al., 1998). vas^{1} is a hypomorphic allele, generated by ethyl methanesulfonate (EMS) mutagenesis (Schupbach and Wieschaus, 1986b). This allele does not carry any mutation in the coding sequence of *vas*; however, expression of Vas in the ovaries is undetectable after the germarium stage (Lasko and Ashburner, 1990; Liang et al., 1994).

vas^{W660E} was generated using CRISPR-cas9 mediated homologous recombination (HR). Three target sequences from the genomic region within 50 base pairs up- or down-stream of Trp660 codon were compared for their efficiency to induce small deletions. To do this, the oligonucleotides corresponding to these sequences were ligated into pCFD3 plasmid and injected into *nos::cas9* embryos (Port et al., 2014). In the second generation, flies were screened using a T7 endonuclease assay, and the genomic modifications were confirmed by sequencing. The most efficient guide RNA (*gRNA*), targeting a sequence encompassing Trp660 codon (AGCAATGGGATTGAAATGTA) was then used for *PhiC31* integrase-mediated transgenesis (Bischof et al., 2007). The resulting transgenic males were then crossed to *nos::cas9* females to

produce embryos expressing both the *gRNA* and Cas9 in their germline. These embryos were then used to inject a donor DNA encoding the W660E mutation. The donor DNA was composed of a synthetic ssDNA oligonucleotide (Integrated DNA Technologies inc.) corresponding to 80 base pairs on either side of Trp660 codon and substituting it with GAA (Glu). The PAM sequence was mutated from TGG to TGA in the donor DNA to prevent its degradation by Cas9 in the cells.

Transgenic *vas* alleles

egfp-vas^{+} and *egfp-vas*$^{\Delta 15\text{-}75}$ have been described previously (Dehghani and Lasko, 2015). *egfp-vas*^{W660E} and *egfp-vas*^{W660F} were generated from a *vas* cDNA construct according to a method described before.

in vivo analysis

All the functional assays including egg-laying and hatching tests as well as the statistical analysis of the data were conducted according to Dehghani & Lasko (2015). Immunostaining, *in situ* hybridization and reverse transcription quantitative PCR (RT-qPCR) tests, were also described previously.

Yeast two-hybrid assays

The Matchmaker Gold Yeast Two-Hybrid System (Clontech Inc.) was used to test direct interactions between *vas* and other candidate genes. The coding sequences of *vas*$^{+}$, *vas*$^{\Delta 655\text{-}661}$ and *vas*^{W660E} were cloned between NdeI and PstI restriction sites in pGBKT7 vector. The entire coding sequences of *gus*, *aret*, *dlic*, *fsn*, *piwi*, *bar*, *cap-D2*, *spn-E*, *aub*, *ago3*, *csul* and *vls* were amplified and inserted between NdeI and BamHI restriction sites in PGADT7 AD vector. For *osk* short isoform, *eIF5B* and *qin* fragments encoding amino acids 5-468, 491-1144 and 275-1857 were cloned, respectively. Media preparation and yeast handling were performed according to Clontech Yeast Protocol Handbook (Clontech Laboratories, Inc. 2009). The interaction assays were conducted according to the instructions for Matchmaker Gold Yeast Two-Hybrid System.

Results

Vas orthologs from different species contain an invariant tryptophan in their C-terminal region

A comparison between amino acid sequences from different Vas orthologs, as far apart as sea urchins and humans, reveals that a Trp residue (W660 in *Drosophila*) is highly conserved within the last two amino acids of these sequences (Fig. 3.1 A). This Trp is also surrounded by several acidic residues in the C-terminal end; for example, in *D. melanogaster* there are four Glu and one Asp preceding and following the conserved Trp, respectively.

Such a conserved Trp however, is rarely found among the other DEAD-box proteins outside of the Vas orthologs (Fig. 3.1 B). In *D. melanogaster*, for example, Belle (Bel), which has the highest sequence similarity to Vas is the only other DEAD-box protein that contains Trp in its C-terminus. Rm62 is another DEAD-box protein in *D. melanogaster* which contains a phenylalanine (Phe; F), another aromatic amino acid in the similar position of its sequence. However, this Phe is not conserved across species; for example DDX5 (P68), the human or mouse orthologs of Rm62, do not carry a Phe within the last five amino acids of their sequences.

Taking advantage of the CRISPR/Cas9 technology together with the homologous recombination, we generated an endogenous allele of *vas* which encodes for Glu substituting Trp660. We also generated a transgenic allele of vas^{W660E}, which encodes a protein N-terminally tagged with eGFP, to examine localization pattern of the mutant protein in the ovaries. Similarly, we constructed an *egfp-*vas^{W660F} allele which encodes Phe as a conservative substitution for Trp660.

Substitution of Trp660 with Glu abolishes germ cell formation and embryonic viability.

The successful substitution of TGG (W) with GAA (E) in vas^{W660E} allele, generated by CRISPR/Cas9, was verified through sequencing. Moreover, the expression of a full length protein was confirmed by western blots (Fig. 3.2 B).

Fig. 3.1 Vas orthologs share a conserved tryptophan (W) in their C-terminal end, which is not common between the other DEAD-box proteins. **A**. Sequence alignment of the last 15 amino acids in Vas orthologs from different species indicates an invariant Trp in the penultimate or ultimate position. **B**. The last 15 amino acids in the C-terminal end from different DEAD-box proteins in *Drosophila melanogaster* are compared. Bel is the only other DEAD-box protein containing two consecutive Trp residues in its C-terminus, which also exhibits the highest sequence homology to Vas.

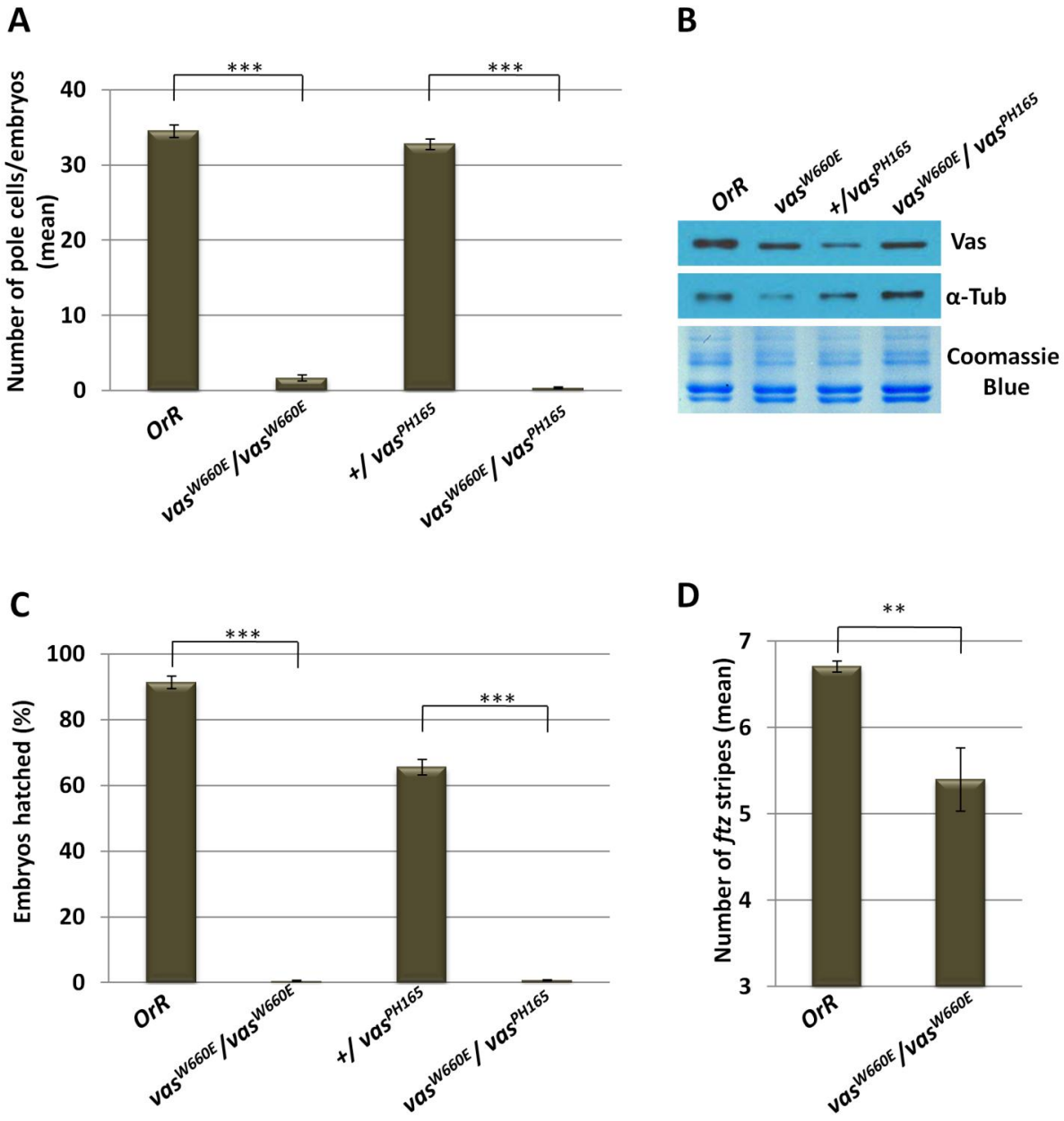


Fig. 3.2 Substitution of tryptophan (W) with glutamic acid (E), which is an abundant residue in the C-terminus of Vas, is associated with the failure in germ cell formation and abolished viability. **A**. the number of pole cells in stage 5-6 embryos is compared between *OrR* (*wt* control) and homozygous vas^{W660E}. A similar difference is also observed between $+/vas^{PH165}$ and vas^{W660E}/ vas^{PH165}. **B**. Western blot indicates a band corresponding to the full-length Vas for the four genotypes shown in A , α-Tubulin (α-Tub) and Coomassie Blue staining are shown as the loading controls. **C**. Embryos produced by vas^{W660E}/vas^{W660E} or vas^{W660E}/vas^{PH165} females have severely reduced hatching rates compared to *OrR* and $+/vas^{PH165}$, respectively. **D**. A significant decrease is also observed in the number of *ftz* stripes in vas^{W660E}/vas^{W660E} embryos compared to *wt*. Three and two asterisks represent P<0.0005 and P<0.005, respectively.

We investigated the significance of Trp660 for Vas function in germ cell development by counting germ cells in the embryos produced by *vas*^{W660E} homozygous females. Stage 5-6 embryos with two copies of wild type *vas* (*OrR*) contain an average of 35 pole cells (Fig. 3.2 A). We found that *vas*^{W660E} homozygous embryos (referred to by the maternal genotype) have an average of only 1.7 pole cells (n=83). To confirm that this is specifically associated with the *vas* mutation we also counted the number of germ cells in *vas*^{W660E}/*vas*^{PH165} embryos, and observed a similarly significant decrease compared to +/*vas*^{PH165} (P= 2E-53).

Vas is also involved in abdominal segmentation by mediating a Nanos (Nos) gradient to be established across the posterior-anterior axis. Therefore, most *vas* alleles produce embryos that do not hatch. We also tested the viability of *vas*^{W660E} homozygous and *vas*^{W660E}/*vas*^{PH165} heterozygous embryos (Fig. 3.2 C). Our analyses show that no more than 1% of the embryos from either of these two groups are viable; whereas +/+ or +/*vas*^{PH165} embryos have a hatching rate of 90% and 60%, respectively (n>1000). We confirmed this result by examining the expression pattern of *fushi tarazu (ftz)*, a paired rule gene, in *vas*^{W660E} embryos (Fig. 3.2 D). The average number of *ftz* stripes in stage 5-6 *vas*^{W660E} embryos was 5.4 which is significantly less than the average of 6.7 stripes in *wt* (P=0.001). According to these analyses, 43% of *vas*^{W660E} embryos, undergoing development, contained less than 6 *ftz* stripes which is consistent with the low viability of these mutants. However, we also found that the majority of *vas*^{W660E} or *vas*^{W660E}/*vas*^{PH165} embryos do not undergo any development suggesting that *vas*^{W660E} females have additional defects in oogenesis and many of their embryos are not fertilized.

Vas localization is independent of Trp660.

Since early stages of oogenesis Vas localizes to the perinuclear region of the nurse cells, also known as the nuage. From stage 10 Vas starts to be transported to the oocyte and accumulates at the pole plasm. In this study we used eGFP-tagged Vas fusions to examine localization pattern of Vas^{W660E} in the ovaries. Confocal images illustrate that eGFP-Vas^{W660E} localizes to the nuage and to the pole plasm similar to eGFP-Vas^{+}. Also when examined in a *wt* background this localization persists through oogenesis (Fig. 3.3 A).

To further examine if eGFP-Vas^{W660E} could maintain its localization in the absence of *wt* Vas, we inspected *vas*^{1} oocytes expressing *egfp-vas*^{W660E} and found that the GFP signal was

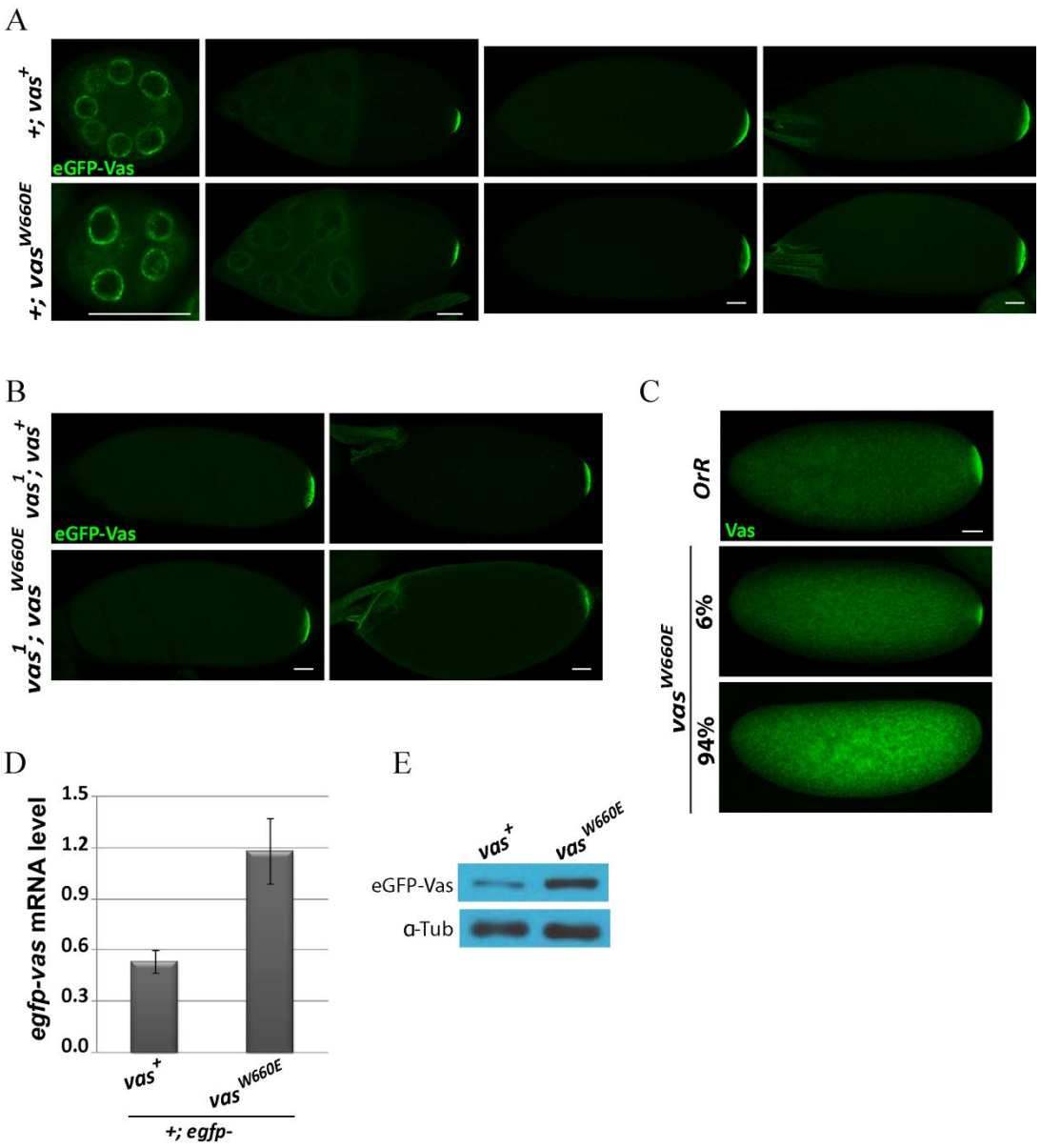


Fig. 3.3 eGFP-Vas^{W660E} mimics the localization pattern of eGFP-Vas^{+} in ovaries. Vas^{W660E} remains stable through oogenesis in a *wt* background, but stage 14 oocytes or early embryos from females who do not carry a *wt* copy of *vas* exhibit reduced levels of Vas^{W660E}. **A**. eGFP-Vas^{W660E} localization in a *wt* background is comparable to eGFP-Vas^{+}, **B**. However, a comparison between stage 13 and stage 14 oocytes from *vas^{1}; egfp-vas^{W660E}* females indicates that the pole plasm localization of eGFP-Vas decreases over time; whereas, this level is relatively stable for eGFP-vas^{+}. **C**. Similarly immunostaining of early embryos with anti-Vas shows that *vas^{W660E}/vas^{W660E}* embryos contain a reduced amount of Vas in their posterior region compared to the *wt* embryos, scale bars indicate 50 μm. **D**, **E**. A Comparison between transcript and protein levels of *egfp-vas^{+}* and *egfp-vas^{W660E}* (RT-qPCR and western blot) also indicates that Vas^{W660E} protein level is relatively stable in a *vas^{1}*/+ background. α-Tubulin were used as the loading control for western blot; *actin-5C* and *rp-49* were used as the reference genes in RT-qPCR.

reduced in stage 14 oocytes compared to the stage 13 (Fig. 3.3 B). This was on the contrary to *egfp-vas*$^{+}$ which maintains its localization throughout oogenesis. Similarly a vast majority (94%) of stage 1-2 embryos from *vas*^{W660E} contained no detectable Vas in their posterior region while the others had a considerably decreased levels compared to *wt* (Fig. 3.3 C). This data suggests that the functional defects of Vas^{W660E} ultimately compromise its localization during the late stages of oogenesis.

We also verified if W660E affects protein stability by quantifying levels of *egfp-vas*$^{+}$ and *egfp-vas*^{W660E} transcripts, using RT-qPCR, and comparing that with the protein levels on the western blot (Fig. 3.3 D,E). These experiments show that the level of *egfp-vas*^{W660E} transcript is about two times higher than *egfp-vas*$^{+}$ in the corresponding transgenic lines which were tested. A similar difference is also observed between eGFP-Vas levels through western blots indicating that the mutation does not have major effects on protein stability. As we have also noticed inconsistent decreases in the amount of Vas in ovary extracts from *vas*^{W660E}/+ compared to +/+, we cannot exclude the possibility that this mutation has subtle effects on protein stability.

Mutation of Trp660 to Glu reduces Vas function in piRNA pathway

Vas is required for the nuage assembly, and its critical role in piRNA biogenesis has been confirmed both in *Drosophila* and in mammals. Females homozygous for a *vas* null allele, *vas*^{PH165}, over-express several transposable elements including Het-A (Zhang et al., 2012). Elevated levels of HeT-A also have been reported for most *vas* alleles with mutations in their conserved helicase domains (Dehghani and Lasko, 2015). To investigate whether substitution of Trp660 with Glu affects Vas function in piRNA pathway, and therefore transposon silencing, we compared levels of HeT-A in the ovaries of *wt* and *vas*^{W660E} (Fig. 3.4 A). These analyses indicate that HeT-A expression in the ovaries of *vas*^{W660E} females is more than ten times higher than in *wt* (P=0.006). The HeT-A levels in *vas*^{W660E} fall within the same range as those observed for females which carry mutations in the conserved helicase motifs of Vas, or carrying a deletion of the seven C-terminal amino acids (Dehghani and Lasko, 2015).

We confirmed this result by measuring the levels of HeT-A in ovaries of *vas*^{PH165} females expressing either eGFP-Vas^{+} or eGFP-Vas^{W660E} (Fig. 3.4 B). In contrast to the *wt* construct,

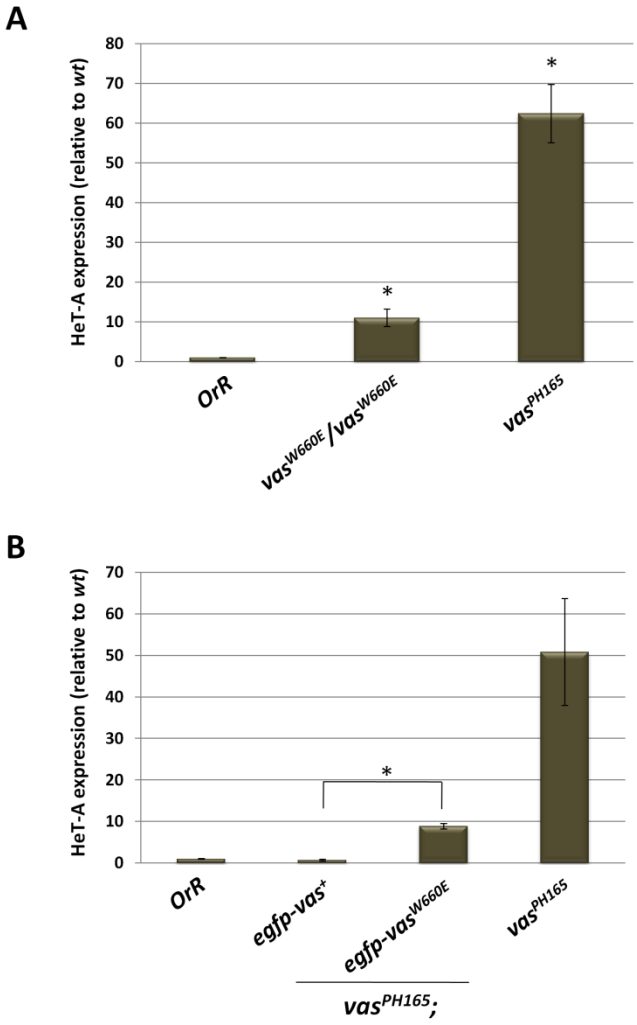


Fig. 3.4 Substitution of Trp660 with Glu impairs Vas function in transposable element silencing. **A**. The retrotransposon, HeT-A, is dramatically over-expressed in *vas*^{PH165} ovaries compared to *wt*. A consistent over-expression, albeit to a lesser extent than in *vas*^{PH165}, was also observed in *vas*^{W660E} ovaries. HeT-A levels in different genotypes are normalized to the *wt* level. Pre-*rp49* and *18s-rRNA* are used as the reference genes. **B**. an *egfp-vas*$^{+}$ construct is able to fully suppress HeT-A over-expression in *vas*^{PH165} ovaries. On the contrary, ovaries expressing *egfp-vas*^{W660E} still exhibit a significant over-expression of HeT-A. Each bar represents the average from at least three biological replicates. Error bars indicate the SEM.

which is able to fully suppress HeT-A over-expression, *vas*^{PH165} ovaries expressing *egfp-vas*^{W660E} exhibit increased levels of HeT-A at about ten times higher than *wt* (P=0.005).

vas^{W660E} females produce a similar number of embryos to wild type; these embryos however, have defects in dorsal appendages.

Vas has a critical role in oogenesis progression by regulating multiple processes including germline mitotic divisions, stem cell differentiation and dorsal ventral patterning of the oocytes (Liu et al., 2009; Pek and Kai, 2011a; Tomancak et al., 1998). In *vas*^{PH165} females all of these pathways are perturbed; these mutants produce very few eggs and their eggs do not form normal dorsal appendages. In addition, our previous analyses indicate that several mutations in the conserved motifs of Vas that are critical for its enzymatic activity do not abolish female fecundity, even though these residues are indispensable for other functions of Vas such as piRNA biogenesis or germ cell formation (Dehghani and Lasko, 2015).

To examine if W660E mutation, similar to its effect on germ cell formation, embryonic patterning and transposon silencing, abrogates female fecundity, we counted the number of eggs laid by individual females within the first 3 days from eclosion (Fig. 3.5 A). *vas*^{W660E} homozygous or hemizygous (over *vas*^{PH165}) produce embryos in numbers comparable to *OrR* females. However, we observed that more than 50% of the embryos produced by *vas*^{W660E} homozygous females do not produce two separate dorsal appendages (P=5.4E-09; Fig. 3.5 B). This phenotype was aggravated in hemizygous *vas*^{W660E} females in that only 15% of the eggs had two separate dorsal appendages versus 90% in +/ *vas*^{PH165} (P=2.4E-05). This indicates that Vas^{W660E} has a significantly decreased activity in establishing dorsal-ventral axis, which unlike Vas^{+}, results in a dosage sensitive situation. Immunostaining of ovaries shows that the defects in dorsal appendages of *vas*^{W660E} are correlated with decreased amounts of Grk in stage 8 oocytes (Fig. 3.5 C). Impaired dorsal-ventral patterning in *vas*^{W660E} embryos could also account for the large number of these embryos which remain unfertilized.

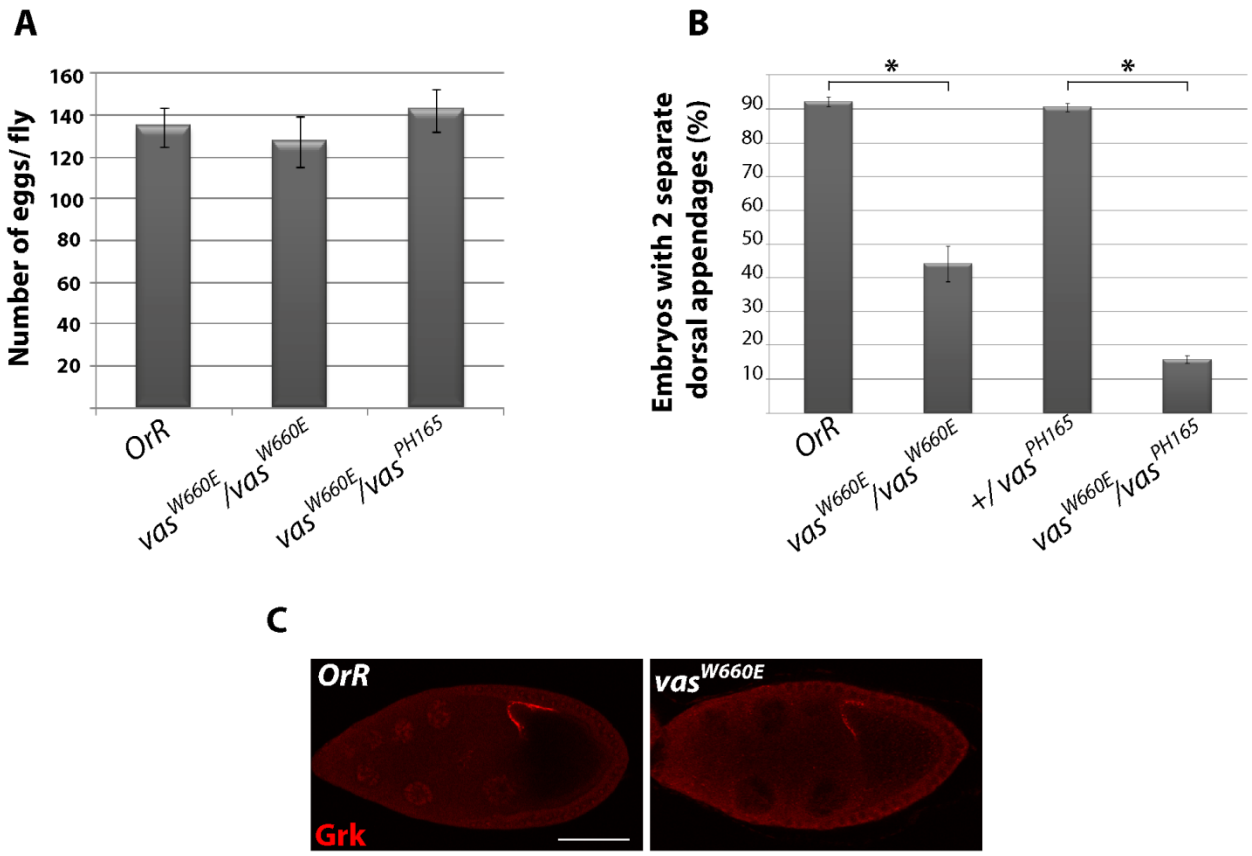


Fig. 3.5 *vas*^{W660E} females produce a wild type number of embryos. The W660E mutation, however, has a significant effect on the formation of dorsal appendages. **A**. females carrying two copies of wild type *vas* (*OrR*), two copies of *vas*^{W660E} or only one copy of *vas*^{W660E} (over *vas*^{PH165}) produce a comparable number of embryos. **B**. The percentage of embryos with two separate dorsal appendages is significantly reduced in *vas*^{W660E} compared to *wt*; this effect is even more pronounced when females are hemizygous for *vas*^{W660E} (over *vas*^{PH165}). n (the number of females tested)>20, error bars represent SEM, asterisks represent significant differences (P<0.05). **C**. representative images of Grk immunostaining in *OrR* and homozygous *vas*^{W660E} oocytes. Error bar indicates 50 µm.

Phenylalanine in position of Trp660 supports germ cell formation, abdominal segmentation and dorsal-ventral patterning.

We next investigated if Phe, a different hydrophobic amino acid with an aromatic ring, could replace Trp in position 660 of Vas. To address this question we used an *egfp-vas*^{W660F} transgene to rescue germ cell formation in a *vas*^{1} background.

We compared this construct with *egfp-vas*$^{\Delta 15\text{-}75}$, as a positive control, since both lines expressed eGFP-Vas at the similar levels (Fig. 3.6 A) and it has been shown previously that *egfp-vas*$^{\Delta 15\text{-}75}$ mimics full-length *vas* in various *in vivo* functional assays (Dehghani and Lasko, 2015). Our analyses indicate that the average number of pole cells in *vas*^{1} embryos expressing *egfp-vas*^{W660F} is 8.2 (Fig. 3.6 B; n=114) which is significantly different from *egfp-vas*^{W660E} embryos (n=185) with no germ cells (P=0.01E-27). In contrast, the number of germ cells present in *egfp-vas*^{W660F} is much closer to, if still significantly lower than, those in *egfp-vas*$^{\Delta 15\text{-}75}$ control embryos (n=323, p=0.01). We conclude that Phe can substitute for Trp660 with respect to the function of Vas in germ cell specification, and that the residual decrease in pole cell numbers in *egfp-vas*^{W660F} embryos can be attributed to the reduced expression level of this transgene.

We also investigated whether *egfp-vas*^{W660F} can support posterior patterning similar to *egfp-vas*$^{\Delta 15\text{-}75}$ control. Since homozygous *vas*^{1}*;egfp-vas*^{W660F} exhibited 90% lethality that we suspected to be non-specifically associated to the transgene insertion site, we decided to compare only one copy of *egfp-vas*^{W660F} with *egfp-vas*$^{\Delta 15\text{-}75}$ in a sensitized background of *vas*^{1}*/vas*^{W660E}. *ftz* staining of stage 5-6 embryos indicates that *egfp-vas*^{W660F} and *egfp-vas*$^{\Delta 15\text{-}75}$ both significantly increase the number of *ftz* stripes from 5.5 in *vas*^{1}*/vas*^{W660E} embryos to about 6.5 (P≤4.9E-12; Fig. 3.6 C).

To determine whether *egfp-vas*^{W660F} can support dorsal-ventral patterning we examined the dorsal appendages in *vas*^{PH165} embryos expressing different *egfp-vas* transgenes (Fig. 3.6 D). However, unlike *egfp-vas*^{W660E} embryos, which in 70% of the cases do not produce two separate or semi-fused dorsal appendages, 84% of *egfp-vas*^{W660F} embryos have two dorsal appendages comparable to 70% observed for *egfp-vas*$^{\Delta 15\text{-}75}$ (n; number of females tested >20). Consistent with this, *egfp-vas*^{W660F} could restore Grk expression in the ovaries of *vas*^{PH165} females similar to *egfp-vas*$^{\Delta 15\text{-}75}$ and more effectively than *efgp-vas*^{W660E}, indicating that Phe can mostly

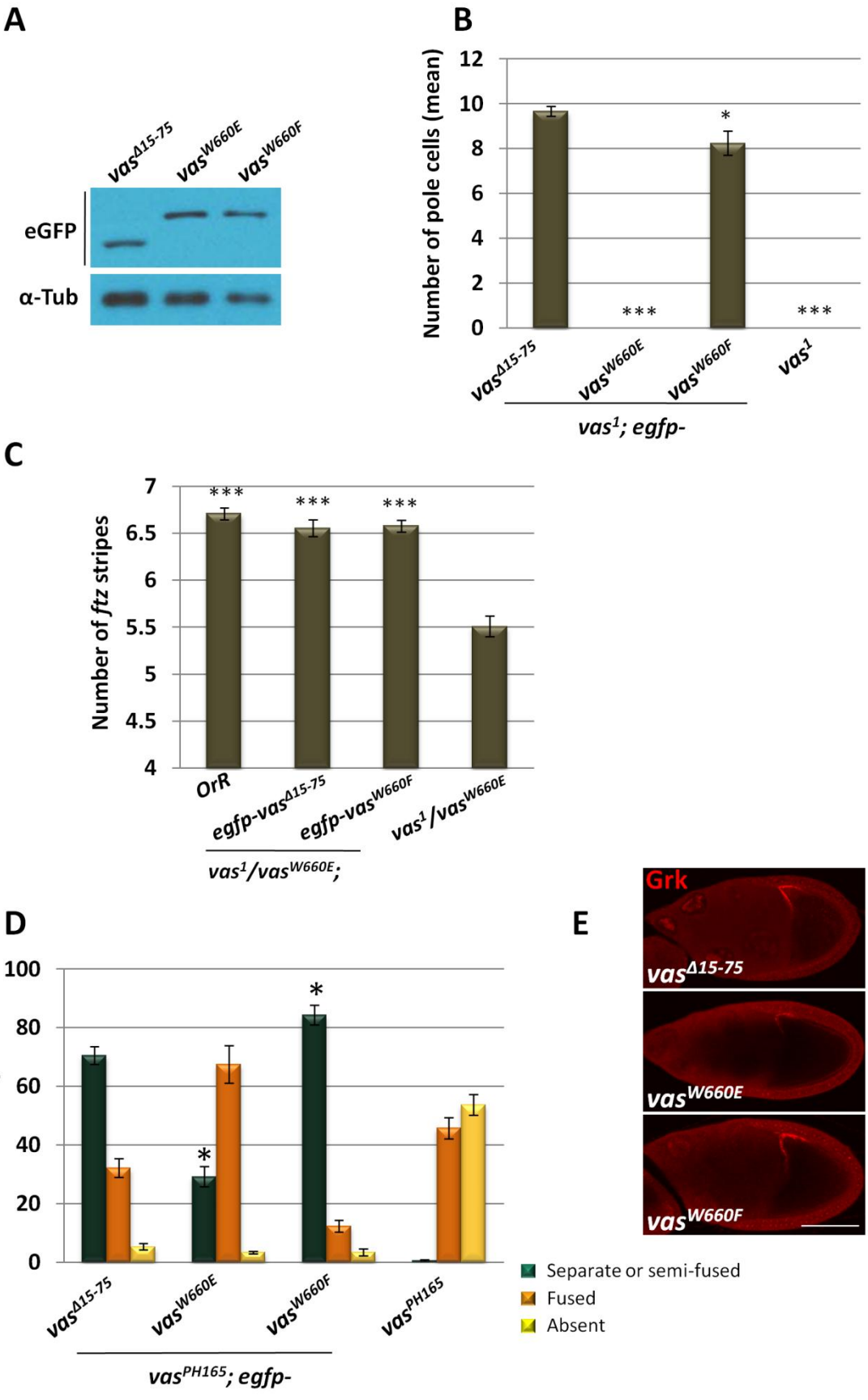

A
vasΔ15-75
vasW660E
vasW660F
eGFP
α-Tub
B
Number of pole cells (mean)
12
10
8
6
4
2
0
*
***
***
vasΔ15-75
vasW660E
vasW660F
vas1
vas1; egfp-
C
Number of ftz stripes
7
6.5
6
5.5
5
4.5
4
***
***
***
OrR
egfp-vasΔ15-75
egfp-vasW660F
vas1/vasW660E
vas1/vasW660E;
D
% of embryos
100
80
60
40
20
0
*
*
vasΔ15-75
vasW660E
vasW660F
vasPH165
vasPH165; egfp-
Separate or semi-fused
Fused
Absent
E
Grk
vasΔ15-75
vasW660E
vasW660F

Fig. 3.6 Conservative substitution of Trp with Phe does not affect the *vas* functions in germ cell formation, embryonic patterning or Grk translation. **A**. Western blot shows the expression levels of different eGFP-Vas proteins, detected by a GFP antibody. *egfp-vas*$^{\Delta 15\text{-}75}$ is considered as a positive control. α-Tub is used as the loading control. **B**. *vas*^{1} embryos expressing *egfp-vas*$^{\Delta 15\text{-}75}$ contain an average of 9.9 pole cells/embryo (n=323). Substitution of Trp with Glu, non-conservatively, abolishes germ cell formation (n=185). However, an *egfp-vas* encoding Phe instead of Trp is able to induce formation of 8.2 pole cells per embryo (n=114). One or three asterisks indicate $P<0.05$ or $P<0.0005$, respectively. **C**. Expression of *egfp-vas*$^{\Delta 15\text{-}75}$ or *egfp-vas*^{W660F} in *vas*^{1}/*vas*^{W660E} embryos both result in similarly significant increases in the number of *ftz* stripes. **D**. The percentage of embryos with two separate or semi-fused dorsal appendages increases from 0 to 70% by expression of *egfp-vas*$^{\Delta 15\text{-}75}$ in the *vas*^{PH165} background. Similarly, *vas*^{PH165} females expressing *egfp-vas*^{W660F} produce 84% embryos with two dorsal appendages. On the contrary, the population of such embryos from *vas*^{PH165}*; egfp-vas*^{W660E} females comprises only 29% of the total, n (number of females tested)>20. **E**. Representative images of Grk immunostaining in stage 7-8 egg chambers of *vas*^{PH165} females with *egfp-vas*$^{\Delta 15\text{-}75}$, *egfp-vas*^{W660E} or *egfp-vas*^{W660F}. Error bar indicates 50 μm.

substitute Trp660 in this regard (Fig. 3.6 E). Taken together, these results reveal that *vas* functions in germ cell formation, dorsal-ventral patterning and abdominal segmentation could tolerate a conservative substitution of Trp660 to Phe.

The direct binding of Vas to its known interacting partners is independent of Trp660.

Previous studies, using different *in vitro* and *in vivo* assays, have identified several proteins that are associated with Vas in the same complexes (Table. 3.S1). To investigate possible correlations between these protein interactions and the functional defects observed for vas^{W660E} and $vas^{\Delta 655\text{-}661}$ we first tried to co-immunoprecipitate Vas-protein complexes from ovaries or embryos. With this method, however, the known interacting partners of Vas, such as Aubergine (Aub), Oskar (Osk) and Valois (Vls), were not detected in the IP eluates of either eGFP-Vas or the endogenous Vas. We reasoned this might be due to weak or transient nature of the interactions or their low abundance in the whole ovary/embryo lysates. Thus as an alternative approach we used a yeast two-hybrid (Y2H) assay, by which Vas^{W660E} and $Vas^{\Delta 655\text{-}661}$ were compared to Vas^{+} for their interactions with sixteen other proteins. Our experiments confirmed direct binding of Vas to Osk, Gus and eIF5B, which have been previously reported (Breitwieser et al., 1996; Carrera et al., 2000; Styhler et al., 2002). These interactions, nevertheless, remained unchanged for vas^{W660E} and $vas^{\Delta 655\text{-}661}$ (Fig. 3.S1). For the other thirteen genes, which were mostly tested as full length proteins, we did not find a direct interaction with *vas* through yeast two-hybrid assays.

Discussion

A recent structure function analysis of *vas* revealed that a motif composed of the last seven amino acids (655-661) in Vas sequence is critical for its activity in germ cell formation and posterior patterning (Dehghani and Lasko, 2015). In addition, deletion of this motif dramatically reduces Vas functions in piRNA pathway and Grk translation. Motif 655-661 exhibits two conserved features among *vas* orthologs: it contains several acidic residues, ranging from three in humans to five in *D. melanogaster*, and it also harbors a highly conserved Trp in its penultimate position. In the current study we found that a non-conservative substitution of Trp to Glu, similar

to the deletion of motif 655-661, severely compromises most functions of Vas including germ cell specification, oocyte axis determination and transposon silencing. This mutation alone, however, did not decrease the number of embryos produced by females.

Trp660 has a fundamental role in Vas activity: a non-conservative substitution of this residue abolishes germ cell formation and embryonic viability.

In this study, for the first time, we edited the endogenous allele of *vas*, and this allowed us to examine the effect of a point mutation by direct comparisons with the *wt* allele. Our analyses indicate that vas^{W660E} phenotypes mimic some of the vas^{1} defects in that females produce a *wt* number of embryos; however, these embryos do not form germ cells and are not viable. On the contrary to vas^{1} females though, we observed that vas^{W660E} females have increased levels of HeT-A retrotransposon in their ovaries indicating that vas^{W660E} fails to fully support piRNA biogenesis. Similarly, vas^{W660E} oocytes display defects in their dorsal appendages suggesting that this mutant form of Vas does not activate Grk translation at sufficient levels.

We also found that W660E is slightly different from Δ655-661 in that it has no effect on the number of embryos produced by females. Vas role in oogenesis progression is exerted during several stages which include regulation of the germline mitotic divisions, stem cell differentiation, piRNA biogenesis and dorsoventral axis formation in the oocyte (Liu et al., 2009; Pek and Kai, 2011a; Tomancak et al., 1998; Zhang et al., 2012). Our previous study showed that many *vas* alleles that contain mutations in their conserved helicase domains still produce a number of eggs, which is significantly higher than *vas* null (Dehghani and Lasko, 2015). This suggests that for some of the functions during early oogenesis, perhaps for germline mitotic progression, *vas* acts as an inert scaffold, which does not require an enzymatic activity. Such scaffolding function of Vas must be also preserved in vas^{W660E} mutants. In addition, and unlike *vas* core mutants, which although higher than *vas* null still produce significantly fewer eggs than *wt*, the number of embryos produced by vas^{W660E} females remains comparable to *wt*. This might be due to the fact that Trp660 is not a conserved feature of DEAD-box proteins and presumably not critical for the enzymatic activity of Vas. Therefore, Vas^{W660E} still confers more activity than an enzymatically dead protein merely acting as a scaffold, which supports a *wt* number of oocytes completing oogenesis.

Non-conservative substitutions, such as Trp to Glu, which might be unfavorable for the protein conformation, could result in fast degradation of the mutant protein. Our analyses exclude the possibility that W660E mutation has a dramatic effect on protein stability, although we observed a slight decrease in the level of Vas protein when comparing ovarian extracts from heterozygous *+/vas*^{W660E} with +/+ (data not shown). The mild effect of W660E on protein stability, however, cannot explain the severe phenotypes observed for *vas*^{W660E} mutants as heterozygous females for *vas* null, which express nearly half the *wt* levels of Vas in their ovaries, still produce embryos with a similar number of germ cells to *wt*.

A *vas* transgene encoding Phe instead of Trp660 could restore Vas functions in the mutant background.

Our study using an *egfp-vas*^{W660F} indicates that many Vas functions in oogenesis and germ cell formation could tolerate a conservative substitution of Trp to Phe. Hydrophobic residues such as Trp are often buried in the protein core, where they contact other non-polar amino acids (Betts and Russell, 2003; Santiveri and Angeles Jimenez, 2010). In addition, Trp could interact with the other amino acids containing an aromatic side chain inside the protein or on the surface of other proteins (Betts and Russell, 2003). Phe and Tyrosine (Tyr) are the two other amino acids, containing an aromatic ring, which resemble Trp in many of its biochemical properties. However, presence of a nitrogen atom in the indole side chain of Trp confers some unique properties to this residue; for example, Trp occasionally mediates interactions with non-protein molecules such as lipid components of the cell membrane (de Jesus and Allen, 2013; Dougherty, 2007).

Since Trp660 has remained invariant through evolution we predicted that some of the functions shared between *vas* orthologs will be impaired by even a conservative substitution to Phe. Surprisingly, such mutation did not result in a dramatic decrease in the number of germ cells. Similarly the ability of *egfp-vas*^{W660F} to restore Grk translation in *vas*^{PH165} ovaries, or embryonic patterning in *vas* hypomorph embryos was comparable to the *wt* construct. Yet, it is possible that W660F results in small but biologically significant changes in Vas functions, which remain to be examined through other approaches. An endogenous allele of *vas*^{W660F} would allow accurate quantitative comparisons with the wild type *vas* allele.

The nature of interactions between Trp660 and other parts of Vas and its possible role in protein conformation remains to be investigated through crystal structure analyses that would include the C-terminal region. We observed, however, that the non-conservative substitution to Glu, despite its severe effects on most Vas functions, had mild or no effect on female fecundity, suggesting that the mutant protein still retains a conformation that supports some function. Yet it is possible that a subtle change in Vas folding affects certain functions more than the others.

The yeast two-hybrid assays confirmed direct binding of Vas to a small number of other proteins, which are however not affected by either Δ655-661 or W660E. On the other hand the wide spectrum of Vas functions suggests that Vas is engaged in many more protein complexes than what is currently known and leaves it possible that W660E compromises some of these interactions.

It also remains unknown whether the acidic residues in the C-terminal region of Vas have a functional significance, although mutation of Trp660 alone resulted in the defects that were almost as severe as those in *vas*$^{\Delta 655\text{-}661}$ (Dehghani and Lasko, 2015). Among the other DEAD-box proteins the C-terminal region of Bel, which exhibits the highest sequence homology to Vas, contains two consecutive Trp, but only one acidic residue within the last ten amino acids. Therefore, one possible experiment is to replace the last seven amino acids of Vas with those from Bel and study the resulting chimeric protein *in vivo*.

Acknowledgments

We are grateful to Beili Hu for microinjection into the embryos. All the images were taken using the Cellular Imaging and Analysis Network (CIAN) facility at McGill University. We would also like to thank Fillip Port for his technical support during CRISPR gene editing and for sharing some of his unpublished experiences. This work was supported by NSERC and CIHR grants to P. L.

Supplementary figures

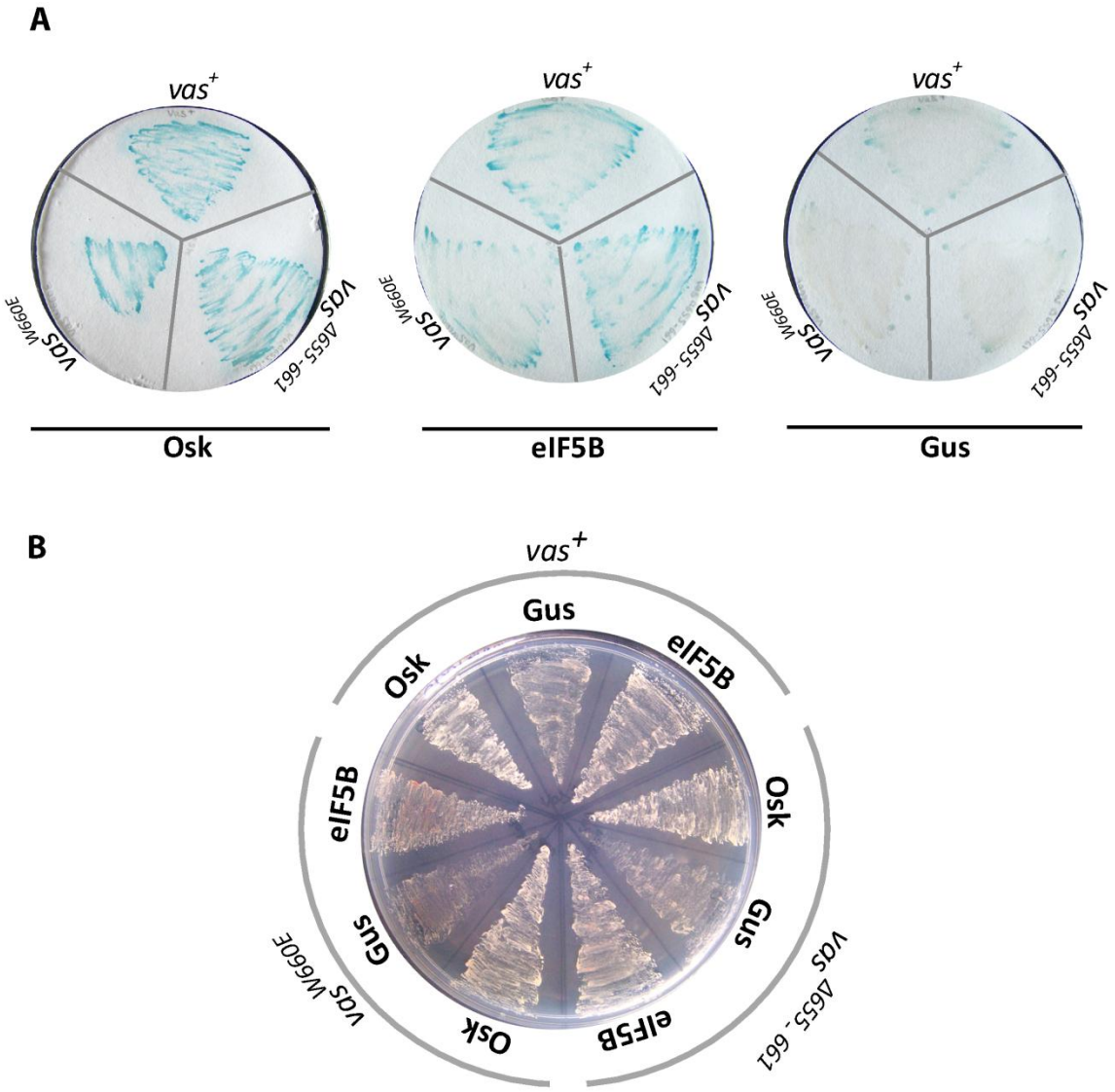


Fig. 3.S1 Interactions of Vas with Osk, Gus and eIF5B are not abolished by Δ655-661 or W660E. **A**. A β-Galactosidase filter assay shows similar results for vas^{+}, $vas^{\Delta 655\text{-}661}$ and vas^{W660E} in terms of interactions with Osk, Gus and eIF5B. **B**. Positive interactions would enable colonies to grow on the quadruple drop-out (-Trp, -Leu, -His and –Ade) media containing 200 ng/ml Aureobasidin A. This test further confirmed that the interactions with Osk, Gus and eIF5B are not notably different among vas^{+}, $vas^{\Delta 655\text{-}661}$ and vas^{W660E}.

Table 3.S1 The list of the proteins that have been previously found to associate with Vas through different assays, including yeast two-hybrid (Y2H), GST pull-down or co-immunoprecipitation (co-IP). The latter has been performed using the endogenous or tagged proteins from *Drosophila* ovaries or *Bombyx mori* BmN4 cell lines. We expressed each one of these candidates in yeast, either as a full length protein or a fragment. Our Y2H assays only showed a direct interaction between full-length Vas and Osk, Gus and eIF5B, consistent with the previous studies. 1) Breitwieser et al., 1996, 2) Anne, 2010, 3) Styhler et al., 2002, 4) Kugler et al., 2010, 5) Carrera et al., 2000, 6) Webster et al., 1997, 7) Lerit and Gavis, 2011, 8) Patil and Kai, 2010, 9) Megosh et al., 2006, 10) Pek and Kai, 2011a, 11) Anand and Kai, 2012, 12) Xiol et al., 2014

Interacting protein	Previous evidence	Cellular process	The fragment expressed in yeast	Direct interaction
Oskar	Y2H [1], Pull-down [2]	Embryonic patterning, germ cell formation	Short isoform (aa 5-468)	Yes
Gustavous	Y2H, pull-down [3], tagged co-IP [4]	Regulation of Vas stability	Full length	Yes
eIF5B	Y2H, Pull-down [5]	Translational control	aa 491-1144	Yes
Arrest/Bruno	Pull-down [6]	Oogenesis, embryonic patterning	Full length	No
Dynein Intermediate chain (Dlic)	Endogenous co-IP [7]	Transport of germ plasm	Full length	No
F-box synaptic protein (FSN)	Tagged co-IP [4]	Regulation of Vas stability	Full length	No
Tejas	Tagged co-IP [8]	piRNA processing	Full length	No
Piwi	Tagged co-IP [9]	piRNA processing	Full length	No
Barren	Endogenous co-IP [10]	Chromosome condensation in GSCs	Full length	No
CAP-D2	Endogenous co-IP [10]	Chromosome condensation in GSCs	Full length	No
Spindle-E	Endogenous co-IP [10]	piRNA processing	Full length	No
Qin	Tagged and endogenous co-IP [11, 12]	piRNA processing	aa 275-1857	No
Aubergine	endogenous co-IP [12]	piRNA processing	Full length	No
Argonaute 3	endogenous co-IP [12]	piRNA processing	Full length	No
Capsuleen	endogenous co-IP [12]	Arginine methyltransferase	Full length	No
Valois	endogenous co-IP [12]	Co –factor of Capsuleen	Full length	No

Appendix B

Introduction

The germline specific RNA helicase, Vas, is essential for several steps of oogenesis and for germ cell specification. This suggests Vas involvement in numerous RNP complexes associated with different pathways, a subset of which have been explored so far through different genetic and biochemical approaches. However, to fully understand mechanisms of Vas functions much needs to be learned about other interacting partners of Vas, the nature of interactions with the known partners and the domains of Vas being involved.

To investigate the effect of different Vas mutations, which have been phenotypically studied before, on the interactions of Vas with other proteins I first applied a co-immunoprecipitation (co-IP) approach. This technique, however, could not efficiently capture Vas interactions with the known partners, most likely, due to the transient nature of these interactions. Therefore, I next applied yeast two-hybrid (Y2H) assays to verify direct interactions between Vas and its previously known partners. Among sixteen proteins tested only Osk, Gus and eIF5B, were found to directly bind Vas through Y2H system, as previously reported. To identify additional Vas partners, whose binding could be examined by Y2H, I carried out an Y2H screen using a *Drosophila* cDNA library. This screen resulted in identification of several new putative interactions, which need to be further investigated by other assays and to be functionally characterized.

Materials and methods

Co-IP

Ovary and embryo extracts were prepared using a lysis buffer composed of 20mM HEPES-KOH pH 7.5, 150 mM KCl, 4 mM $MgCl_2$, 0.15% NP-40 and 1x HALT protease inhibitor (Thermo Scientific). This lysis buffer was used in all co-IP experiments shown, unless otherwise specified. For protein cross-linking, lysis buffer was supplemented with 3 mM DSP, and extracts were incubated at room temperature for 30 min with mixing. Cross-linking was stopped by adding 50 mM Tris-HCl pH 7.8 and 15 min incubation. Lysates were pre-cleared with Protein A beads (Roche) for 1 h. For immunoprecipitation, extracts were incubated either with pre-blocked

GFP-Trap (Chromotek) for 3 h or with anti-Vas (1:400; Liu et al., 2003), anti-GFP (Invitrogen, 1:250) or anti-eIF4E (1:50; Lachance et al., 2002) overnight at 4˚C with mixing. When antibody used, lysates were incubated with pre-blocked Protein A beads for 3 h. The beads were washed four times, 10 min each, with wash buffer (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 4 mM $MgCl_2$ and 1x HALT protease inhibitor). The beads were boiled in 1X SDS buffer (2% SDS, 62.5 mM Tris base, 10% glycerol, 35 mM DTT) for 10 min. Antibodies against Aub (1:10000; P. Lasko's lab), Osk (1:2000; Yoshida et al., 2004), α-Tub (1:15000, Sigma), PABP (1:5000; Nishihara et al., 2013), Cup (1:500; Nelson et al., 2004) or eIF4G (1:2000; Ghosh and Lasko, 2015) were used for western blots.

To prepare samples for mass-spectrometry analyses, proteins were loaded onto a 10% SDS polyacrylamide gel and run for only 2 cm. The entire area containing the proteins was excised as one sample and sent for LC-MS/MS analysis.

Yeast two-hybrid

Plasmids: To amplify the coding sequence of different *vas* partners, the following cDNA clones were used: *bru: LD29068, fsn: LD47425, qin: AT14886, barr: RE15383, CAP-D2: LD40412, ago3: LD15865, vls: AT09449, dlic: GH06357, piwi: RE21038*. The open reading frames of *csul* and *tej* were amplified from the total cDNA generated from ovaries. *aub* and *spn-E* were sub-cloned into pGADT7-AD from the plasmids generated by Patil and Kai (2010). Similarly the plasmids described by Carrera et al. (2000) and Kugler et al. (2010) were used for cloning *osk, eIF5B* and *gus*. Further details of construct preparation could be found in chapter 3.

Protein expression test: To verify expression of AD-fused proteins in yeast by western blots two protein extraction methods were applied. These two protocols use different lysis buffers: 1- 200 mM NaCl, 1mM EGTA, 5mM MgSO4, 50mM Tris-HCL, 5% glycerol, 2% TritonX-100 and 0.1% SDS (Vogel et al., 2001) and 2- 8M Urea, 40mM Tris-HCL, 0.1mM EDTA, 5% SDS (Printen and Sprague, 1994). A rat anti-HA (Cat No. 11867423001, Roche) was used to probe fusion proteins on the western blots.

Interaction assays: Protein interactions were tested according to the Matchmaker Gold Yeast Two-Hybrid System (Clontech Inc.). The mating culture was incubated at 30˚C, 250 RPM, for 24 h, and then spread on double drop-out (DDO) SD medium lacking Trp and Leu, and

containing X-α-Gal (DDO/X). Positive interactions were further tested through an X-β-gal filter lift assay (Breeden and Nasmyth, 1985) as well as by the growth on medium lacking Trp, Leu, His and Ade (quadruple drop-out; QDO) but containing 200 ng/ml Aureobasidin A (QDO/A).

Library screen: A Mate & Plate Universal *Drosophila* Library, which is normalized to reduce the amount of highly abundant transcripts (Cat No. 630485, Clontech Inc.), was used to screen for the interacting partners of Vas. The library screens were carried out according to the instructions for Matchmaker Gold Yeast Two-hybrid System (Clontech Inc.). The mating cultures were examined under the microscope for the presence of 3-lobed zygotes after 22-24 h. DDO/X medium containing Aureobasidin A (DDO/X/A) or triple drop-out (TDO) medium, lacking Trp, Leu and His but containing X-α-Gal (TDO/X) was used for spreading the overnight culture in the first and the second screens, respectively. To identify positively interacting genes, plasmids were extracted from yeast by inoculating yeast colonies in 30 µl NaOH (20 mM) and incubating at 95˚C for 45 min. After centrifugation the supernatant was directly used in a PCR reaction with Matchmaker AD LD-Insert Screening Amplifier primers. The PCR product was sequenced using T7 primer.

Results

The co-immunoprecipitation approach to identify Vas-interacting partners is impeded by weak or transient nature of the interactions.

To explore molecular mechanisms underlying the defects observed in various *vas* mutants I used a co-IP approach. With this method, proteins that co-IP with wild-type and mutant forms of Vas could be compared through western blot or mass-spectrometry analyses; the former is limited by the availability of the antibodies while the latter identifies novel proteins in an unbiased manner.

To purify eGFP-Vas from the ovarian extracts I tested a polyclonal anti-Vas and reagents against the GFP tag: a polyclonal anti-GFP antibody and GFP-Trap beads. For simplicity immunoprecipitation is used as a general term, which refers to affinity purification where GFP-trap beads are used. In all the following experiments *egfp-vas* refers to *egfp-vas*$^{\Delta 15\text{-}75}$, which was

used due to its higher expression level compared to *egfp-vas*$^{+}$ (Dehghani and Lasko, 2015). As shown in Fig. B.1, the GFP-Trap beads were considerably more efficient in immunoprecipitating Vas (enrichment >15%, Fig. B.1 C) as compared with the two antibodies (Fig. B.1 A, B). The GFP-Trap beads consist of a GFP binding protein covalently coupled to the surface of agarose beads. This is also advantageous for the western blot and mass spectrometry analyses as the immunoprecipitate is devoid of the IgG chains; thus all subsequent IPs were carried out using the GFP-Trap beads.

To determine if the immunoprecipitate contains protein interactions that exist in known Vas-containing complexes, I stained the blot for the presence of Aub and Osk, two proteins that have been previously identified to be associated with Vas (Xiol et al. 2014, Breitwieser et al. 1996). However, these proteins were not detected in the immunoprecipitate from the ovarian lysates (Fig. B.2 A). As a control for the co-IP conditions I used the same procedure to precipitate some of the proteins associated with the eukaryotic initiation factor 4E (eIF4E; Nelson et al., 2004; Richter and Sonenberg, 2005; Tarun and Sachs, 1996; Zapata et al., 1994). This method could efficiently precipitate eIF4G and Cup as proteins which interact with eIF4E in an RNA-independent manner, as well as PABP whose interaction with eIF4E is through RNA (Fig. B.2 B).

One possible reason for the failure to co-IP proteins in a complex could be due to weak or transient interaction among the partners. To overcome this, protein interactions can be stabilized through chemical cross-linking methods (Mattson et al., 1993). I chose dithiobis [succinimidyl propionate] (DSP) which is a membrane-permeable homobifunctional N-hydroxysuccimide ester (NHS-ester), commonly used to cross-link intracellular or intramembrane proteins (Baskin and Yang, 1982; Hordern et al., 1979; Joshi and Burrows, 1990; Lomant and Fairbanks, 1976). DSP reacts with primary amines, found at the N-terminus of each polypeptide chain and in lysine residues, to form thiol-cleavable bonds. For this procedure tissues were homogenized in the presence of DSP in the lysis buffer, and IP was performed as before. As a control for cross-linking efficiency I compared migration of eIF4G, eIF4E and Vas through a 10% SDS polyacrylamide gel among non-cross-linked, cross-linked, and cross-link reversed samples (Fig. B.2 C). In this control experiment eIF4G band was absent in the lane corresponding to the cross-

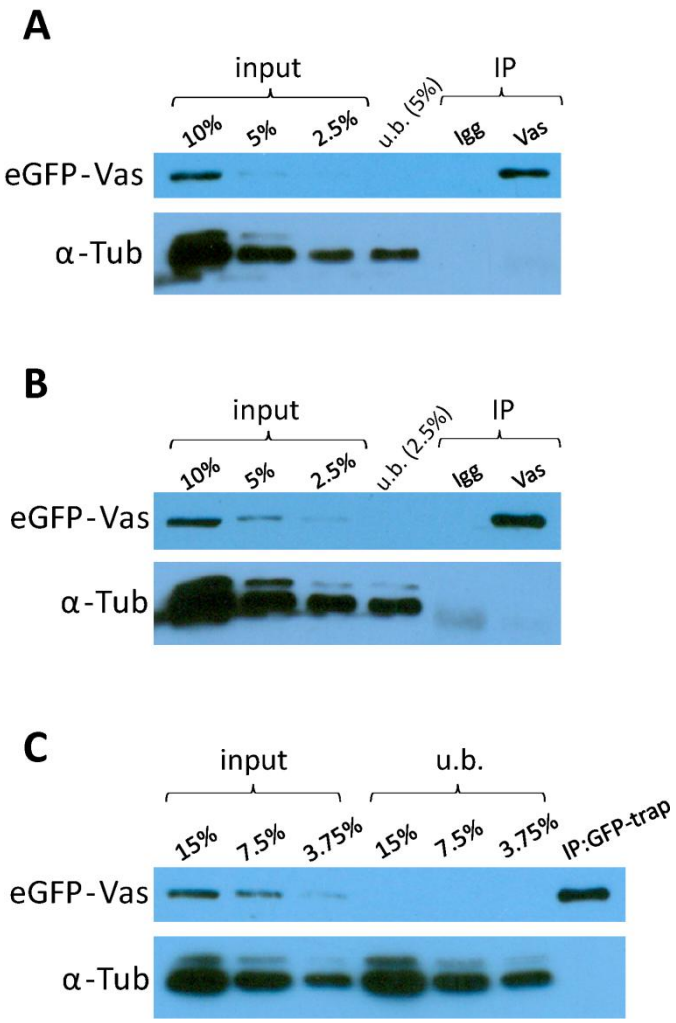


Fig. B.1 A comparison between a Vas antibody (**A**), a GFP antibody (**B**) and GFP-Trap beads (**C**) for their efficiency to immunoprecipitate eGFP-Vas from embryos. All three methods immunoprecipitated eGFP-Vas specifically compared to α-Tub. The IP efficiency was higher using GFP-Trap beads or anti-GFP (notably higher than 15% or 10% recovery, respectively) compared to anti-Vas (~10% recovery). The input fractions are relative to the amount loaded in the IP lanes. u.b.: unbound fraction

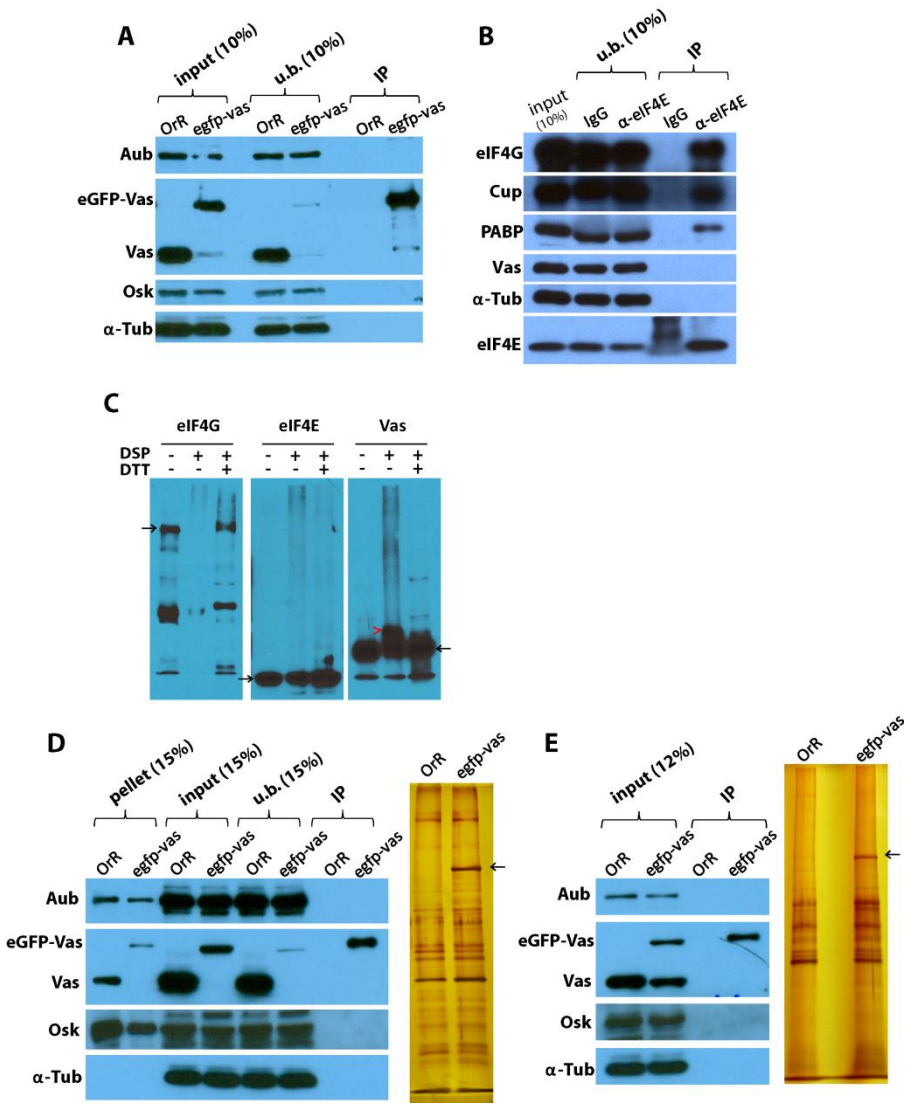


Fig. B.2 A co-IP approach to isolate Vas associated protein complexes from *Drosophila* ovaries and embryos **A**. IP using ovaries from *vas*^{1}*; egfp-vas* (*egfp-vas*) or *OrR* control females, western blots are stained with anti-Vas to control for the IP efficiency. Osk and Aub are examined as the two known partners of Vas. α-Tub serves as the negative control for the co-IP, u.b.: unbound fraction, **B**. Co-IP conditions were tested with an IP from *wt* ovary extracts using eIF4E antibody. A control IP was performed using normal rabbit serum (IgG). eIF4G, Cup and PABP are detected in the eIF4E co-IP. Vas and α-Tub are included as the negative controls. **C**. eIF4G, eIF4E and Vas complexes were tested between samples, prepared in the presence (+) or absence (-) of DSP. 35 mM DTT was used to cleave the disulfide bonds resulted from cross-linking. Arrows indicate the band corresponding to the size of the specific protein tested in each blot. The red arrowhead shows a Vas complex slightly larger than free Vas protein. **D**. co-IP using the same tissues as in B, and DSP as the cross-linking reagent: to ensure that the proteins are efficiently extracted by the extraction buffer, the tissue pellets were also tested by western blots. Except Osk, the other proteins tested were efficiently solubilized. Silver staining compares the same co-IP eluates tested by western blots. Arrow indicates eGFP-Vas. **E**. Similar co-IP as in D but using *vas*^{1}*/+; egfp-vas* or *OrR* embryos.

linked sample, suggesting that cross-linking has stabilized eIF4G in large complexes, which cannot enter the gel matrix. Also Vas antibody, in addition to a band corresponding to Vas, detected a prominent band about 10 kDa higher, and a smear of higher molecular weight complexes specifically in the cross-linked sample. However, the eluates from *vas*^{1}*; egfp-vas* ovarian or *vas*^{1}/+*; egfp-vas* embryo extracts, did not show the presence of Aub or Osk even after cross-linking (Fig. B.2 D, E).

It was still possible that the amount of these proteins in the immunoprecipitate is below the detection limit of the western blots. Therefore, to confirm this and to identify yet uncharacterized proteins that interact specifically with Vas in the pole plasm, I performed IP using 0-2 h embryo extracts from *vas*^{1}/+*; egfp-vas* and *OrR* (*wt* control) under conditions as before and using DSP. The immunoprecipitates were then subjected to LC-MS/MS analysis (Table B.1). This experiment also did not detect any known interacting partners of Vas in the co-IP eluate from *egfp-vas* embryos. However, the small number of peptide reads indicated that the samples did not contain sufficient amounts of protein, and thus these analyses remained inconclusive.

A previous study demonstrated that substitution of a single amino acid (E to Q) in the conserved DEAD box motif of *Bombyx mori* Vas (*Bm*Vas) reduces the dynamics of Vas *in vivo*, thereby stabilizing complexes containing it (Xiol et al., 2014). When extracts from a *B. mori* ovarian cell line, transfected with epitope-tagged *Bm*Vas or *Bm*Vas^{DQAD} were used for co-IP, the piRNA pathway components Ago3, Siwi (*Bombyx* ortholog of Aub) and Qin (a Tudor domain protein) immunoprecipitated with Vas^{DQAD} but not with the wild-type Vas. This suggests that Vas in its ATP-bound closed conformation provides a platform for stable assembly of a piRNA amplifier complex. To test if a similar approach could stabilize Vas-containing complexes in the ovaries I used a DEAD-box mutant allele, *egfp-vas*^{D399L}, which I had generated previously and which should have the same characteristics as *vas*^{DQAD}. With this approach and also following the same co-IP protocol as Xiol et al. (2014), only a faint band corresponding to Vls, another protein previously shown to interact with Vas, was detected on the western blots. However, Aub and Osk remained undetectable (Fig. B.3 A). Although silver staining indicates many bands in *egfp-vas*^{D399L} lane, which are absent in *wt*, western blots using anti-GFP suggest that egfp-Vas^{D399L} is highly unstable and that many of these bands may correspond to the truncated forms

Table B.1 The top 27 proteins identified by LC/MS/MS analyses in the co-IP elutes from *OrR* (*wt* control) and *egfp-vas* embryos using GFP-Trap beads. For this list of proteins peptide threshold and protein threshold were set at 90% and 95% confidences, respectively. Values displayed for each sample are the total number of peptide counts for the corresponding proteins.

#	Identified Proteins	*OrR*	*egfp-vas*
1	ATP-dependent RNA helicase vasa	0	2
2	Heat shock protein 83	6	0
3	Vitellogenin-2 precursor	1	0
4	Vitellogenin-3 precursor	3	1
5	*REV* Protein 4.1 homolog	1	0
6	Vitellogenin-1 precursor	0	1
7	*REV* Kinesin-like protein Klp68D	2	0
8	CG8977-PB, isoform B	2	0
9	Netrin-B precursor	1	0
10	Transitional endoplasmic reticulum ATPase TER94	2	0
11	CG1721-PA, isoform A	1	0
12	Tropomyosin-1, isoforms 9A/A/B	1	0
13	CG10701-PE, isoform E	1	0
14	CG13389-PB, isoform B	1	0
15	40S ribosomal protein S18	1	0
16	IP15838p	1	0
17	CG2512-PB, isoform B	1	0
18	CG12101-PB, isoform B	1	0
19	26S proteasome regulatory complex subunit p48B	1	0
20	GH07384p	1	0
21	LD42771p	1	0
22	CG2852-PB, isoform B	1	0
23	CG6203-PD, isoform D	1	0
24	Zinc finger protein on ecdysone puffs	1	0
25	CG3213-PA	1	0
26	CG1821-PC, isoform C	1	0
27	CG3762-PB, isoform B	1	0

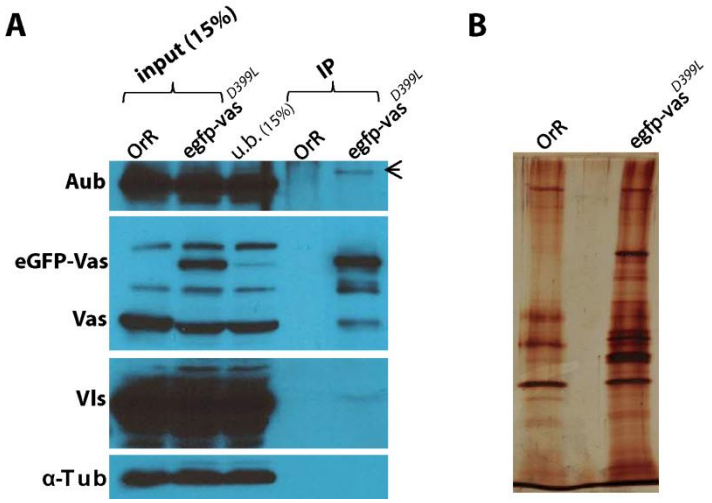
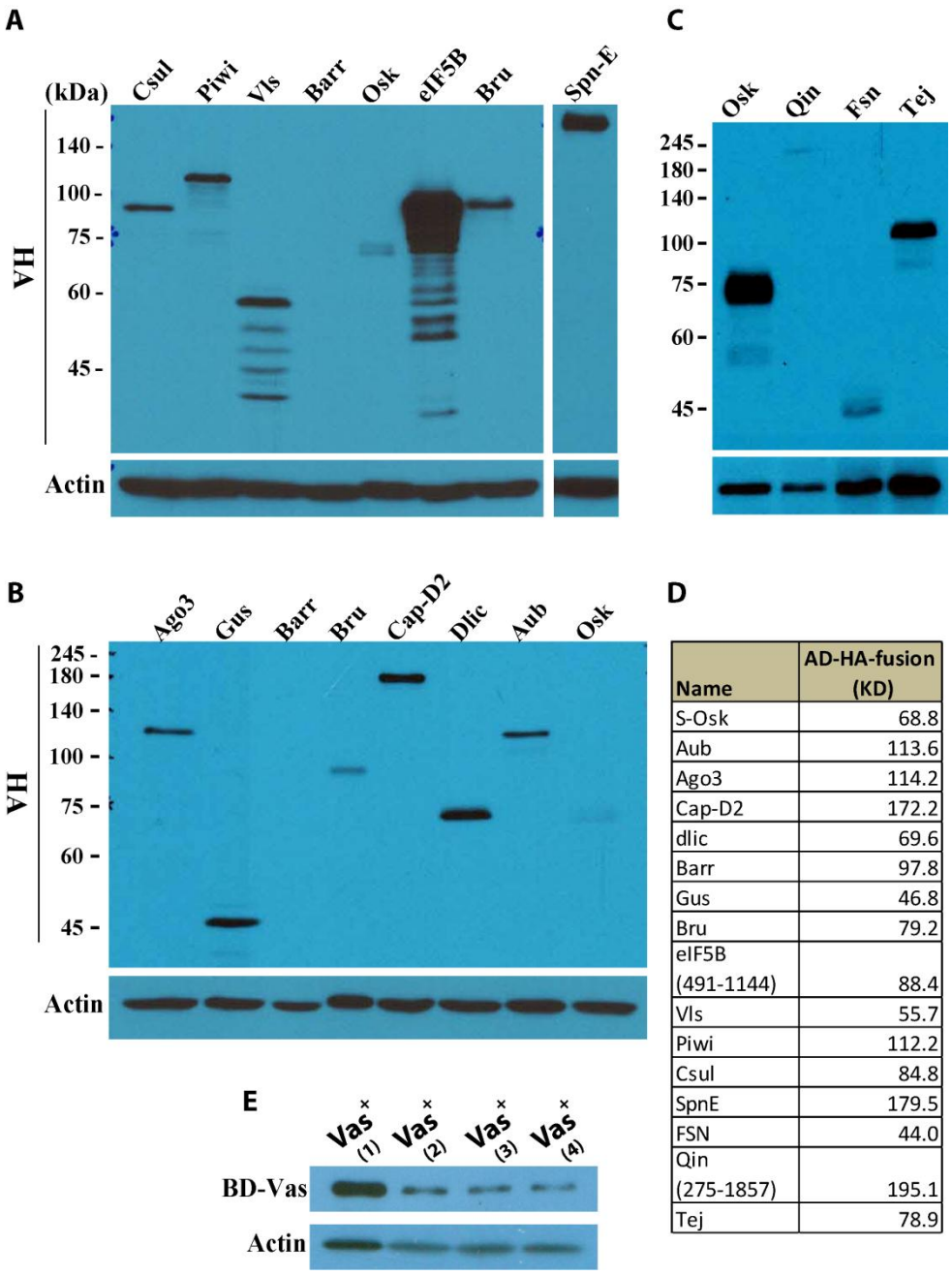


Fig. B.3 an *egfp-vas* allele with mutation in DEAD motif (D399L) was used to form stable interactions in the RNP complexes (Xiol et al. 2014). **A**. Western blot indicates a band corresponding to Vls in the IP eluate from *vas*^{1}*/+; egfp-vas*^{D399L} ovaries that is not present in *OrR* sample. The band detected by Aub antibody in the IP from *egfp-vas*^{D399L} (arrowhead) is non-specific and corresponds to the size of eGFP-Vas^{D399L} bait. Also both western blot and silver staining (**B**) indicate that a significant portion of eGFP-Vas^{D399L} in the IP eluate is degraded.

of protein (Fig B.3 A,B). It remains possible that D399, in addition to its known ATP-binding role (Sengoku et al., 2006), is critical for Vas folding and stability, and thus DQAD, which does not have this additional function serves as a better substitution for stabilizing Vas complexes.

Y2H confirmed previously known interactions of Vas with Osk, Gus and eIF5B, but did not indicate direct interactions with the other 13 partners

Since co-IP experiments did not effectively capture interactions between Vas and its known protein partners, and also due to the limited number of the reagents available, I chose yeast two-hybrid (Y2H) as an alternative. This method is widely used to examine direct protein-protein interactions, which are likely to be affected by mutations or deletions in the proteins. A previous study used Vas as the bait in an Y2H screen against an ovarian cDNA library (Carrera et al., 2000). The putative interactions identified in that screen mostly fell into five classes, each being composed of different fragments from: *osk, eIF5B, gus, DNA polymerase interacting tpr containing protein of 47kD* (*Dpit47*) and a *Drosophila* virus gene (Carrera et al. 2000; Styhler et al. 2002; Paul Lasko, unpublished information). Moreover, the direct interaction between Vas and Osk, which is an essential step in the formation of pole plasm, was found for the first time through Y2H assays (Breitwieser et al., 1996). Several other proteins have been found associated with Vas through different *in vivo* or *in vitro* assays; however for the majority of these proteins a direct interaction has not been demonstrated (Table 3.S1). To investigate if different mutations in Vas sequence, such as Δ655-661 or W660E, affect Vas binding to its partners I first verified the direct interaction between vas^{+} and all the other genes that *vas* is known to associate with. The coding regions from 16 genes, including *osk*, *eIF5B* and *gus*, were cloned into a pGADT7 vector expressing the prey protein fused at the N-terminus to the activation domain (AD) of Gal4 and an HA tag. The accuracy of the inserts was verified by sequencing, and expression of the HA-tagged proteins in yeast was tested through western blots (Fig. B.4 A, B, C, D). Two yeast protein extraction methods were applied with noticeable differences in their capacity to solubilize various fusion proteins (Vogel et al., 2001; Printen and Sprague, 1994). Of the 16 proteins, AD-Barr was the only one whose expression was not detected following either of these protocols. It is possible that *Drosophila* Barr could not be expressed in yeast or alternatively this protein could not be solubilized using the standard protein extraction methods.



Name	AD-HA-fusion (KD)
S-Osk	68.8
Aub	113.6
Ago3	114.2
Cap-D2	172.2
dlic	69.6
Barr	97.8
Gus	46.8
Bru	79.2
eIF5B (491-1144)	88.4
Vls	55.7
Piwi	112.2
Csul	84.8
SpnE	179.5
FSN	44.0
Qin (275-1857)	195.1
Tej	78.9

Fig. B.4 Candidate genes from *Drosophila* were N-terminally fused to the activation domain (AD: 12.5 kDa) of Gal4 and expressed in yeast. Each fusion protein also contains an HA tag following the AD sequence. Protein extracts used in **A**, **C** were prepared according to Vogel et al. (2001), whereas for the western blot shown in **B** proteins were extracted according to Printen & Sprague (1994). The two buffers significantly differed in their capacity to solubilize AD-Osk. Actin is used as a loading control. **D**. Table indicates the expected sizes of different AD-HA-fusion proteins. **E**. Western blot shows expression of Vas fused to the DNA binding domain (BD) of Gal4. Protein expression was confirmed in four separate colonies.

Similarly, I expressed Vas N-terminally tagged with the binding domain (BD) of Gal4 in a Y2H-Gold strain and verified its expression by western blots (Fig. B.4 E). I also confirmed that neither *vas*$^{+}$ nor any of the interacting fusion proteins could individually activate reporter genes in the Y2H system to produce false-positive signals via X-α-Gal test or X-β-Gal filter lift assay.

Next I examined a direct interaction between *vas*$^{+}$ and each one of the 16 candidate genes (Fig. B.5). A robust interaction could only be observed for the previously reported genes: *osk*, *eIF5B* and *gus*. Some other genes, such as *Cap-D2* and *bru*, were positive in the X-α-Gal tests but not in others. A direct interaction has been previously shown between Bru and Vas through GST pull-down assays (Webster et al., 1997); it is however possible that certain interactions remain undetectable in Y2H assays (Van Criekinge and Beyaert, 1999). Also noteworthy is that the majority of these candidate genes were tested as full-length proteins. Large proteins sometimes are not properly folded in yeast or may also contain inhibitory domains which affect interactions (Auerbach and Stagljar, 2005; Van Criekinge and Beyaert, 1999). Therefore, fragments of such proteins might reveal genuine interactions stronger than the full-length.

A Y2H screen identifies new putative interacting partners of Vas

To investigate if there are other proteins whose interactions with Vas are detectable through Y2H assays I performed a screen using full-length Vas as the bait and a cDNA library prepared from a pool of mRNA isolated from *Drosophila* embryos, larvae, and adults. This cDNA library was expected to include genes that are specifically expressed in the ovaries. Furthermore a Gal4-based Y2H system, such as the one used in these experiments and a LexA system used by Carrera et al. (2000) have different capacities for detecting various types of interactions (Van Criekinge and Beyaert, 1999). Thus, I anticipated this screen would identify interactions which might have remained unknown in the previous study.

In the first round of screening I spread the mated culture on DDO/X/A medium. Only two potential interactions were captured by this screen, from two genes identified as *pasha* and *CG33978*. Since none of the known interacting partners of *vas* including *osk*, *gus* and *eIF5B* were found, I repeated the screen using a less stringent condition by spreading the overnight culture on TDO/X plates. *HIS3* reporter gene produces high levels of His as a result of interaction between bait and prey proteins; however, in many yeast strains this construct also

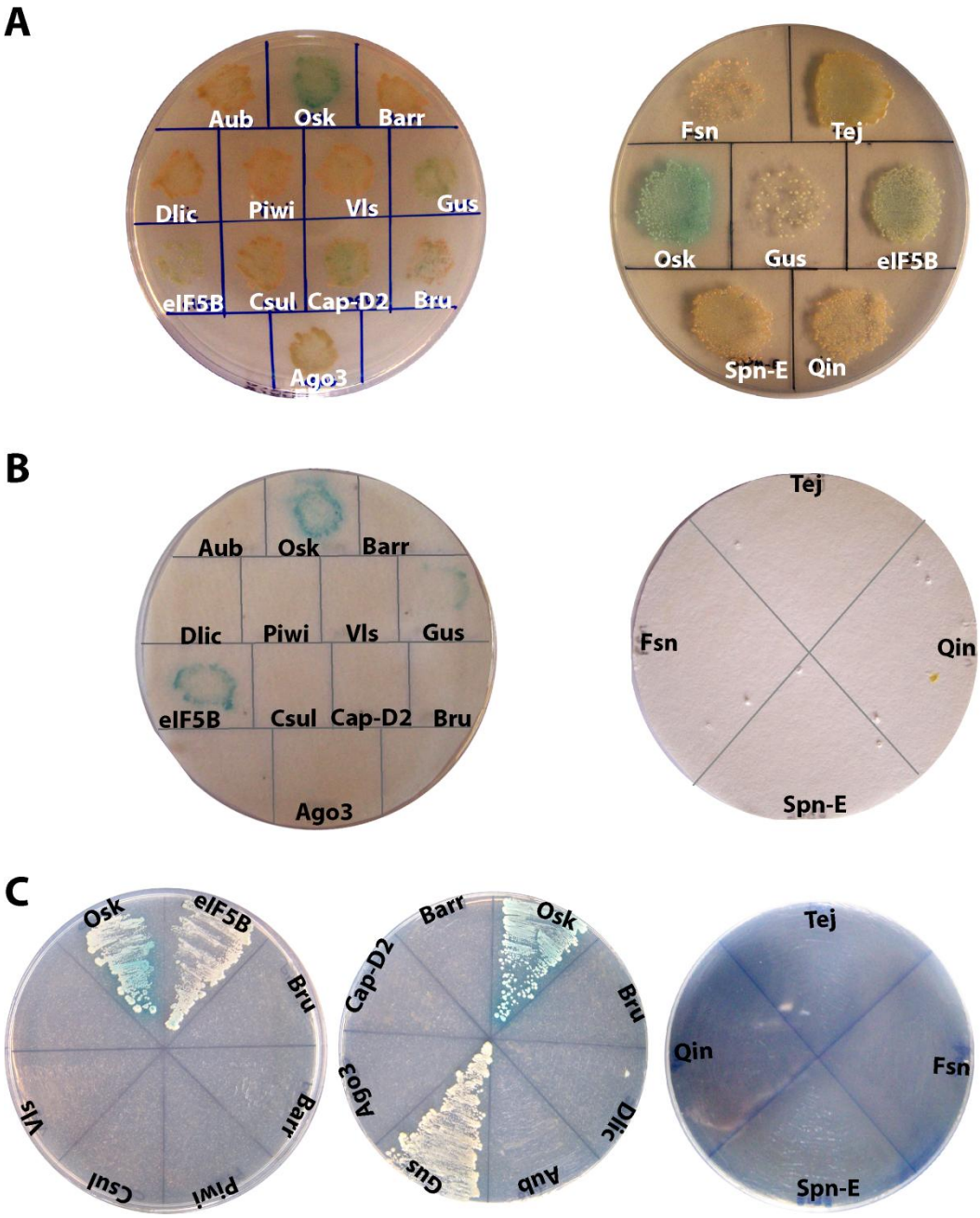


Fig. B.5 Direct interactions between candidate genes (Table 3.S1) and *vas*$^{+}$ were tested through a yeast two-hybrid assay. **A**. Alpha-galactosidase (*MEL1*) is expressed in yeast cells as a result of interaction between bait and prey fusion proteins; thus, the positive colonies turn blue in the presence of X-α-Gal in the medium. **B**. An X-Gal filter lift assay confirmed positive interactions of Vas^{+} with Osk, Gus and eIF5B. **C**. The yeast cells containing Osk, eIF5B or Gus together with Vas^{+} are the only ones able to grow on QDO/A plates.

expresses some background levels of His, which result in false positive signals. For this reason I further relied on the *MEL1* reporter gene and selected the blue colonies on the X-α-gal containing medium. By this method, together with the results from the first screen, I detected 70 potential interactions, which were subjected to additional confirmation tests (Fig. B.6).

Thirty five colonies among these were also found to be β-gal positive or/and to grow on the highly stringent QDO/A medium. Through colony PCR and sequencing I identified 19 different genes as candidates with potentially genuine interactions with *vas* (Table B.2). From this list nine genes have a high or moderate expression, whereas the remaining genes exhibit low or no expression in ovaries (Cherbas et al., 2011; Chintapalli et al., 2007; Graveley et al., 2011). As in the first screen, however, I did not find *osk*, *gus* or *eIF5B* among the interacting genes. Three different DNA samples isolated from cDNA library in yeast were tested through PCR for the presence of *vas, osk, gus* and *eIF5B* (Fig. B.7). Results of this experiment suggest that the library is mostly composed of full-length cDNA from short transcripts or 3' end cDNA fragments of 1kb or smaller. As colony PCR from positive interactions identified in our screen shows, the library also contains cDNA between 1.5-2 kb, which however might constitute a small subset of the colons corresponding to each gene, and thus are only detectable for highly expressed genes. This could account for our failure to detect interaction between Vas and eIF5B or Osk in the current screen, since the cDNA clones that encode necessary motifs for these interactions may not be sufficiently present in the library. The mating efficiency of the bait and the library strains was calculated by counting the number of haploid cells, only able to grow on single drop-out plates (-Trp or –Leu) and the diploid cells being able to also grow on DDO plates. With the mating efficiency of 1.8%, this screen was estimated to have tested $1.4x10^6$ diploids, which is above the minimum number recommended by the library supplier (Clontech Inc.). However, given the complexity of the library used in this screen, low abundant or weak interactions may still require a larger number of diploids to be tested by repeating the screen. Alternatively a tissue specific library, which is enriched for the potential partners, could be generated and used.

Furthermore I tested *mRpl44, Rpp20, CG2678, pasha, CG2091, CG1239, eas, Pka-C1* and *cip4*, the genes identified in the screen, which are also reported to be expressed in the ovaries, for their interaction with $vas^{\Delta 655\text{-}661}$. The pGADT7 plasmids containing these genes were isolated from diploid yeast cells, and following amplification in *E. coli* they were transformed back to the

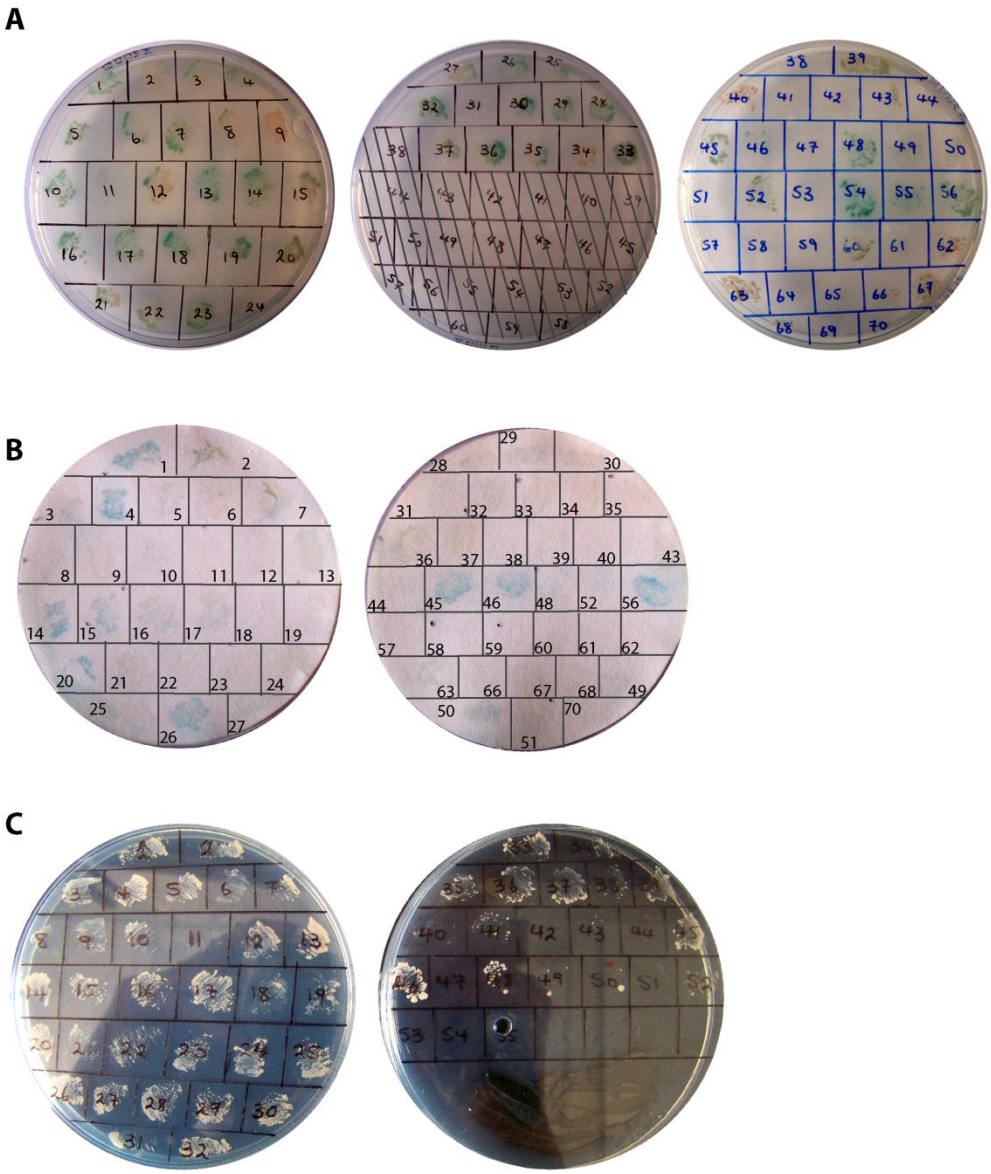


Fig. B.6 Three different assays were used to confirm positive interactions identified in the yeast two-hybrid screen using *vas*$^{+}$ as the bait and a *Drosophila* cDNA library as the prey. **A**. Expression of α-galactosidase is visualized on DDO/X plates for the majority of these clones consistent with their initial identification in the screen. Clones 14 and 15 were from the first round of screening, whereas the remaining clones were identified in the second screen. **B**. Majority of these clones were β-gal positive to various degrees. A number of clones are missing in this test: clones 47, 53, 64, 65 and 69 did not grow on DDO plates when re-streaked, possibly due to the toxicity of the library gene, and therefore were eliminated from further analyses. Clones 41, 42, 54 and 55 (later found similar to 20) although missing from β-gal filter-lift assay have been tested by the two other assays. **C**. Interactions were further tested on QDO plates. Colonies 57-70, which did not exhibit a strong interaction through X-α-gal or β-gal filter-lift assays, and were mostly not viable on DDO plates, were not further tested on QDO plates.

Table B.2 list of the genes identified in a yeast two-hybrid screen using the full-length *vas* as the bait and a *Drosophila* cDNA library (Clontech Inc.) as the prey, the clone numbers correspond to the numbers shown in Fig. B.4. A relatively high expression refers to a moderate expression level that is higher than in most other tissues tested. 1: FlyBase Curators et al., 2004-, 2: FlyBase Genome Annotators, 2012, 3: Koc et al., 2001, 4: Pho et al., 1996, 5: Palm et al., 2012, 6: Duncan et al., 1999, 7: Sisson et al., 2000, 8: Panáková et al., 2005, 9: Hua and Zhou, 2004, 10: Ayme-Southgate et al., 2004, 11: Clark et al., 1998, 12: Denli et al., 2004, 13: Landthaler et al., 2004, 14: Martin et al., 2009, 15: Smibert et al., 2011, 16: Azzam et al., 2012, 17: Lagueux et al., 2000, 18: Stroschein-Stevenson et al., 2005, 19: Pavlidis et al., 1994, 20: Engel and Wu, 1994, 21: Wojtowicz et al., 2004, 22: Watson et al., 2005, 23: Kalderon and Rubin, 1988, 24: Lane and Kalderon, 1995, 25: Nahm et al., 2010, 26: Chintapalli et al., 2007, 27: Graveley et al., 2011, 28: Cherbas et al., 2011

Clones#	Gene symbol	Molecular function	Expression in the ovaries
1,15,16, 30,32,33	*CG33978*	Predicted EGF-like calcium ion binding [1]	No expression [26]
5,21,36, 37,46	*CR44309*	Probable long-non-coding RNA gene; may encode small polypeptide(s) [2]	No expression [27]
24,27,34, 48	*mRpL44*	Structural constituent of mitochondrial large ribosome subunit [3]	Relatively high [26]
13, 17,29	*Rfabg*	Lipid transporter activity [4, 5] Heme binding [6] Microtubule binding [7] Required for morphogens wingless (wg) and hedgehog (hh) diffusion [8]	Low [26]
4,56	*Rpp20*	Predicted ribonuclease P activity [9]; tRNA and rRNA processing [1]	Very high [27]
10,19	*aqrs*	Predicted serine-type endopeptidase activity, proteolysis [1]	No expression [26]
20,55	*bt*	Structural constituent of muscle [10], protein kinase activity [1]	Low [26]
7	*CG2678*	DNA binding transcription factor activity [11]	Relatively high [26]
14	*pasha*	Primary miRNA processing [12,13, 14, 15] (implicated in processes such as germarium-derived female germ-line cyst formation [16])	Relatively high [26]
2	*CG42808*	No experimental evidence; no prediction	No expression [28]
22	*CG31886*	No experimental evidence; no prediction	Low [26]
26	*CG2091*	Predicted hydrolase activity; deadenylation-dependent decapping of nuclear-transcribed mRNA[1]	Relatively high [26]
28	*Tep2*	Predicted endopeptidase inhibitor activity [1](implicated in anti-bacterial defense and phagositosis [17, 18]).	No expression [26]
35	*CG1239*	Predicted methyltransferase activity, with bicoid interacting 3 features[1]	Relatively high [26]
39	*CG14696*	No experimental evidence; no prediction	No expression [26]
54	*eas*	Ethanolamine kinase activity [19], choline kinase activity [20]	Moderate [26]
23	*Dscam1*	Axon guidance receptor activity [21], antigen binding [22]	No expression [26]
25	*Pka-C1*	cAMP-dependent protein kinase 1 [23] (implicated in several processes including oogenesis [24])	Moderate [26]
50	*cip4*	GTPase activating protein binding (implicated in several processes including cellular protein localization [25])	Relatively high [26]

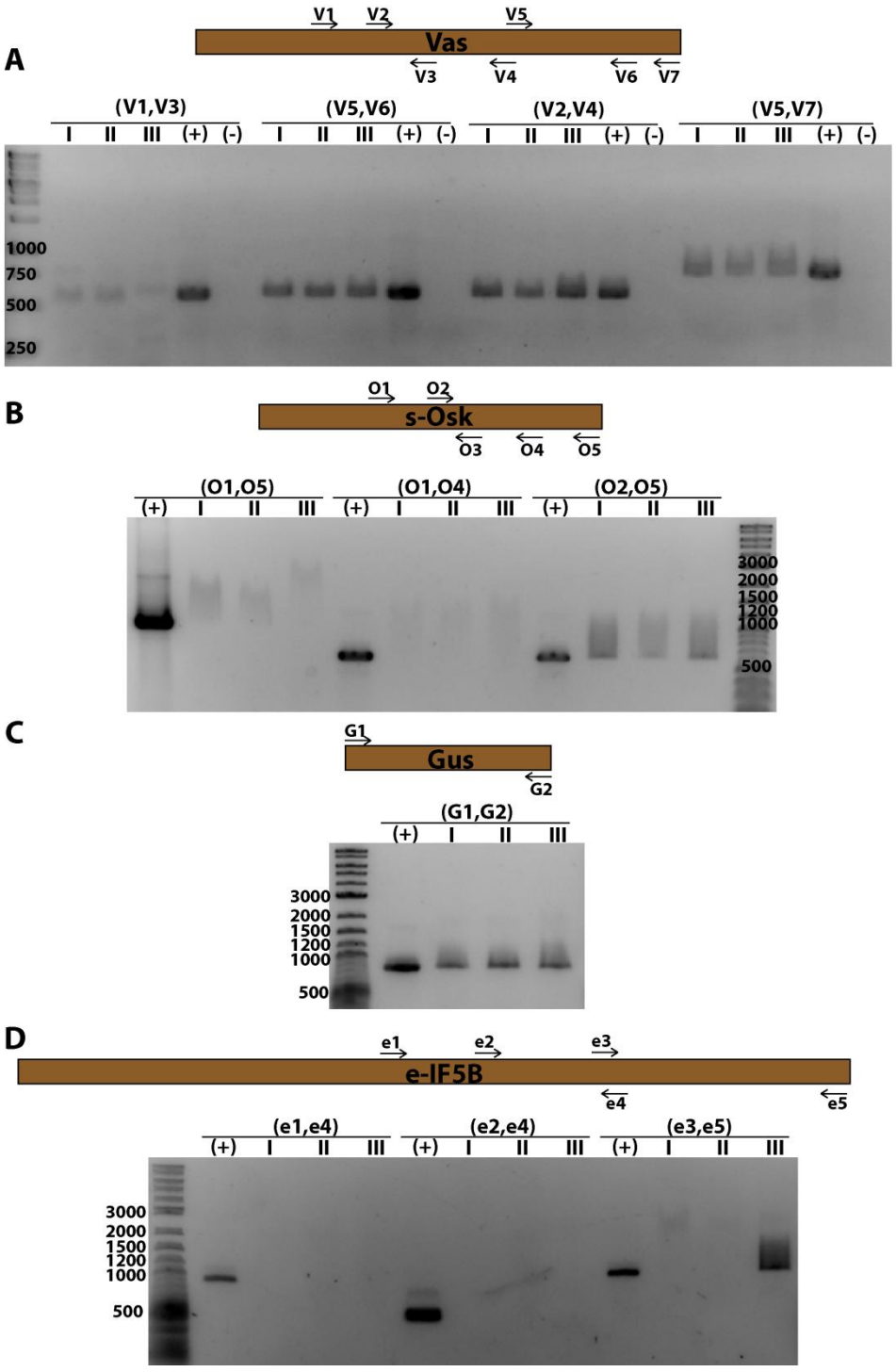


Fig. B.7 The cDNA library was tested for the presence of *vas* and its known interacting partners. Three DNA samples (I, II, III), independently extracted from the cDNA library were tested for presence of *vas* (**A**), *osk* (**B**), *gus* (**C**) and *eIF5B* (**D**). Schematics illustrate position of the primers used in different PCR reactions relative to the coding sequence of the tested isoform. DNA extracted from Y187 yeast cells with or without the corresponding cDNA plasmid was used as positive (+) or negative (-) control, respectively.

prey yeast strain, Y187. Additional sequencing revealed that all but one of these plasmids contained the same genes identified in the initial screen. One of the colonies, identified by PCR, to contain *CG2091,* appeared to also carry another plasmid encoding *unc13,* which was the only plasmid recovered. To confirm that the initial interactions observed in our screen can be reproduced, I first reexamined the interactions with vas^{+} (Fig. B.8 A, B). Next these nine genes were tested against $vas^{\Delta 655\text{-}661}$; however no difference was observed between vas^{+} and $vas^{\Delta 655\text{-}661}$ for their interactions with these genes (Fig. B.8 C). I further confirmed that the interactions of Pasha, mRpl44, CG2678, Rpp20, CG33978 and CR44309 with Vas is specific by testing them against DB-fused p53, as a negative control (Fig. B.8 D).

Discussion

Examining Vas interactions through a co-IP approach

The co-IP approach that I applied to isolate proteins interacting with Vas from *Drosophila* ovaries or embryos, although could efficiently purify eGFP-Vas, did not detect known interactions with Aub or Osk. This could be due to the transient nature of interactions that RNA helicases, such as Vas, have with other components of the RNP complexes (Xiol et al. 2014), supported by a control experiment showing that the same protocol could efficiently isolate other protein complexes. As a solution to this, a mutant form of Vas, defective in ATP hydrolysis, which stably clamps RNA and remains in a closed conformation, has been successfully used to co-IP several components of piRNA pathway in a *B. mori* ovarian cell line (Xiol et al. 2014). A similar approach with *Drosophila* ovaries enabled me to visualize Vls, one Vas-interacting protein, but did not detectably co-precipitate others that would be expected such as Aub or Osk. It is however not surprising that the interacting proteins are easier to isolate from a homogenous population of cells compared to more heterogeneous tissues, especially if only a small fraction of the tissue is composed of the cells containing particular protein complexes. For example, interactions between Vas and components of condensin I complex, such as Barren (Barr), which together with piRNA components Spn-E and Aub regulate mitotic progression in the germline stem cells (GSC) and cystoblasts (CB), are restricted to germanium (Pek et al. 2011). These Vas-containing complexes however were identified in *bag-of-marbles* (*bam*) mutant ovaries, which are enriched for GSC-like cells. Furthermore, the RNA-clamping mechanism described by Xiol et al. (2014), which provides a closed conformation of Vas thereby nucleating the assembly of an

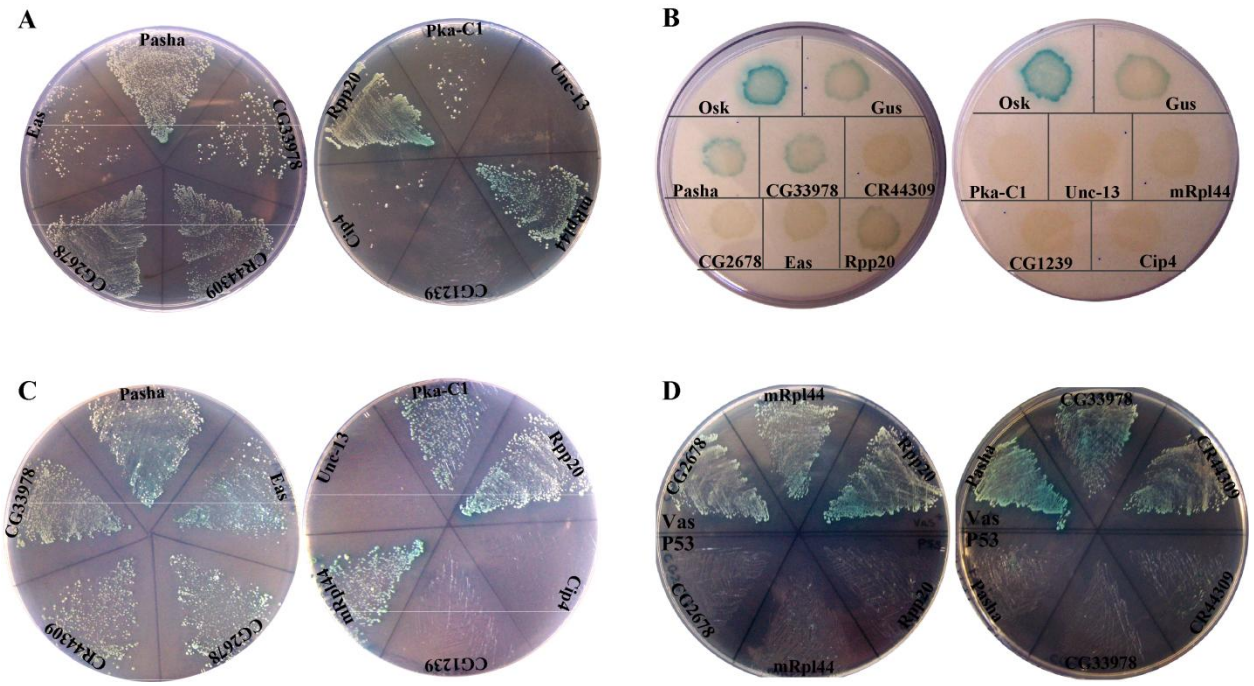


Fig. B.8 Among the genes identified to interact with *vas* through my yeast two-hybrid screens, nine that are expressed in the ovaries (*mRpl44, Rpp20, CG2678, pasha, CG2091, CG1239, Pka-C1, eas* and *cip4*) were isolated from yeast, amplified in *E. coli* and transformed back into Y187 yeast strain to subsequently test their interactions with *vas* mutant proteins. *CG33978* and *CR44309*, which are not reported as being expressed in ovaries, are also shown. *gus* was included as a positive control. Note that the original colony containing *CG2091* appeared to contain another plasmid, *unc13*, which was the only I could recover from yeast. **A**., **B**. To test if the initial results from my library screens are reproducible, I first tested each one of the interactions with vas^{+} by growing colonies on QDO plates (A), and by an X-Gal filter lift assay (B). **C**. These genes could also interact with $vas^{\Delta 655\text{-}661}$, similar to vas^{+}, inferred from their growth on QDO/A/X plates. The difference between vas^{+} and $vas^{\Delta 655\text{-}661}$ for their interactions with *cip4*, observed here, was not found reproducible. **D**. *Rpp20, mRpl44, CG2678, pasha, CG33978* and *CG44309* did not interact with P53, as the negative control.

amplifier complex, may not be a general mechanism for the interactions between Vas and its other partners such as Osk. Vas binding to Osk is proposed as the primary mechanism for anchoring Vas to pole plasm, suggesting that the two proteins have a relatively stable interaction (Breitwieser et al., 1996). Although, the difficulty to co-IP Osk with Vas, argues against this model, it is also worth mentioning that only a fraction of Vas in the ovaries is localized to the pole plasm, and maintenance of localization after initial anchoring may not require Vas to remain bound to Osk. Consistent with this, localization of RNP particles in the pole plasm is maintained through persistent trafficking on cortical microtubules mediated by motor proteins (Sinsimer et al., 2013). In addition, a large number of Vas interactions occur inside the pole plasm, which due to its distinct structure provides a separate cytoplasmic phase for localization of germ cell determinants (Mahowald, 1962). The same mechanism that prevents RNP complexes, inside the pole plasm, from being diffused into the surrounding cytoplasm may also prevent them from becoming soluble in most of the co-IP buffers that are designed to protect protein-protein interactions. A relatively high amount of Osk consistently found in the pellet following tissue lysis supports this hypothesis (Fig. B.2 D).

To stabilize transient complexes and weak interactions, *in vitro*, I used DSP. By having two reactive end groups and a spacer arm of 12A˚, DSP cross-links two proteins that are in close proximity (Lomant and Fairbanks, 1976). The primary amines, which DSP reacts with, have a positive charge in the physiological pH, and therefore are predominantly found on the surface of proteins readily accessible to cross-linking reagents. This cross-linking method, however, did not significantly improve co-IP of Vas with its known interacting partners. There are still a wide range of other cross-linking reagents which have different properties mostly based on their reactive groups, spacer arm length and water solubility, and thus could be advantageous for stabilizing different interactions (Kluger and Alagic, 2004). It should be also noted that in most *in vitro* protocols tissues are exposed to the cross-linking reagent at the homogenization step. If the complexes are transient or require the native environment of the cells for their assembly, an *in vivo* cross-linking procedure could stabilize them before the cell membrane is broken (Vasilescu et al., 2004). An *in vivo* cross-linking method using low concentrations of formaldehyde on permeabilized *Drosophila* embryos, for example, has been used to purify Tudor protein complexes (Gao et al., 2014).

In addition, my preliminary mass-spectrometry analyses on eluates from eGFP-Vas co-IP, which due to insufficient materials remained inconclusive, suggest that 35-70 mg total protein in the embryo lysates (instead of 7 mg), under the same co-IP protocol, would allow identification of low abundance proteins and accurate quantitative analyses.

The Y2H screen suggests new functions for Vas in miRNA pathway and ribosome biogenesis

Through a yeast two-hybrid screen using *vas* as the bait and a *Drosophil*a cDNA library, I found 19 genes encoding proteins with putative interactions with *vas*. Among this list there are nine candidates, which are also expressed in the ovaries and thus more likely to be genuine partners of Vas in different RNPs. As the next experiment the physical interaction between these proteins and Vas could be further tested through co-IP from ovaries. Similarly genetic interaction assays could reveal if these genes are involved in the same pathways as Vas. Two genes, *pasha* and *Rpp20,* are particularly interesting as, similar to *vas*, they are implicated in RNA processing (Eder et al., 1997; Han et al., 2006; Martin et al., 2009; Stolc et al., 1998).

Pasha (DGCR8 in human) and its partner Drosha have a central role in miRNA maturation process. Pasha specifically binds the double-stranded stem in the hairpin of the primary miRNAs and acts as an anchor to recruit Drosha, an RNAse III enzyme, which then cleaves the hairpin from its single stranded flanks. The resulted pre-miRNA is then exported to the cytoplasm for further processing (Han et al., 2006; Martin et al., 2009). *pasha* mutants in *Drosophila* die mostly as third-instar larva (Martin et al., 2009). In addition, germline clones homozygous for null alleles of *pasha*, similar to some other genes involved in miRNA pathway, exhibit defects in germline cell divisions, and produce egg chambers often composed of only 8 nurse cells and no oocyte (Azzam et al., 2012). Further examination of mutants for other components of miRNA pathway such as *ago1* indicates severely reduced germ cell divisions in the ovaries, resembling the phenotype observed in vas^{PH165} females with atrophied germarium (Azzam et al., 2012; Styhler et al., 1998). Although egg chambers with only 8 nurse cells are rare in vas^{PH165} females, these mutants more often produce egg chambers with 16 nurse cells and not containing any oocyte (Styhler et al. 1998). *vas* has been shown to bind piRNA precursors and facilitate their handover between components of piRNA pathway for further maturation (Xiol et al. 2014). The

possible role of *vas* in miRNA pathway, which is suggested by the direct interaction between *vas* and *pasha,* has never been explored. Pasha however is mostly found in the nucleus where the first step of miRNA maturation occurs. Thus cooperation between Vas and Pasha in this process would require co-localization of these two proteins in the nucleus. Such nuclear localization of Vas in the germline stem cells remains to be investigated by high resolution microscopy. Alternatively Pasha might have an unknown cytoplasmic function, which is associated to its interaction with Vas. It would be also important to quantify the canonical miRNAs for their decreased levels in vas^{PH165} ovaries similar to that in *pasha* null mutants (Martin et al., 2009).

Rpp20 in human and its yeast homolog *Rpp2* co-immunoprecipitate with the components of RNase P and RNase MRP, involved in processing of precursor tRNA and rRNA, respectively (Eder et al., 1997; Jarrous et al., 1998; Stolc et al., 1998). These two processes occur in the nucleus, whereas Vas is mainly found in the cytoplasm, making the association between Vas and Rpp20 rather puzzling. However, another study indicates that there are also small fractions of Rpp20 diffusely distributed throughout the cytoplasm, which, in response to stress, aggregate into granules (Hua and Zhou, 2004). The function of Rpp20 in cytoplasm is not known, but is proposed to be important for protection of RNA under stress condition, or in the cytoplasmic regions, such as axonal growth cones where mRNA is transported for local protein synthesis. Therefore, it would be interesting to further investigate the association of Vas with Rpp20 in *Drosophila* ovaries, which are another classical example of RNA localization and localized translation.

The previously known genes interacting with *vas*, namely *osk*, *gus* and *eIF5B*, were not found in the current Y2H screen. However, when these genes were directly tested against *vas*, using the same Y2H system, I observed a robust interaction. Our further examinations of the cDNA library also suggest that *osk* and perhaps some other ovarian partners of *vas*, are not enriched in this library. To effectively screen, through the ovarian cDNA clones, for the genes interacting with *vas*, a tissue specific cDNA library could be constructed and used. The cDNA libraries designed to be used with a *Gal4* based Y2H system, may result in discovering new proteins associated with Vas which have not been found in the previous screen using a *LexA*-based version of Y2H (Carrera et al., 2000; Sobhanifar, 2003; Van Criekinge and Beyaert, 1999).

Chapter 4: Conclusion and future directions

The research presented in this thesis addresses the role of different motifs of Vas in its various functions. We found that the least explored N- (amino acids 1-200) and C- (amino acids 623-661) terminal regions, despite being highly divergent, contain critical motifs for Vas functions *in vivo*. These studies confirm significance of the N-terminal RGG motifs for Vas function in germ cell development. Our analyses of eGFP-Vas$^{\Delta 636\text{-}646}$, lacking a C-terminal motif conserved among closely related *Drosophila* species, suggest that Vas plays a role in pole plasm condensation at the posterior region. *vas*$^{\Delta 636\text{-}646}$ mutants exhibit a diffused localization of Vas and other components of pole plasm at the posterior cortex, which does not allow pole buds to adopt the germ cell fate. Furthermore, we investigated the functional significance of a highly conserved Trp in the second last position of Vas sequence. These experiments using endogenous alleles of *vas* indicate that substitution of Trp660 with Glu abolishes germ cell formation and embryonic viability, and significantly reduces DV patterning and piRNA biogenesis. Our analyses also reveal that females carrying mutations in conserved helicase motifs of *vas* produce a number of eggs significantly higher than *vas* null, suggesting that Vas plays some functions in oogenesis independent of its helicase activity. In the following sections I will discuss some of the unanswered questions related to this research and propose directions for future studies.

Additional quantitative analyses on the role of RGG motifs

Our study shows that RGG motifs in the N-terminal region greatly contribute to Vas function in germ cell development; however, the specific RGG motifs being involved are not mapped. A systematic mutagenesis approach that targets different RGG motifs in the full length protein and changes those to AGG or KGG could address this question. The repeat-rich nature of the N-terminal sequence causes a challenge in site-directed mutagenesis, or CRISPR-Cas9 targeting; thus a more efficient approach would be to use synthetic DNA fragments encoding these mutations for generating transgenic alleles. To accurately quantify functional contribution of single RGG motifs, which could be subtle but significant, proteins need to be compared at the same expression level. Site-specific transgene integration is a popular method, which overcomes variability of position effects by inserting the transgene in a predefined site in the genome

(Bischof et al., 2007). In our initial experiments using this method *egfp-vas*$^{+}$, which was inserted in the commonly used landing site, *ZH-attP-86Fb* (Bischof et al., 2007), did not allow sufficient expression of the protein to support germ cell formation at levels approaching wild-type. Other studies, which have compared 20 different landing sites, using a ubiquitously expressed luciferase reporter, show that the expression differences among the majority of available lines are less than two fold (Markstein et al., 2008). This study further shows that expression levels in most tissues are boosted several fold when transgenes are flanked by a gypsy retrovirus insulator (Markstein et al., 2008). However, I did not find a significant change in eGFP-Vas expression upon inserting a gypsy insulator on both sides of the expression cassette (data not shown), suggesting that specific features of *egfp-vas* construct, such as the lack of introns, are responsible for the limited expression levels. Introns could regulate gene expression at different steps, including transcription, nuclear export, protein translation and mRNA stability; thus genomic constructs often result in more efficient gene expression than the cDNA constructs (Chorev and Carmel, 2012; Liu, 2013). Indeed a *vas* construct cloned into p{Pmat-tub67c:gfp} vector, which contains maternal *tubulin 67c* promoter followed by a 487 bp intron, has been shown to fully rescue germ cell formation in a vas^{1}/vas^{PH165} background (Micklem et al., 1997; Wang et al., 2015; our unpublished observations). Further highlighting the role of introns in enhancing Vas expression, I found that even using the maternal tubulin Gal4 to drive a *UAS-egfp-vas*, lacking introns, does not result in a robust expression (our unpublished data).

Another interesting experiment is to directly test the effect of RGG motifs on the interaction between Vas and its mRNA targets. I investigated the effect of different N-terminal truncations on the specific interaction between Vas and a (U)-rich motif derived from *mei-p26* 3′UTR (Liu et al., 2009). Similar *in vitro* assays using full-length mRNAs, containing all the sequences involved in Vas interaction, could be informative but are technically challenging. An alternative experiment is to quantify these target transcripts after co-immunoprecipitation with wild type or mutant forms of Vas from ovaries. RGG motifs could be also involved in protein-protein interactions, as several yeast proteins including Scd6, Sbp1 and Npl3, which accumulate in P bodies and repress mRNA translation, contain RGG motifs that mediate interaction with eIF4G (Rajyaguru et al., 2012). An interaction between Vas and eIF4G is also possible, and might have remained unknown due to technical limitation of co-IP and Y2H screens (Carrera et al., 2000; Xiol et al., 2014; my current study).

The role of Vas in pole plasm concentration

Our analysis of *egfp-vas*$^{\Delta 636\text{-}646}$ in *vas*^{1} background indicates that a focused localization of Vas, and perhaps the entire pole plasm, at the posterior region is essential for efficient germ cell formation. Our further examination in a wild type background shows that the localization defects of eGFP-Vas$^{\Delta 636\text{-}646}$ are only observed in the absence of wild type Vas, and thus are due to the impaired function of Vas$^{\Delta 636\text{-}646}$. Two mechanisms could be envisaged, through which Δ636-646 affects pole plasm concentration at the posterior: according to the first mechanism pole plasm components reach the posterior region of the embryo and their subsequent diffusion along the cortex is inhibited through a process that involves Vas; the second model would propose that pole plasm is initially diffuse in the posterior region but later becomes condensed through an active mechanism which is impaired in *vas*$^{\Delta 636\text{-}646}$. Our preliminary live imaging of the early embryos favours the second mechanism; however this needs to be confirmed through high resolution live imaging of pole plasm in late stage oocytes combined with automated image analysis techniques (Parton et al., 2011). Furthermore, and supporting either of the above mechanisms, stable anchoring of pole plasm components at the posterior cortex is shown to be a dynamic process, which requires persistent trafficking of germ plasm RNP particles on the posterior microtubules (Sinsimer et al., 2013). A screen for modifiers of *vas*$^{\Delta 636\text{-}646}$ phenotype could help to identify other factors interacting with Vas in fine tuning pole plasm concentration. In addition, immuno-electron microscopy analyses of Vas-containing particles in the pole plasm could compare *vas*$^{\Delta 636\text{-}646}$ with *vas*$^{+}$ at ultrastructural level.

Exploring Vas-protein and -RNA interactions

A particular focus of future studies should be to further examine Vas interacting partners. Our co-IP experiments did not efficiently capture Vas protein interactions in ovaries or embryos, as judged by the absence of known interacting partners such as Osk and Aub in the co-IP eluate, possibly due to the weak or transient nature of Vas interactions. One approach, which has been successfully used to identify Vas partners in *B. mori* BmN4 cell line, is to stabilize Vas complexes through mutations that impair ATP product dissociation, and thus generate an immobile RNA clamp (Xiol et al., 2014). This method, however, may not be ideal for analyzing the role of different motifs in Vas interactions, due to possible impact of the stabilizing mutation

on the interactions. Using the same principle a non-hydrolyzable ATP analog may also stabilize Vas complexes (Liu et al., 2014). The protocol then needs to be optimized as whether the ATP analog is added to the lysis buffer before tissue homogenization or if ovaries need to be cultured in a medium containing this compound.

Proximity-dependent biotin identification (BioID) is another approach, developed to identify neighboring and potentially interacting proteins that are targeted by a biotin-ligase fusion protein (Roux et al., 2012). The biotin-ligase used in this system is a modified version of the *E. coli* enzyme, BirA, which in its wild type form specifically biotinylates an acceptor peptide in the target proteins (Beckett et al., 1999). Co-expression of BirA and the protein of interest tagged with the biotin acceptor in mammalian cells, supplied with biotin, or *Drosophila* tissues, which have endogenous levels of biotin, allows subsequent affinity purification of biotinylated complexes using streptavidin conjugated resins (Kim et al., 2009; Struebbe et al., 2011). BioID uses a BirA mutant form (BirA*) that prematurely releases the highly reactive bioAMP and thus allows promiscuous protein biotinylation (Roux et al., 2012). The biotinylated proteins could be then purified and their identity could be determined by mass spectrometry. BioID is a promising technique for identifying novel interacting partners of Vas in ovaries or embryos expressing BirA*-Vas compared to BirA* alone. This method could also compare different mutant forms of Vas, which are fused to BirA*.

Some protein interactions are difficult to capture due to the low abundance of one or more proteins involved. For example, Vas interacts with condensin I complex components, Barr and Cap-D2, in only GSCs and CBs, which constitute a small portion of the ovaries (Pek and Kai, 2011a). These interactions have been identified through Vas co-IP from *bam* ovaries, which contain a large population of GSCs and CBs. To biochemically examine the effect of different *vas* mutations on such germarium-limited interactions, different *egfp-vas* transgenes need to be expressed ideally in double mutant ovaries for *bam* and vas^{PH165}. Alternatively endogenous alleles of *vas* could be generated by CRISPR-Cas9 method and tested in a *bam* mutant background.

Our yeast two-hybrid assays only indicated direct interactions between Vas and previously known partners, Osk, Gus and eIF5B; however, it remains possible that certain interactions strictly depend on the cellular environment in which they normally occur and thus are refractory

to Y2H assays. To dissect specific pathways that are affected by each *vas* mutation, genetic interaction assays could be carried out through a candidate approach. Depending on the availability of endogenous alleles, *vas* mutations could be tested in transheterozygous females also carrying a loss-of-function allele of the candidate partner.

The multiplicity of its functions suggests that Vas is involved in many RNP complexes associated with various pathways, and thus targeting different RNAs. Strong evidence indicates that Vas activates translation of *grk* and *mei-P26* and it also binds piRNA precursors (Carrera et al., 2000; Liu et al., 2009; Xiol et al., 2014; Zhang et al., 2012). In addition, a previous study, using 0-2 h embryo extracts, has identified ~ 200 mRNAs that co-immunoprecipitate with Vas (Liu et al., 2009). These mRNA targets include *gcl, bicD, cib*, *Bsg25D*, *pxt* and *aret*, which like Vas localize to the posterior region of early embryos (Lecuyer et al., 2007; Liu et al., 2009; Tomancak et al., 2007). The Vas core region alone confers helicase activity *in vitro* (Sengoku et al., 2006), however, our data indicate that the flanking N- and C-terminal sequences contribute to Vas function, possibly through unknown auxiliary motifs that regulate Vas affinity to mRNA molecules. To investigate this, RNA immunoprecipitation (RIP) followed by deep sequencing could compare different Vas mutants with wild type for their RNA targets in ovaries and embryos. High sensitivity of the current RNA sequencing techniques would also allow identification of new targets which might have remained unknown in the previous screen (Liu et al., 2009).

Vas was initially thought to bind RNA in a non-sequence-specific manner (Sengoku et al., 2006). However, Vas directly targets a (U)-rich motif in *mei-p26* 3′ UTR and this interaction is required for translational activation of Mei-P26 by Vas (Liu et al., 2009). Similar (U)-rich motifs are present in many mRNA co-purified with Vas, supporting that Vas is directly involved in sequence-specific targeting of mRNAs. A combination of cross-linking immunoprecipitation (CLIP) and RNase-mediated Vas footprint sequencing could help to further explore Vas-binding sites across transcriptome (Silverman et al., 2014; Singh et al., 2014; Zhang and Darnell, 2011). Yet, it is possible that the sequence specificity for a subpopulation of Vas targets is conferred by other proteins interacting with Vas, as the 12-nt Vas footprints on TE transcripts lack any nucleotide bias (Xiol et al., 2014).

Post-translational modifications of Vas

Different post-translational modifications (PTMs) of Vas orthologs, are implicated in regulating translational activity, localization and protein turn-over. Activation of meiotic check-points, in response to unrepaired DSBs, results in decreased accumulation of Grk, mediated by Vas phosphorylation, and causes dorsoventral patterning defects during oogenesis (Abdu et al., 2002; Ghabrial and Schupbach, 1999). Vas phosphorylation may also have regulatory effects on its other functions, which remains to be examined through phosphorylation site mapping and phosphomimetic mutation analyses.

In addition, Vas orthologs from mouse, *Xenopus*, and *D. melanogaster* carry both symmetrically and asymmetrically dimethylated arginines (sDMA, aDMA; Kirino et al., 2010). In *Drosophila* sDMA modifications depend on arginine protein methyltransferase 5 (dPRMT5/Csul/Dart5); whereas, aDMAs are produced by type I PRMTs, such as dPRMT1 (Dart1) and dPRMT4 (Dart4). aDMA modifications often occur in RGG motifs, which are abundant in the Vas N-terminus, and sDMAs occur in the motifs composed of an Arg flanked by Ala or Gly (R17, R65, R588, R644 in Vas; Kirino et al., 2010; Bedford and Richard, 2005; Krause et al., 2007). Methylarginine sites often serve as the binding motifs for Tud proteins; in *Drosophila,* for example, sDMA modifications of Aub are required for its binding to Tud, which in turn mediates localization of Aub to the pole plasm (Chen et al., 2011; Kirino et al., 2010b). Csul co-immunoprecipitates with wild type Vas transiently expressed in BmN4 cells (Xiol et al., 2014). In addition, Csul and its cofactor, Vls, are required for the assembly of the nuage and the pole plasm; however, the initial localization of Vas to *csul* genetically null oocytes appears normal (Anne and Mechler, 2005; Anne et al., 2007; Cavey et al., 2005; Kirino et al., 2010a). To address the significance of Vas arginine methylation in germ cell formation, I studied embryos maternally transheterozygous for vas^{1} and a mutant allele of *vls*, *csul* or *dart1*. Combination of vas^{1} and vls^{2}, which encodes a truncated protein of 52 amino acids (Anne and Mechler, 2005), resulted in embryos containing at least 22% fewer germ cells than either vls^{2}/+ or vas^{1}/+ (p=2.2E-10; data not shown). Transheterozygous embryos for vas^{1} and $csul^{e00797}$, which encodes a truncated protein missing the entire conserved methyltransferase domain (Gonsalvez et al., 2006) or a *dart1*-null allele (Kimura et al., 2008) exhibited less dramatic but significant decreases in germ cell numbers (13% and 10%, respectively). These results, however, should be

interpreted cautiously, as the combinatorial effects on germ cell number might be simply associated with the additive defects in pole plasm assembly, which requires most or all of these genes. Other preliminary experiments also show that the arginines 65, 73, 101, 108 and 150 are the actual targets for the arginine methyltransferases, but as a result of insufficient sample, the MALDI-TOF data are not statistically significant (Rappsilber, unpublished data). Thus in future, large scale immunopurification of Vas from ovaries followed by mass-spectrometry could accurately map methylated arginines in Vas. Moreover, our current functional analyses suggest that possibly critical arginine methylations occur on R163 and/or R588 as deletion or mutation of the other Arg residues did not have significant effects on Vas functions.

Vas accumulation in pole plasm depends on activities of Gus, Fsn, and Faf through regulating its ubiquitination (sections 1.4.4). Furthermore, acetylation of mouse Vas homologue, Mvh, modulates its RNA binding and is linked to its temporal regulation during spermatogenesis (sections 1.9.5). A comprehensive post-translational modification map of Vas would not only facilitate future mutation analysis but could also reveal yet unknown modifications regulating Vas functions in different pathways.

Further investigation on the role of Vas in cancer

Several germ cell specific genes are upregulated in *l(3)mbt* brain tumors, and a subset of those including *vas* significantly contribute to malignancy (Janic et al., 2010). Although the exact mechanism by which Vas promotes tumor formation is not clear, it is shown that acquisition of germline traits in somatic cells results in their increased fitness and survival consistent with the immortality observed in both germ cells and cancer cells (Curran et al., 2009). Our structure-function analyses provide useful information about different motifs of Vas with respect to its various functions in oogenesis and germ cell formation. These analyses could be extended to the role of Vas in tumor formation by expressing different mutant forms of Vas in $l(3)mbt^{ts1}$ and vas^{PH165} double mutants, and testing their ability to restore brain tumor formation. A comparison between the results of such analyses and our current data would provide useful links between *vas* function in tumorigenesis and in various germ cell specific pathways. I approached these experiments, however, I found highly variable brain overgrowths in $l(3)mbt^{ts1}$ homozygous larvae raised at restrictive temperature, which does not allow using the brain size as an accurate readout in these analyses. The first step in future studies would be to optimize the experimental

condition for obtaining a consistent and robust phenotype in *l(3)mbt* mutants. Alternatively, *l(3)mbt* brains could be further characterized for expression of different tumor markers, such as cell cycle regulating genes, which could be then used to assay Vas function in promoting malignancy (Saini and Reichert, 2012; Whitfield et al., 2006).

To investigate if ectopic expression of Vas is sufficient to transform normal cells into cancer cells the effect of Vas overexpression could be examined in different tissues using a Gal4/UAS system. Based on the results from these experiments Vas could be then overexpressed in the relevant cell lines to further study mechanism of its function at the molecular level. It would be also interesting to examine different human cell lines for over-expression of germline specific genes. The significance of these genes in malignancy could be then investigated through RNAi knock-down experiments or CRISPR-Cas9 gene targeting.

Vas role in inter-specific hybrid dysgenesis

In *Drosophila* inter- and intra-specific hybrid sterility is associated with TE derepression in F1 germline (Brennecke et al., 2008; Kelleher et al., 2012). The intra-specific hybrid sterility often occurs when males carrying active copies of TEs mate with females from other strain in which those transposons are absent or no longer active. In these cases the maternal piRNA cannot silence the paternal-specific TEs in F1 germline (Brennecke et al., 2008). In contrast, the hybrid dysgenesis observed between *Drosophila* species, for example *D. melanogaster* and *D. simulans*, is mostly correlated with divergence of ping-pong amplification components, such as Aub, and is not due to the differences in piRNA pool (Kelleher et al., 2012). The sequence divergence of Vas, as another critical component of piRNA pathway, may also contribute to sterility of *D. melanogaster* and *D. simulans* hybrids, although Vas localization to the nuage appears normal. One possible experiment is to express *vas* orthologs from other *Drosophila* species in vas^{PH165} ovaries and examine their ability to suppress transposons, such as HeT-A. The N-terminal region of Vas, which contains essential elements for its localization, is highly diverse even among closely related *Drosophila* species, thus I would expect defects in piRNA pathway at least due to mislocalization of some Vas orthologs in *D. melanogaster* ovaries. For those Vas orthologs that fail to suppress TEs, despite their normal localization, chimeric constructs could be tested, where different motifs are replaced by the counterpart domains from *D. melanogaster*

vas. These analyses would provide valuable information about contribution of Vas to inter-specific hybrid sterility, specifically associated with piRNA pathway, and about the critical motifs of Vas being involved.

References

Abdu, U., Brodsky, M., and Schupbach, T. (2002). Activation of a meiotic checkpoint during *Drosophila* oogenesis regulates the translation of *gurken* through Chk2/Mnk. Curr. Biol. *12*, 1645-1651.

Ables, E.T. (2015). *Drosophila* oocytes as a model for understanding meiosis: An educational primer to accompany "Corolla is a novel protein that contributes to the architecture of the synaptonemal complex of *Drosophila*". Genetics *199*, 17-23.

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. *19*, 568-576.

Alie, A., Hayashi, T., Sugimura, I., Manuel, M., Sugano, W., Mano, A., Satoh, N., Agata, K., and Funayama, N. (2015). The ancestral gene repertoire of animal stem cells. Proc Natl Acad Sci *112*, E7093-E7100.

Alie, A., Leclere, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., Queinnec, E., and Manuel, M. (2011). Somatic stem cells express *piwi* and *vasa* genes in an adult ctenophore: Ancient association of "germline genes" with stemness. Dev. Biol. *350*, 183-197.

Altmann, M., Muller, P.P., Wittmer, B., Ruchti, F., Lanker, S., and Trachsel, H. (1993). A *Sacharomyces cerevisiae* homolog of mammalian translation initiation factor-4B contributes to RNA helicase activity. Embo J. *12*, 3997-4003.

Anand, A., and Kai, T. (2012). The tudor domain protein Kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. Embo J. *31*, 870-882.

Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science *313*, 1968-1972.

Anne, J. (2010). Targeting and anchoring Tudor in the pole plasm of the *Drosophila* oocyte. Plos One *5*. e14362

Anne, J., and Mechler, B.M. (2005). Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuleen. Development *132*, 2167-2177.

Anne, J., Ollo, R., Ephrussi, A., and Mechler, B.M. (2007). Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in *Drosophila*. Development *134*, 137-146.

Arnold, S.J., Maretto, S., Islam, A., Bikoff, E.K., and Robertson, E.J. (2006). Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. Dev. Biol. *296*, 104-118.

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., and Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. Nat. Cell Biol. *1*, 431-437.

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*, 1161-1169.

Auerbach, D., and Stagljar, I. (2005). Yeast two-hybrid protein-protein interaction networks. In Waksman G. (ed), Proteomics and Protein–Protein Interactions: Biology, Chemistry, Bioinformatics, and Drug Design 19-31 (New York, Springer).

Ayme-Southgate, A., Bounaix, C., Riebe, T.E., and Southgate, R. (2004). Assembly of the giant protein projectin during myofibrillogenesis in *Drosophila* indirect flight muscles. BMC cell biol. *5*, 1.

Azzam, G., Smibert, P., Lai, E.C., and Liu, J.L. (2012). *Drosophila* Argonaute 1 and its miRNA biogenesis partners are required for oocyte formation and germline cell division. Dev. Biol. *365*, 384-394.

Babu, K., Cai, Y., Bahri, S., Yang, X.H., and Chia, W. (2004). Roles of Bifocal, Homer, and F-actin in anchoring Oskar to the posterior cortex of *Drosophila* oocytes. Genes Dev. *18*, 138-143.

Barbosa, C., Peixeiro, I., and Romao, L. (2013). Gene expression regulation by upstream open reading frames and human disease. Plos Genetics *9*.

Bartfai, R., and Orban, L. (2003). The *vasa* locus in zebrafish: Multiple RGG boxes from intragenic duplications. DNA and Cell Biol. *22*, 47-54.

Baskin, L.S., and Yang, C.S. (1982). Cross-linking studies of the protein topography of rat liver microsomes. Biochim. Biophys. Acta. *684*, 263-271.

Bates, G.J., Nicol, S.M., Wilson, B.J., Jacobs, A.M.F., Bourdon, J.C., Wardrop, J., Gregory, D.J., Lane, D.P., Perkins, N.D., and Fuller-Pace, F.V. (2005). The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. Embo J. *24*, 543-553.

Becalska, A.N., and Gavis, E.R. (2009). Lighting up mRNA localization in *Drosophila* oogenesis. Development *136*, 2493-2503.

Beckett, D., Kovaleva, E., and Schatz, P.J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci. *8*, 921-929.

Bedford, M.T., and Richard, S. (2005). Arginine methylation: An emerging regulator of protein function. Mol. Cell *18*, 263-272.

Berg, C.A. (2005). The *Drosophila* shell game: patterning genes and morphological change. Trends Genet. *21*, 346-355.

Berghmans, S., Murphey, R.D., Wienholds, E., Neuberg, D., Kutok, J.L., Fletcher, C.D.M., Morris, J.P., Liu, T.X., Schulte-Merker, S., Kanki, J.P., *et al.* (2005). tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. Proc Natl Acad Sci *102*, 407-412.

Bergsten, S.E., and Gavis, E.R. (1999). Role for mRNA localization in translational activation but not spatial restriction of nanos RNA. Development *126*, 659-669.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryos. Embo J. *7*, 1749-1756.

Beshore, E.L., McEwen, T.J., Jud, M.C., Marshall, J.K., Schisa, J.A., and Bennett, K.L. (2011). *C. elegans* Dicer interacts with the P-granule component GLH-1 and both regulate germline RNPs. Dev. Biol. *350*, 370-381.

Betts, M.J., and Russell, R.B. (2003). Amino acid properties and consequences of substitutions. Bioinformatics for geneticists *317*, 289.

Bhat, M.A., Philp, A.V., Glover, D.M., and Bellen, H.J. (1996). Chromatid segregation at anaphase requires the barren product, a novel chromosome-associated protein that interacts with topoisomerase II. Cell *87*, 1103-1114.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phi C31 integrases. Proc Natl Acad Sci *104*, 3312-3317.

Boisvert, F.M., Cote, J., Boulanger, M.C., Cleroux, P., Bachand, F., Autexier, C., and Richard, S. (2002). Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. J. Cell Biol. *159*, 957-969.

Boisvert, F.M., Cote, J., Boulanger, M.C., and Richard, S. (2003). A proteomic analysis of arginine-methylated protein complexes. Mol. Cell. Proteomics *2*, 1319-1330.

Borman, A.M., Michel, Y.M., and Kean, K.M. (2000). Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. Nucleic Acids Res *28*, 4068-4075.

Boswell, R.E., and Mahowald, A.P. (1985). *tuduor*, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. Cell *43*, 97-104.

Braat, A.K., van de Water, S., Korving, J., and Zivkovic, D. (2001). A zebrafish *vasa* morphant abolishes Vasa protein but does not affect the establishment of the germline. Genesis *30*, 183-185.

Branciforte, D., and Martin, S.L. (1994). Developmental and cell type specificity of line-1 expression in mouse testis - implications for transposition. Mol. and Cell. Biol. *14*, 2584-2592.

Breeden, L., and Nasmyth, K. (1985). Regulation of the yeast HO gene. Cold Spring Harb. Symp. Quant. Biol. *50,* 643-50

Breitwieser, W., Markussen, F.H., Horstmann, H., and Ephrussi, A. (1996). Oskar protein interaction with Vasa represents an essential step in polar granule assembly. Genes Deve. *10*, 2179-2188.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell *128*, 1089-1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387-1392.

Bullock, S.L., and Ish-Horowicz, D. (2001). Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis. Nature *414*, 611-616.

Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of function of RNA-binding proteins. Science *265*, 615-621.

Buszczak, M., and Spradling, A.C. (2006). The *Drosophila* P68 RNA helicase regulates transcriptional deactivation by promoting RNA release from chromatin. Genes Dev. *20*, 977-989.

Campos-Ortega, J.A., Hartenstein, V. (1985). The Embryonic Development of *Drosophila melanogaster* (Heidelberg, Springer-Verlag).

Capowski, E.E., Martin, P., Garvin, C., and Strome, S. (1991). Identificatin of grandchildless loci whose products are required for normal germline development in the nematode *Caenorhabditis elegans*. Genetics *129*, 1061-1072.

Caretti, G., Schiltz, R.L., Dilworth, F.J., Di Padova, M., Zhao, P., Ogryzko, V., Fuller-Pace, F.V., Hoffman, E.P., Tapscott, S.J., and Sartorelli, V. (2006). The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. Dev. Cell *11*, 547-560.

Carmel, A.B., and Matthews, B.W. (2004). Crystal structure of the BstDEAD N-terminal domain: A novel DEAD protein from *Bacillus stearothermophilus*. RNA *10*, 66-74.

Carre, D., Djediat, C., and Sardet, C. (2002). Formation of a large Vasa-positive germ granule and its inheritance by germ cells in the enigmatic *Chaetognaths*. Development *129*, 661-670.

Carrera, P., Johnstone, O., Nakamura, A., Casanova, J., Jackle, H., and Lasko, P. (2000). Vasa mediates translation through interaction with a *Drosophila* yIF2 homolog. Mol. Cell *5*, 181-187.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.

Caruthers, J.M., Johnson, E.R., and McKay, D.B. (2000). Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. Proc Natl Acad Sci *97*, 13080-13085.

Castrillon, D.H., Quade, B.J., Wang, T.Y., Quigley, C., and Crum, C.P. (2000). The human *vasa* gene is specifically expressed in the germ cell lineage. Proc Natl Acad Sci *97*, 9585-9590.

Cavaliere, V., Taddei, C., and Gargiulo, G. (1998). Apoptosis of nurse cells at the late stages of oogenesis of *Drosophila melanogaster*. Dev. Genes Evol. *208*, 106-112.

Cavey, M., Hijal, S., Zhang, X.L., and Suter, B. (2005). *Drosophila valois* encodes a divergent WD protein that is required for Vasa localization and Oskar protein accumulation. Development *132*, 459-468.

Chagnovich, D., and Lehmann, R. (2001). Poly(A)-independent regulation of maternal *hunchback* translation in the *Drosophila* embryo. Proc Natl Acad Sci *98*, 11359-11364.

Chan, C.C., Dostie, J., Diem, M.D., Feng, W.Q., Mann, M., Rappsilber, J., and Dreyfuss, G. (2004). eIF4A3 is a novel component of the exon junction complex. RNA *10*, 200-209.

Chan, S.K., and Struhl, G. (1997). Sequence-specific RNA binding by Bicoid. Nature *388*, 634-634.

Chan, T.A., Hermeking, H., Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1999). 14-3-3 sigma is required to prevent mitotic catastrophe after DNA damage. Nature *401*, 616-620.

Chang, H., and Matzuk, M.M. (2001). Smad5 is required for mouse primordial germ cell development. Mech. Dev. *104*, 61-67.

Chang, J.S., Tan, L.H., and Schedl, P. (1999). The *Drosophila* CPEB homolog, Orb, is required for Oskar protein expression in oocytes. Dev. Biol. *215*, 91-106.

Chang, J.S., Tan, L.H., Wolf, M.R., and Schedl, P. (2001). Functioning of the *Drosophila orb* gene in *gurken* mRNA localization and translation. Development *128*, 3169-3177.

Chatterjee, S., Azad, B.B., and Nimmagadda, S. (2014). The intricate role of CXCR4 in cancer. Emerging Applications of Molecular Imaging to Oncology *124*, 31-82.

Chen, C., Nott, T.J., Jin, J., and Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. Nat. Rev. Mol. Cell. Biol. *12*, 629-642.

Chen, J.Y.F., Stands, L., Staley, J.P., Jackups, R.R., Latus, L.J., and Chang, T.H. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol. Cell *7*, 227-232.

Chen, Y., Potratz, J.P., Tijerina, P., Del Campo, M., Lambowitz, A.M., and Russell, R. (2008). DEAD-box proteins can completely separate an RNA duplex using a single ATP. Proc Natl Acad Sci *105*, 20203-20208.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., and Dumais, J. (2011). The transcriptional diversity of 25 *Drosophila* cell lines. Genome res. *21*, 301-314.

Chicoine, J., Benoit, P., Gamberi, C., Paliouras, M., Simonelig, M., and Lasko, P. (2007). Bicaudal-C recruits CCR4-NOT deadenylase to target mRNAs and regulates oogenesis, cytoskeletal organization, and its own expression. Dev. Cell *13*, 691-704.

Chintapalli, V.R., Wang, J., and Dow, J.A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. *39*, 715-720.

Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., and Sonenberg, N. (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos. Curr. Biol *16*, 2035-2041.

Chorev, M., and Carmel, L. (2012). The function of introns. Front. Genet. *3*, 55-55.

Chuang, R.Y., Weaver, P.L., Liu, Z., and Chang, T.H. (1997). Requirement of the DEAD-box protein Ded1p for messenger RNA translation. Science *275*, 1468-1471.

Cinalli, R.M., and Lehmann, R. (2013). A spindle-independent cleavage pathway controls germ cell formation in *Drosophila*. Nat. Cell Biol. *15*, 839-845.

Clancy, S. (2008). RNA splicing: introns, exons and spliceosome. Nat. Educ. *1*, 31.

Clark, A., Meignin, C., and Davis, I. (2007). A Dynein-dependent shortcut rapidly delivers axis determination transcripts into the *Drosophila* oocyte. Development *134*, 1955-1965.

Clark, D.V., Sabl, J.F., and Henikoff, S. (1998). Repetitive arrays containing a housekeeping gene have altered polytene chromosome morphology in *Drosophila*. Chromosoma *107*, 96-104.

Clouse, K.N., Ferguson, S.B., and Schupbach, T. (2008). Squid, Cup, and PABP55B function together to regulate *gurken* translation in *Drosophila*. Dev. Biol. *313*, 713-724.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. RNA *7*, 1717-1727.

Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17-37.

Coward, S.J. (1974). Chromatoid bodies in somatic cells of planarian – observations on their behavior during mitosis. Anat. Rec. *180*, 533-546.

Craig, A.W.B., Haghighat, A., Yu, A.T.K., and Sonenberg, N. (1998). Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. Nature *392*, 520-523.

Crittenden, S.L., Bernstein, D.S., Bachorik, J.L., Thompson, B.E., Gallegos, M., Petcherski, A.G., Moulder, G., Barstead, R., Wickens, M., and Kimble, J. (2002). A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*. Nature *417*, 660-663.

Curran, S.P., Wu, X., Riedel, C.G., and Ruvkun, G. (2009). A soma-to-germline transformation in long-lived *Caenorhabditis elegans* mutants. Nature *459*, 1079-U1060.

D'Agostino, I., Merritt, C., Chen, P.L., Seydoux, G., and Subramaniam, K. (2006). Translational repression restricts expression of the *C. elegans* Nanos homolog Nos-2 to the embryonic germline. Dev. Biol. *292*, 244-252.

Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T., and Darnell, R.B. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell *107*, 489-499.

de Jesus, A.J., and Allen, T.W. (2013). The role of tryptophan side chains in membrane protein anchoring and hydrophobic mismatch. Biochim. Biophys. Acta *1828*, 864-876.

Dearolf, C.R., Topol, J., and Parker, C.S. (1989). The *caudal* gene product is a direct activator of *fushi tarazu* transcription during *Drosophila* embryogenesis. Nature *341*, 340-343.

Dehghani, M., and Lasko, P. (2015). *In vivo* mapping of the functional regions of the DEAD-box helicase Vasa. Biol. Open *4*, 450-U457.

Del Campo, M., and Lambowitz, A.M. (2009). Structure of the yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. Mol. Cell *35*, 598-609.

Deng, W., and Lin, H.F. (2002). *miwi*, a murine homolog of *piw*i, encodes a cytoplasmic protein essential for spermatogenesis. Dev. Cell *2*, 819-830.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the microprocessor complex. Nature *432*, 231-235.

Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of poly(A)-binding protein through PABP-interacting proteins. Cold Spring Harb. Symp. Quant. Biol. *71*, 537-543.

Deshler, J.O., Highett, M.I., Abramson, T., and Schnapp, B.J. (1998). A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. Curr. Biol. *8*, 489-496.

Deshpande, G., Calhoun, G., Yanowitz, J.L., and Schedl, P.D. (1999). Novel functions of *nanos* in downregulating mitosis and transcription during the development of the *Drosophila* germline. Cell *99*, 271-281.

Desset, S., Meignin, C., Dastugue, B., and Vaury, C. (2003). COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. Genetics *164*, 501-509.

Diges, C.M., and Uhlenbeck, O.C. (2001). *Escherichia coli* DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. Embo J. *20*, 5503-5512.

Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. Curr. Biol. *12*, 1060-1067.

Dougherty, D.A. (2007). Cation-n interactions involving aromatic amino acids. J. Nutr. *137*, 15045-15085.

Driever, W., and Nüsslein-Volhard, C. (1989). The Bicoid protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. Nature *337*, 138-143.

Duncan, T., Osawa, Y., Kutty, R.K., Kutty, G., and Wiggert, B. (1999). Heme-binding by *Drosophila* retinoid-and fatty acid-binding glycoprotein (RFABG), a member of the proapolipophorin gene family. J. Lipid Res. *40*, 1222-1228.

Eder, P.S., Kekuda, R., Stolc, V., and Altman, S. (1997). Characterization of two scleroderma autoimmune antigens that copurify with human ribonuclease P. Proc Natl Acad Sci *94*, 1101-1106.

Engel, J., and Wu, C.F. (1994). Altered mechanoreceptor response in *Drosophila* bang-sensitive mutants. J. Comp. Physiol. *175*, 267-278.

Ephrussi, A., Dickinson, L.K., and Lehmann, R. (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant *nanos*. Cell *66*, 37-50.

Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by Oskar. Nature *358*, 387-392.

Feichtinger, J., Larcombe, L., and McFarlane, R.J. (2014). Meta- analysis of expression of *l(3)mbt* tumor- associated germline genes supports the model that a soma- to- germline transition is a hallmark of human cancers. Int. J. Cancer *134*, 2359-2365.

Fekete, C.A., Applefield, D.J., Blakely, S.A., Shirokikh, N., Pestova, T., Lorsch, J.R., and Hinnebusch, A.G. (2005). The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection *in vivo*. Embo J. *24*, 3588-3601.

Findley, S.D., Tamanaha, M., Clegg, N.J., and Ruohola-Baker, H. (2003). Maelstrom, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development *130*, 859-871.

FlyBase Curators, Swiss-Prot Project Members, InterPro Project Members (2004-). Gene Ontology annotation in FlyBase through association of InterPro records with GO terms.

FlyBase Genome Annotators (2012). Changes affecting gene model number or type in release 5.49 of the annotated *D. melanogaster* genome.

Forbes, A., and Lehmann, R. (1998). Nanos and Pumilio have critical roles in the development and function of *Drosophila* germline stem cells. Development *125*, 679-690.

Forrest, K.M., and Gavis, E.R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for *nanos* mRNA localization in *Drosophila*. Curr. Biol *13*, 1159-1168.

Forristall, C., Pondel, M., Chen, L.H., and King, M.L. (1995). Patterns of localization and cytoskeletal association of 2 vegetally localized RNAs, VG1 and XCAT-2. Development *121*, 201-208.

Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H., Furusawa, M., and Noce, T. (1994). Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila* Vasa and its specific expression in germ cell lineage. Proc Natl Acad Sci *91*, 12258-12262.

Fuller-Pace, F.V., and Ali, S. (2008). The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators. Biochim. Soc. Trans. *36*, 609-612.

Furuhashi, H., Takasaki, T., Rechtsteiner, A., Li, T., Kimura, H., Checchi, P.M., Strome, S., and Kelly, W.G. (2010). Trans-generational epigenetic regulation of *C. elegans* primordial germ cells. Epigenet. Chromatin *3,* 15.

Gallo, C.M., Wang, J.T., Motegi, F., and Seydoux, G. (2010). Cytoplasmic partitioning of P granule components is not required to specify the germline in *C. elegans*. Science *330*, 1685-1689.

Gao, M., McCluskey, P., Loganathan, S.N., and Arkov, A.L. (2014). An *in vivo* crosslinking approach to isolate protein complexes from *Drosophila* embryos. J. Vis. Exp. 2014 Apr 23;(86).

Gavis, E.R., and Lehmann, R. (1992). Localization of *nanos* RNA controls embryonic polarity. Cell *71*, 301-313.

Geigy, R. (1931). Action de l'ultra-violet sur le pole germinal dans l'oeuf de *Drosophila melanogaster*. Rev. suisse zool. *38*, 187-288.

Ghabrial, A., and Schupbach, T. (1999). Activation of a meiotic checkpoint regulates translation of Gurken during *Drosophila* oogenesis. Nat. Cell Biol. *1*, 354-357.

Ghosh, S., and Lasko, P. (2015). Loss-of-function analysis reveals distinct requirements of the translation initiation factors eIF4E, eIF4E-3, eIF4G and eIF4G2 in *Drosophila* spermatogenesis. Plos One *10*, e0122519

Gill, M.E., Hu, Y.C., Lin, Y., and Page, D.C. (2011). Licensing of gametogenesis, dependent on RNA binding protein Dazl, as a gateway to sexual differentiation of fetal germ cells. Proc Natl Acad Sci *108*, 7443-7448.

Gingras, A.C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. Annu. Rev. Biochem. *68*, 913-963.

Gönczy, P. and Rose, L. S. Asymmetric cell division and axis formation in the embryo. In WormBook, The *C. elegans* Research Community ed., doi/10.1895/wormbook.1.30.1.

Gonsalvez, G.B., Rajendra, T.K., Tian, L.P., and Matera, A.G. (2006). The Sm-protein methyltransferase, Dart5, is essential for germ-cell specification and maintenance. Curr. Biol *16*, 1077-1089.

González-Reyes, A., Elliott, H., and Stjohnston, D. (1995). Polarization of both major body axes in *Drosophila* by Gurken-Torpedo signaling. Nature *375*, 654-658.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., and Booth, B.W. (2011). The developmental transcriptome of *Drosophila melanogaster*. Nature *471*, 473-479.

Grieder, N.C., de Cuevas, M., and Spradling, A.C. (2000). The fusome organizes the microtubule network during oocyte differentiation in *Drosophila*. Development *127*, 4253-4264.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-U1115.

Grohman, J.K., Del Campo, M., Bhaskaran, H., Tijerina, P., Lambowitz, A.M., and Russell, R. (2007). Probing the mechanisms of DEAD-Box proteins as general RNA chaperones: The C-terminal domain of CYT-19 mediates general recognition of RNA. Biochemistry *46*, 3013-3022.

Gruidl, M.E., Smith, P.A., Kuznicki, K.A., McCrone, J.S., Kirchner, J., Roussell, D.L., Strome, S., and Bennett, K.L. (1996). Multiple potential germline helicases are components of the germline-specific P granules of *Caenorhabditis elegans*. Proc Natl Acad Sci *93*, 13837-13842.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. Science *315*, 1587-1590.

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 Dev. *12*, 1652-1664.

Guo, T., Peters, A.H.F.M., and Newmark, P.A. (2006). A *bruno*-like gene is required for stem cell maintenance in planarians. Dev. Cell *11*, 159-169.

Gustafson, E.A., Yajima, M., Juliano, C.E., and Wessel, G.M. (2011). Post-translational regulation by Gustavus contributes to selective Vasa protein accumulation in multipotent cells during embryogenesis. Dev. Biol. *349*, 440-450.

Haase, A.D. (2016). A small RNA-based immune system defends germ cells against mobile genetic elements. Stem Cells Int. *2016*, 7595791-7595791.

Hachet, O., and Ephrussi, A. (2004). Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. Nature *428*, 959-963.

Hake, L.E., and Richter, J.D. (1994). CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. Cell *79*, 617-627.

Han, J.J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y.J., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). *Drosophila* Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature *451*, 730-U737.

Hardin, J.W., Hu, Y.X., and McKay, D.B. (2010). Structure of the RNA binding domain of a DEAD-box helicase bound to its ribosomal RNA target reveals a novel mode of recognition by an RNA recognition motif. . *402*, 412-427.

Harris, A.N., and Macdonald, P.M. (2001). *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. Development *128*, 2823-2832.

Harrison, R.E., and Huebner, E. (1997). Unipolar microtubule array is directly involved in nurse cell-oocyte transport. Cell Motil. Cytoskel. *36*, 355-362.

Hartung, O., Forbes, M.M., and Marlow, F.L. (2014). Zebrafish Vasa is required for germ-cell differentiation and maintenance. Mol. Reprod. Dev. *81*, 946-961.

Hashimoto, H., Sudo, T., Mikami, Y., Otani, M., Takano, M., Tsuda, H., Itamochi, H., Katabuchi, H., Ito, M., and Nishimura, R. (2008). Germ cell specific protein Vasa is over-expressed in epithelial ovarian cancer and disrupts DNA damage-induced G2 checkpoint. Gynecol. Oncol. *111*, 312-319.

Hashimoto, Y., Maegawa, S., Nagai, T., Yamaha, E., Suzuki, H., Yasuda, K., and Inoue, K. (2004). Localized maternal factors are required for zebrafish germ cell formation. Dev. Biol. *268*, 152-161.

Hashiyama, K., Hayashi, Y., and Kobayashi, S. (2011). *Drosophila sex lethal* gene initiates female development in germline progenitors. Science *333*, 885-888.

Hausen, P., and Riebesell, M. (1991). The early development of *Xenopus Laevis*: An atlas of histology (Heidelberg, Springer-Verlag).

Havin, L., Git, A., Elisha, Z., Oberman, F., Yaniv, K., Schwartz, S.P., Standart, N., and Yisraeli, J.K. (1998). RNA-binding protein conserved in both microtubule- and microfilament-based RNA localization. Genes Dev. *12*, 1593-1598.

Hay, B., Ackerman, L., Barbel, S., Jan, L.Y., and Jan, Y.N. (1988a). Identification of a component of *Drosophila* polar granules. Development *103*, 625-640.

Hay, B., Jan, L.Y., and Jan, Y.N. (1988b). A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. Cell *55*, 577-587.

Hay, B., Jan, L.Y., and Jan, Y.N. (1990). Localization of Vasa, a component of *Drosophila* polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity. Development *109*, 425-433.

Hayashi, K., de Sousa Lopes, S.M.C., and Surani, M.A. (2007). Germ cell specification in mice. Science *316*, 394-396.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519-532.

Hayashi, Y., Hayashi, M., and Kobayashi, S. (2004). Nanos suppresses somatic cell fate in *Drosophila* germ line. Proc Natl Acad Sci *101*, 10338-10342.

Heasman, J., Quarmby, J., and Wylie, C.C. (1984). The mitochondrial cloud of *Xenopus* oocyte – the source of germinal granule material. Dev. Biol. *105*, 458-469.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.F., Chakraborty, A., and Gleizes, P.E. (2015). An overview of pre-rRNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA *6*, 225-242.

Hernandez, G., Altmann, M., Sierra, J.M., Urlaub, H., del Corral, R.D., Schwartz, P., and Rivera-Pomar, R. (2006). Functional analysis of seven genes encoding eight translation initiation factor 4E (eIF4E) isoforms in *Drosophila*. Mech. Dev. *122*, 529-43.

Hilbert, M., Karow, A.R., and Klostermeier, D. (2009). The mechanism of ATP-dependent RNA unwinding by DEAD-box proteins. Biol. Chem. *390*, 1237-1250.

Hilbert, M., Kebbel, F., Gubaev, A., and Klostermeier, D. (2011). eIF4G stimulates the activity of the DEAD box protein eIF4A by a conformational guidance mechanism. Nucleic Acids Res. *39*, 2260-2270.

Hilliker, A., Gao, Z., Jankowsky, E., and Parker, R. (2011). The DEAD-Box protein Ded1 modulates translation by the formation and resolution of an elF4F-mRNA complex. Mol. Cell *43*, 962-972.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. *83*, 779–812

Hirakata, S., and Siomi, M.C. (2016). piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. Biochim. Biophys. Acta. *1859*, 82-92.

Hoey, T., Doyle, H.J., Harding, K., Wedeen, C., and Levine, M. (1986). Homeo box gene expression in anterior and posterior regions of the *Drosophila* embryo. Proc Natl Acad Sci *83*, 4809-4813.

Hordern, J.S., Leonard, J.D., and Scraba, D.G. (1979). Structure of the mengo virion .VI. Spatial relationships of the capsid polypeptides as determined by chemical cross-linking analyses. Virology *97*, 131-140.

Hori, I. (1982). An ultrastructural study of the chromatoid body in planarian regenerative cells. J Electron Microsc. *31*, 63-72.

Horner, V.L., and Wolfner, M.F. (2008). Transitioning from egg to embryo: Triggers and mechanisms of egg activation. Dev. Dyn. *237*, 527-544.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., *et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell *129*, 69-82.

Hua, Y.M., and Zhou, J.H. (2004). Rpp20 interacts with SMN and is re-distributed into SMN granules in response to stress. Biochem. Biophys. Res. Commun. *314*, 268-276.

Huang, X.A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A major epigenetic programming mechanism guided by piRNAs. Dev. Cell *24*, 502-516.

Huang, Y.Z., Baker, R.T., and Fischervize, J.A. (1995). Control of cell fate by a deubiquitinating enzyme encoded by the *fat facets* gene. Science *270*, 1828-1831.

Hubbard, E.J.A., and Greenstein, D. (2005). Introduction to the germline. In WormBook, The *C. elegans* Research Community *ed.,* doi: 10.1895/wormbook.1.18.1.

Ikenishi, K. (1998). Germ plasm in *Caenorhabditis elegans*, *Drosophila* and *Xenopus*. Dev. Growth Differ. *40*, 1-10.

Ikenishi, K., and Tanaka, T.S. (1997). Involvement of the protein of *Xenopus vasa* homolog (*Xenopus vasa-like gene 1, XVLG1*) in the differentiation of primordial germ cells. Dev. Growth Differ. *39*, 625-633.

Ikenishi, K., and Tanaka, T.S. (2000). Spatio-temporal expression of *Xenopus vasa* homolog, *XVLG1*, in oocytes and embryos: The presence of *XVLG1* RNA in somatic cells as well as germline cells. Dev. Growth Differ. *42*, 95-103.

Ikenishi, K., Tanaka, T.S., and Komiya, T. (1996). Spatio-temporal distribution of the protein of *Xenopus vasa* homologue (*Xenopus vasa-like gene 1, XVLG1*) in embryos. Dev. Growth Differ. *38*, 527-535.

Illmensee.K, and Mahowald, A.P. (1974). Transplantation of posterior polar plasm in *Drosophila* - Induction of germ cells at anterior pole of egg. Proc Natl Acad Sci *71*, 1016-1020.

Irish, V., Lehmann, R., and Akam, M. (1989). The *Drosophila* posterior group gene nanos functions by repressing Hunchback activity. Nature *338*, 646-648.

Jadhav, S., Rana, M., and Subramaniam, K. (2008). Multiple maternal proteins coordinate to restrict the translation of *C. elegans nanos-2* to primordial germ cells. Development *135*, 1803-1812.

Jaglarz, M.K., and Howard, K.R. (1995). The active migration of *Drosophila* primordial germ cells. Development *121*, 3495-3503.

Jain, S., and Parker, R. (2013). The discovery and analysis of P bodies. Adv. Exp. Med. Biol. *768*, 23-43.

Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression of germline genes drives malignant brain tumor growth in *Drosophila*. Science *330*, 1824-1827.

Jankovics, F., Sinka, R., Lukacsovich, T., and Erdelyi, M. (2002). Moesin crosslinks actin and cell membrane in *Drosophila* oocytes and is required for Oscar anchoring. Curr. Biol *12*, 2060-2065.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Bioch. Sci. *36*, 19-29.

Jarmoskaite, I., and Russell, R. (2011). DEAD-box proteins as RNA helicases and chaperones. Wiley Interdiscip. Rev. -RNA *2*, 135-152.

Jarrous, N., Eder, P.S., Guerrier-Takada, C., Hoog, C., and Altman, S. (1998). Autoantigenic properties of some protein subunits of catalytically active complexes of human ribonuclease P. RNA *4*, 407-417.

Jaruzelska, J., Kotecki, M., Kusz, K., Spik, A., Firpo, M., and Pera, R.A.R. (2003). Conservation of a Pumilio-Nanos complex from *Drosophila* germ plasm to human germ cells. Dev. Genes Evol. *213,* 120-126.

Johnstone, O., Deuring, R., Bock, R., Linder, P., Fuller, M.T., and Lasko, P. (2005). Belle is a *Drosophila* DEAD-box protein required for viability and in the germ line. Dev. Biol. *277,* 92-101.

Johnstone, O., and Lasko, P. (2004). Interaction with eIF513 is essential for Vasa function during development. Development *131*, 4167-4178.

Jongens, T.A., Ackerman, L.D., Swedlow, J.R., Jan, L.Y., and Jan, Y.N. (1994). *germ cell-less* encodes a cell type specific nuclear pore-associated protein and functions early in the germ cell specification pathway of *Drosophila*. Genes Dev. *8*, 2123-2136.

Jongens, T.A., Hay, B., Jan, L.Y., and Jan, Y.N. (1992). The *germ cell-less* gene product – a posterior localized component necessary for germ-cell development in Drosophila. Cell *70*, 569-584.

Joshi, S., and Burrows, R. (1990). ATP synthase complex from bovine heart mitochondria - subunit arrangement as revealed by nearest neighbor analysis and susceptibility to trypsin. J. Biol. Chem. *265*, 14518-14525.

Juliano, C.E., Swartz, S.Z., and Wessel, G.M. (2010). A conserved germline multipotency program. Development *137*, 4113-4126.

Juliano, C.E., Voronina, E., Stack, C., Aldrich, M., Cameron, A.R., and Wessel, G.M. (2006). Germline determinants are not localized early in sea urchin development, but do accumulate in the small micromere lineage. Dev. Biol. *300*, 406-415.

Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal cyclin B mRNA by *nanos* in the *Drosophila* germline. Development *134*, 1519-1527.

Kalderon, D., and Rubin, G.M. (1988). Isolation and characterization of *Drosophila* cAMP-dependent protein kinase genes. Genes Dev. *2*, 1539-1556.

Kalifa, Y., Huang, T., Rosen, L.N., Chatterjee, S., and Gavis, E.R. (2006). Glorund, a *Drosophila* hnRNP F/H homolog, is an ovarian repressor of *nanos* translation. Dev. Cell *10*, 291-301.

Karginov, F.V., Caruthers, J.M., Hu, Y.X., McKay, D.B., and Uhlenbeck, O.C. (2005). YxiN is a modular protein combining a DEx(D)/(H) core and a specific RNA-binding domain. J. Biol. Chem. *280*, 35499-35505.

Kataoka, K., Yamaguchi, T., Orii, H., Tazaki, A., Watanabe, K., and Mochii, M. (2006). Visualization of the *Xenopus* primordial germ cells using a green fluorescent protein

controlled by cis elements of the 3' untranslated region of the *DEADSouth* gene. Mech. Dev. *123*, 746-760.

Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The *Bombyx* ovary-derived cell line endogenously expresses Piwi/Piwi-interacting RNA complexes. RNA *15*, 1258-1264.

Kelleher, E.S., Edelman, N.B., and Barbash, D.A. (2012). *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. Plos Biol. 10*:* e1001428.

Khemici, V., Poljak, L., Toesca, I., and Carpousis, A.J. (2005). Evidence *in vivo* that the DEAD-box RNA helicase RhIB facilitates the degradation of ribosome-free mRNA by RNase E. Proc Natl Acad Sci *102*, 6913-6918.

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 Genet. 11(2):e1004992.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of *in vivo* biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. at. Protoc. *4*, 506-517.

Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C., and Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: The crystal structure provides insights into the mode of unwinding. Structure *6*, 89-100.

Kim, K.H., Kang, Y.-J., Jo, J.-O., Ock, M.S., Moon, S.H., Suh, D.S., Yoon, M.S., Park, E.-S., Jeong, N., Eo, W.-K., *et al.* (2014). DDX4 (DEAD box polypeptide 4) colocalizes with cancer stem cell marker CD133 in ovarian cancers. Biochem. Biophys. Res. Commun. *447*, 315-322.

Kimble, J. & Crittenden, S. L. (2005). Germline proliferation and its control. In WormBook, The *C. elegans* Research Community ed., doi/doi: 10.1895/worm-book.1.13.1.

Kim-Ha, J., Kerr, K., and Macdonald, P.M. (1995). Translational regulation of oskar messenger RNA by Bruno, an ovarian RNA-binding protein, is essential. Cell *81*, 403-412.

Kim-Ha, J., Smith, J.L., and Macdonald, P.M. (1991). *oskar* mesenger RNA is localized to the posterior pole of the *Drosophila* oocyte. Cell *66*, 23-34.

Kimura, S., Sawatsubashi, S., Ito, S., Kouzmenko, A., Suzuki, E., Zhao, Y., Yamagata, K., Tanabe, M., Ueda, T., Fujiyama, S., *et al.* (2008). *Drosophila* arginine methyltransferase 1 (DART1) is an ecdysone receptor co-repressor. Biochem. Biophys. Res. Commun. *371*, 889-893.

King M.L. (2014) Germ cell Specification in *Xenopus*. In Kloc, M., and Kubiak, J.Z. (Eds.), *Xenopus* Development 75-100 (Hoboken, John Wiley & Sons, Inc.)

King, M.L., Messitt, T.J., and Mowry, K.L. (2005). Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. Biol. Cell *97*, 19-33.

Kirino, Y., Vourekas, A., Kim, N., Alves, F.D., Rappsilber, J., Klein, P.S., Jongens, T.A., and Mourelatos, Z. (2010a). Arginine methylation of Vasa protein is conserved across phyla. J. Biol. Chem. *285*, 8148-8154.

Kirino, Y., Vourekas, A., Sayed, N., Alves, F.D., Thomson, T., Lasko, P., Rappsilber, J., Jongens, T.A., and Mourelatos, Z. (2010b). Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. RNA *16*, 70-78.

Kitamura, E., Igarashi, J., Morohashi, A., Hida, N., Oinuma, T., Nemoto, N., Song, F., Ghosh, S., Held, W.A., Yoshida-Noro, C., *et al.* (2007). Analysis of tissue-specific differentially methylated regions (TDMs) in humans. Genomics *89*, 326-337.

Klattenhoff, C., and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Development *135*, 3-9.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., *et al.* (2009). The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Kloc, M., Bilinski, S., Chan, A.P., Allen, L.H., Zearfoss, N.R., and Etkin, L.D. (2001). RNA localization and germ cell determination in *Xenopus*. Int. Rev. Cytol., a Survey of Cell Biology, *203*, 63-91.

Klostermeier, D., and Rudolph, M.G. (2009). A novel dimerization motif in the C-terminal domain of the *Thermus thermophilus* DEAD box helicase Hera confers substantial flexibility. Nucleic Acids Res. *37*, 421-430.

Kluger, R., and Alagic, A. (2004). Chemical cross-linking and protein-protein interactions - a review with illustrative protocols. Bioorg. Chem. *32*, 451-472.

Knaut, H., Pelegri, F., Bohmann, K., Schwarz, H., and Nüsslein-Volhard, C. (2000). Zebrafish *vasa* RNA but not its protein is a component of the germ plasm and segregates asymmetrically before germline specification. J. Cell Biol. *149*, 875-888.

Knaut, H., Steinbeisser, H., Schwarz, H., and Nüsslein-Volhard, C. (2002). An evolutionary conserved region in the vasa 3' UTR targets RNA translation to the germ cells in the zebrafish. Curr. Biol. *12*, 454-466.

Knaut, H., Werz, C., Geisler, R., Nüsslein-Volhard, C., and Tubingen Screen, C. (2003). A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor. Nature *421*, 279-282.

Koc, E.C., Burkhart, W., Blackburn, K., Moyer, M.B., Schlatzer, D.M., Moseley, A., and Spremulli, L.L. (2001). The large subunit of the mammalian mitochondrial ribosome: analysis of the complement of ribosomal proteins present. Biol. Chem. *276*, 43958−43969

Komiya, T., Itoh, K., Ikenishi, K., and Furusawa, M. (1994). Isolation and characterization of a novel gene of the DEAD box protein family which is specifically expressed in germ cells of *Xenopus laevis*. Dev. Biol. *162*, 354-363.

Kopeina, G.S., Afonina, Z.A., Gromova, K.V., Shirokov, V.A., Vasiliev, V.D., and Spirin, A.S. (2008). Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. Nucleic Acids Res. *36*, 2476-2488.

Koprunner, M., Thisse, C., Thisse, B., and Raz, E. (2001). A zebrafish *nanos*-related gene is essential for the development of primordial germ cells. Genes Dev. *15*, 2877-2885.

Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G. (1997). Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. Cell *90*, 635-647.

Kotaja, N., Bhattacharyya, S.N., Jaskiewicz, L., Kimmins, S., Parvinen, M., Filipowicz, W., and Sassone-Corsi, P. (2006). The chromatoid body of male germ cells: Similarity with processing bodies and presence of Dicer and microRNA pathway components. Proc Natl Acad Sci *103*, 2647-2652.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Current Opinion in Gens. Dev. *12*, 198-209.

Krause, C.D., Yang, Z.H., Kim, Y.S., Lee, J.H., Cook, J.R., and Pestka, S. (2007). Protein arginine methyltransferases: Evolution and assessment of their pharmacological and therapeutic potential. Pharmacol. Ther. *113*, 50-87.

Kudoh, T., Tsang, M., Hukriede, N.A., Chen, X.F., Dedekian, M., Clarke, C.J., Kiang, A., Schultz, S., Epstein, J.A., Toyama, R., *et al.* (2001). A gene expression screen in zebrafish embryogenesis. Genome Res. *11*, 1979-1987.

Kugler, J.M., and Lasko, P. (2009). Localization, anchoring and translational control of *oskar*, *gurken*, *bicoid* and *nanos* mRNA during *Drosophila* oogenesis. Fly *3*, 15-28.

Kugler, J.M., Woo, J.S., Oh, B.H., and Lasko, P. (2010). Regulation of *Drosophila* Vasa *in vivo* through paralogous cullin-RING E3 ligase specificity receptors. Mol. Cell. Biol. *30*, 1769-1782.

Kunwar, P.S., Starz-Gaiano, M., Bainton, R.J., Heberlein, U., and Lehmann, R. (2003). Tre1, a G protein-coupled receptor, directs transepithelial migration of *Drosophila* germ cells. Plos Biol. *1*, 372-384.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., *et al.* (2004). *mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. Development *131*, 839-849.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Takamatsu, K., Chuma, S., Kojima-Kita, K., Shiromoto, Y., Asada, N., Toyoda, A., Fujiyama, A., *et al.* (2010). MVH in piRNA processing and gene silencing of retrotransposons. Genes Dev. *24*, 887-892.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. *22*, 1617-1635.

Kuznicki, K.A., Smith, P.A., Leung-Chiu, W.M.A., Estevez, A.O., Scott, H.C., and Bennett, K.L. (2000). Combinatorial RNA interference indicates GLH-4 can compensate for GLH-1; these two P granule components are critical for fertility in *C. elegans*. Development *127*, 2907-2916.

Lachance, P.E., Miron, M., Raught, B., Sonenberg, N., and Lasko, P. (2002). Phosphorylation of eukaryotic translation initiation factor 4E is critical for growth. Mol. and cell. Biol. *22*, 1656-1663.

Ladomery, M., Wade, E., and Sommerville, J. (1997). Xp54, the *Xenopus* homologue of human RNA helicase p54, is an integral component of stored mRNP particles in oocytes. Nucleic Acids Res. *25*, 965-973.

Lagueux, M., Perrodou, E., Levashina, E.A., Capovilla, M., and Hoffmann, J.A. (2000). Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. Proc Natl Acad Sci *97*, 11427-11432.

Lamm, G.M., Nicol, S.M., FullerPace, F.V., and Lamond, A.I. (1996). p72: A human nuclear DEAD box protein highly related to p68. Nucleic Acids Res. *24*, 3739-3747.

Lamont, L.B., Crittenden, S.L., Bernstein, D., Wickens, M., and Kimble, J. (2004). FBF-1 and FBF-2 regulate the size of the mitotic region in the *C. elegans* germline. Dev. Cell *7*, 697-707.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. Curr. Biol *14*, 2162-2167.

Lane, M.E., and Kalderon, D. (1995). Localization and functions of protein kinase A during *Drosophila* oogenesis. Mech. Dev. *49*, 191-200.

Lasko, P.F. (2012). mRNA localization and translational control in *Drosophila* oogenesis. Cold Spring Harb. Perspect. Biol. *4*, 4

Lasko, P.F. (2013). The DEAD-box helicase Vasa: Evidence for a multiplicity of functions in RNA processes and developmental biology. Biochimica et Biophysica Acta -Gene Regulatory Mechanisms *1829*, 810-816.

Lasko, P.F., and Ashburner, M. (1988). The product of the Drosophila gene *vasa* is very similar to eukaryotic initiation factor-4A. Nature *335*, 611-617.

Lasko, P.F., and Ashburner, M. (1990). Posterior localization of Vasa protein correlates with, but is not sufficient for, pole cell development. Genes Dev. *4*, 905-921.

Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J., and Hogan, B.L.M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev. *13*, 424-436.

Le Hir, H., Sauliere, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. Nat. Rev. Mol. Cell Biol. *17*, 41-54.

Leatherman, J.L., Levin, L., Boero, J., and Jongens, T.A. (2002). Germ cell-less acts to repress transcription during the establishment of the *Drosophila* germ cell lineage. Curr. Biol *12*, 1681-1685.

Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174-187.

Lee, M.H., and Schedl T. (2006). RNA-binding proteins. In WormBook, The *C. elegans* Research Community, ed. 10.1895/wormbook.1.79.1

Lee, C.S., Dias, A.P., Jedrychowski, M., Patel, A.H., Hsu, J.L., and Reed, R. (2008). Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. Nucleic Acids Res. *36*, 4708-4718.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development *129*, 1807-1817.

Lehmann, R. (2012). Germline stem cells: Origin and destiny. Cell Stem Cell *10*, 729-739.

Lehmann, R., and Nüsslein-Volhard, C. (1987). *hunchback*, a gene required for segmentation of an anterior and posterior region of the *Drosophila* embryo. Dev. Biol. *119*, 402-417.

Lehmann, R., and Nüsslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior patterning formation of the *Drosophila* embryo. Development *112*.

Leitch, H.G., and Smith, A. (2013). The mammalian germline as a pluripotency cycle. Development *140*, 2495-2501.

Lerit, D.A., and Gavis, E.R. (2011). Transport of germ plasm on astral microtubules directs germ cell development in *Drosophila*. Curr. Biol *21*, 439-448.

Lesch, B.J., and Page, D.C. (2012). Genetics of germ cell development. Nat. Rev. Genetics *13*, 781-794.

Li, Q.Y., Imataka, H., Morino, S., Rogers, G.W., Richter-Cook, N.J., Merrick, W.C., and Sonenberg, N. (1999). Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. Mol. and Cell. Biol. *19*, 7336-7346.

Li, W., Klovstad, M., and Schuepbach, T. (2014). Repression of Gurken translation by a meiotic checkpoint in *Drosophila* oogenesis is suppressed by a reduction in the dose of eIF1A. Development *141*, 3910-3921.

Liang, L., Diehljones, W., and Lasko, P. (1994). Localization of Vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. Development *120*, 1201-1211.

Lim, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. Proc Natl Acad Sci *104*, 6714-6719.

Linden, M.H., Hartmann, R.K., and Klostermeier, D. (2008). The putative RNase P motif in the DEAD box helicase Hera is dispensable for efficient interaction with RNA and helicase activity. Nucleic Acids Res. *36*, 5800-5811.

Linder, P. (2010). The dynamic life with DEAD-Box RNA helicases. In Linder, P. (ed) RNA Helicases. 32-60 (London, The Royal Society of Chemistry).

Linder, P., and Fuller-Pace, F.V. (2013). Looking back on the birth of DEAD-box RNA helicases. Biochim. Biophys. Acta. -Gene Regulatory Mechanisms *1829*, 750-755.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. *12*, 505-516.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. Nature *337*, 121-122.

Liu, C. (2013). Strategies for designing transgenic DNA constructs. Methods Mol. Biol. *1027*, 183-201.

Liu, F., Putnam, A.A., and Jankowsky, E. (2014). DEAD-box helicases form nucleotide-dependent, long-lived complexes with RNA. Biochemistry *53*, 423-433.

Liu, N.K., Dansereau, D.A., and Lasko, P. (2003). Fat facets interacts with Vasa in the *Drosophila* pole plasm and protects it from degradation. Curr. Biol. *13*, 1905-1909.

Liu, N.K., Han, H., and Lasko, P. (2009). Vasa promotes *Drosophila* germline stem cell differentiation by activating mei-P26 translation by directly interacting with a (U)-rich motif in its 3' UTR. Genes Dev. *23*, 2742-2752.

Loedige, I., Stotz, M., Qamar, S., Kramer, K., Hennig, J., Schubert, T., Loeffler, P., Laengst, G., Merkl, R., Urlaub, H., *et al.* (2014). The NHL domain of Brat is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. Genes Dev. *28*, 749-764.

Lomant, A.J., and Fairbanks, G. (1976). Chemical probes of extended biological structures – Synthesis and properties of cleavable protein cross-linking reagent dithiobis (succinimidyl-S-35 propionate). J. Mol. Biol *104*, 243-261.

Luking, A., Stahl, U., and Schmidt, U. (1998). The protein family of RNA helicases. Crit. Rev. Biochem. Mol. Biol. *33*, 259-296.

MacArthur, H., Houston, D.W., Bubunenko, M., Mosquera, L., and King, M.L. (2000). DEADSouth is a germ plasm specific DEAD-box RNA helicase in *Xenopus* related to eIF4A. Mech. of Dev. *95*, 291-295.

Macdonald, P.M., and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. Nature *324*, 537-545.

Macdonald, P.M., and Struhl, G. (1988). Cis-acting sequences responsible for anterior localization of *bicoid* messenger RNA in *Drosophila* embryos. Nature *336*, 595-598.

Mach, J.M., and Lehmann, R. (1997). An Egalitarian-BicaudalD complex is essential for oocyte specification and axis determination in *Drosophila*. Genes Dev. *11*, 423-435.

Mahowald, A.P. (1962). Fine structure of pole cells and polar granules in *Drosophila melanogaster*. J. Exp. Zool. *151*, 201-215.

Mahowald, A.P. (1968). Polar granules in *Drosophila*: II. Ultrastructural changes during early embryogenesis. J. Exp. Zool. *167*, 237-261.

Mahowald, A.P. (2001). Assembly of the *Drosophila* germ plasm. Int. Rev. Cytol. - a Survey of Cell Biology, *203*, 187-213.

Mahowald, A.P., Goralski, T.J., and Caulton, J.H. (1983). *In vitro* activation of *Drosophila* eggs. Dev. Biol. *98*, 437-445.

Mallam, A.L., Jarmoskaite, I., Tijerina, P., Del Campo, M., Seifert, S., Guo, L., Russell, R., and Lambowitz, A.M. (2011). Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. Proc Natl Acad Sci *108*, 12254-12259.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. Cell *137*, 522-535.

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. Nat. Genet. *40*, 476-483.

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*, 3723-3732.

Martin, R., Smibert, P., Yalcin, A., Tyler, D.M., Schaefer, U., Tuschl, T., and Lai, E.C. (2009). A *Drosophila pasha* mutant distinguishes the canonical microRNA and mirtron pathways. Mol. Cell. Biol. *29*, 861-870.

Martin, R., Straub, A.U., Doebele, C., and Bohnsack, M.T. (2013). DExD/H-box RNA helicases in ribosome biogenesis. RNA Biol. *10*, 4-18.

Martinez-Arroyo, A.M., Medrano, J.V., Remohi, J., and Simon, C. (2014). Germline development: lessons learned from pluripotent stem cells. Curr. Opin. Genetics Dev. *28*, 64-70.

Martinho, R.G., Kunwar, P.S., Casanova, J., and Lehmann, R. (2004). A noncoding RNA is required for the repression of RNA pol ll-dependent transcription in primordial germ cells. Curr. Biol. *14*, 159-165.

Mattson, G., Conklin, E., Desai, S., Nielander, G., Savage, M.D., and Morgensen, S. (1993). A practical approach to cross-linking. Mol. Biol. Rep. *17*, 167-183.

McFarlane, R.J., Feichtinger, J., and Larcombe, L. (2014). Cancer germline gene activation friend or foe? Cell Cycle *13*, 2151-2152.

McLaren, A. (1995). Germ cells and germ cell sex. Philos. Trans. R. Soc. Lond. B Biol. Sci. *350*, 229-233.

Medrano, J.V., Ramathal, C., Nguyen, H.N., Simon, C., and Pera, R.A.R. (2012). Divergent RNA-binding proteins, Dazl and Vasa, induce meiotic progression in human germ cells derived *in vitro*. Stem Cells *30*, 441-451.

Megosh, H.B., Cox, D.N., Campbell, C., and Lin, H. (2006). The role of Piwi and the miRNA machinery in *Drosophila* germline determination. Curr. Biol *16*, 1884-1894.

Meignin, C., and Davis, I. (2010). Transmitting the message: intracellular mRNA localization. Curr. Opin. Cell Biol. *22*, 112-119.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. *14*, 447-459.

Mello, C.C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Priess, J.R. (1996). The PIE-1 protein and germline specification in *C. elegans* embryos. Nature *382*, 710-712.

Merrick, W.C. (2004). Cap-dependent and cap-independent translation in eukaryotic systems. Gene *332*, 1-11.

Micklem, D.R., Adams, J., Grunert, S., and Johnston, D.S. (2000). Distinct roles of two conserved Staufen domains in *oskar* mRNA localization and translation. Embo J. *19*, 1366-1377.

Micklem, D.R., Dasgupta, R., Elliott, H., Gergely, F., Davidson, C., Brand, A., González-Reyes, A., and StJohnston, D. (1997). The *mago nashi* gene is required for the polarisation of the oocyte and the formation of perpendicular axes in *Drosophila*. Curr. Biol. *7*, 468-478.

Mills, R.E., Bennett, E.A., Iskow, R.C., and Devine, S.E. (2007). Which transposable elements are active in the human genome? TIG *23*, 183-191.

Mochizuki, K., Nishimiya-Fujisawa, C., and Fujisawa, T. (2001). Universal occurrence of the *vasa*-related genes among metazoans and their germline expression in Hydra. Dev. Genes Evol. *211*, 299-308.

Mochizuki, K., Sano, H., Kobayashi, S., Nishimiya-Fujisawa, C., and Fujisawa, T. (2000). Expression and evolutionary conservation of *nanos*-related genes in Hydra. Dev. Genes Evol. *210*, 591-602.

Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The Rhino-Deadlock-Cutoff complex Licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. Cell *157*, 1364-1379.

Mohr, G., Del Campo, M., Mohr, S., Yang, Q., Jia, H., Jankowsky, E., and Lambowitz, A.M. (2008). Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed *in vivo* and *in vitro*. J. Mol. Biol. *375*, 1344-1364.

Mohr, S.E., Dillon, S.T., and Boswell, R.E. (2001). The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize *oskar* mRNA during *Drosophila* oogenesis. Genes Dev. *15*, 2886-2899.

Moore, L.A., Broihier, H.T., Van Doren, M., Lunsford, L.B., and Lehmann, R. (1998). Identification of genes controlling germ cell migration and embryonic gonad formation in *Drosophila*. Development *125*, 667-678.

Morlang, S., Weglohner, W., and Franceschi, F. (1999). Hera from *Thermus thermophilus*: The first thermostable DEAD-box helicase with an RNase P protein motif. J. Mol. Biol. *294*, 795-805.

Mosquera, L., Forristall, C., Zhou, Y., and King, M.L. (1993). A messenger RNA localized to the vegetal cortex of *Xenopus* oocytes encodes a protein with a Nanos-like zinc finger domain. Development *117*, 377-386.

Mukherjee, D., and Zhao, J. (2013). The role of chemokine receptor CXCR4 in breast cancer metastasis. Am. J. Cancer Res. *3*, 46-57.

Murata, Y., and Wharton, R.P. (1995). Binding of Pumilio to maternal Hunchback messenger RNA is required for posterior patterning in *Drosophila* embryos. Cell *80*, 747-756.

Nagamori, I., Cruickshank, V.A., and Sassone-Corsi, P. (2011). Regulation of an RNA granule during spermatogenesis: acetylation of MVH in the chromatoid body of germ cells. J. Cell Sci. *124*, 4346-4355.

Nahm, M., Long, A.A., Paik, S.K., Kim, S., Bae, Y.C., Broadie, K., and Lee, S. (2010). The Cdc42-selective GAP rich regulates postsynaptic development and retrograde BMP transsynaptic signaling. J. Cell Biol. *191*, 661-675.

Nakamura, A., Amikura, R., Hanyu, K., and Kobayashi, S. (2001). Me31B silences translation of oocyte-localizing RNAs through the formation of cytoplasmic RNP complex during *Drosophila* oogenesis. Development *128*, 3233-3242.

Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S., and Lasko, P.F. (1996). Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. Science *274*, 2075-2079.

Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). *Drosophila* Cup is an eIF4E binding protein that associates with Bruno and regulates *oskar* mRNA translation in oogenesis. Dev. Cell *6*, 69-78.

Navarro, R.E., Shim, E.Y., Kohara, Y., Singson, A., and Blackwell, T.K. (2001). cgh-1, a conserved predicted RNA helicase required for gametogenesis and protection from physiological germline apoptosis in *C. elegans*. Development *128*, 3221-3232.

Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). *Drosophila* Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. Embo J. *23*, 150-159.

Neuman-Silberberg, F.S., and Schupbach, T. (1993). The *Drosophila* dorsoventral patterning gene *gurken* produces a dorsally localized RNA and encodes a TGF-alpha-like protein. Cell *75*, 165-174.

Neuman-Silberberg, F.S., and Schupbach, T. (1996). The *Drosophila* TGF-alpha-like protein Gurken: Expression and cellular localization during *Drosophila* oogenesis. Mech. Dev. *59*, 105-113.

Neumuller, R.A., Betschinger, J., Fischer, A., Bushati, N., Poernbacher, I., Mechtler, K., Cohen, S.M., and Knoblich, J.A. (2008). Mei-P26 regulates microRNAs and cell growth in the *Drosophila* ovarian stem cell lineage. Nature *454*, 241-U272.

Nielsen, K.H., Chamieh, H., Andersen, C.B.F., Fredslund, F., Hamborg, K., Le Hir, H., and Andersen, G.R. (2009). Mechanism of ATP turnover inhibition in the EJC. RNA *15*, 67-75.

Niessing, D., Dostatni, N., Jackle, H., and Rivera-Pomar, R. (1999). Sequence interval within the PEST motif of Bicoid is important for translational repression of *caudal* mRNA in the anterior region of the *Drosophila* embryo. Embo J. *18*, 1966-1973.

Nishida, K.M., Iwasaki, Y.W., Murota, Y., Nagao, A., Mannen, T., Kato, Y., Siomi, H., and Siomi, M.C. (2015). Respective functions of two distinct Siwi complexes assembled during Piwi-interacting RNA biogenesis in *Bombyx* germ cells. Cell Rep. *10*, 193-203.

Nishihara, T., Zekri, L., Braun, J.E., and Izaurralde, E. (2013). miRISC recruits decapping factors to miRNA targets to enhance their degradation. Nucleic acids res. *41*, 8692-8705.

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M.K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., *et al.* (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature *491*, 284-U157.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., *et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207-213.

Okada, M., Kleinman, I.A., and Schneide.Ha (1974). Restoration of fertility in sterilized *Drosophila* eggs by transplantation of polar cytoplasm. Dev. Biol. *37*, 43-54.

Olsen, L.C., Aasland, R., and Fjose, A. (1997). A *vasa*-like gene in zebrafish identifies putative primordial germ cells. Mech. Dev. *66*, 95-105.

Onohara, Y., Fujiwara, T., Yasukochi, T., Himeno, M., and Yokota, S. (2010). Localization of Mouse Vasa homolog protein in chromatoid body and related nuage structures of mammalian spermatogenic cells during spermatogenesis. Histochem. Cell Biol. *133*, 627-639.

Oyama, A., and Shimizu, T. (2007). Transient occurrence of *vasa*-expressing cells in nongenital segments during embryonic development in the oligochaete annelid *Tubifex tubifex*. Dev. Genes Evol. *217*, 675-690.

Page, S.L., McKim, K.S., Deneen, B., Van Hook, T.L., and Hawley, R.S. (2000). Genetic studies of mei-P26 reveal a link between the processes that control germ cell proliferation in both sexes and those that control meiotic exchange in *Drosophila*. Genetics *155*, 1757-1772.

Paksa, A., and Raz, E. (2015). Zebrafish germ cells: motility and guided migration. Curr. Opin. cell Biol. *36*, 80-85.

Palakodeti, D., Smielewska, M., Lu, Y.-C., Yeo, G.W., and Graveley, B.R. (2008). The Piwi proteins Smedwi-2 and Smedwi-3 are required for stem cell function and piRNA expression in planarians. RNA *14*, 1174-1186.

Palm, W., Sampaio, J.L., Brankatschk, M., Carvalho, M., Mahmoud, A., Shevchenko, A., and Eaton, S. (2012). Lipoproteins in *Drosophila melanogaster* - assembly, function, and influence on tissue lipid composition. PLoS Genet. *8*, e1002828.

Panáková, D., Sprong, H., Marois, E., Thiele, C., and Eaton, S. (2005). Lipoprotein particles are required for Hedgehog and Wingless signalling. Nature *435*, 58-65.

Parisi, M., and Lin, H.F. (2000). Translational repression: A duet of Nanos and Pumilio. Curr. Biol *10*, R81-R83.

Parton, R.M., Hamilton, R.S., Ball, G., Yang, L., Cullen, C.F., Lu, W., Ohkura, H., and Davis, I. (2011). A PAR-1-dependent orientation gradient of dynamic microtubules directs posterior cargo transport in the *Drosophila* oocyte. J. Cell Biol. *194*, 121-135.

Patel, S.S., Belmont, B.J., Sante, J.M., and Rexach, M.F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. Cell *129*, 83-96.

Patil, V.S., and Kai, T. (2010). Repression of retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. Curr. Biol *20*, 724-730.

Patton, E.E., and Zon, L.I. (2001). The art and design of genetic screens: Zebrafish. Nat. Rev. Genet. *2*, 956-966.

Pavlidis, P., Ramaswami, M., and Tanouye, M.A. (1994). The *Drosophila* easily shocked gene: a mutation in a phospholipid synthetic pathway causes seizure, neuronal failure, and paralysis. Cell *79*, 23-33.

Pazdernik, N., and Schedl, T. (2013). Introduction to germ cell development in *Caenorhabditis elegans*. Adv. Exp. Med. Biol. *757*, 1-16.

Pek, J.W., Anand, A., and Kai, T. (2012a). Tudor domain proteins in development. Development *139*, 2255-2266.

Pek, J.W., and Kai, T. (2011a). A Role for Vasa in regulating mitotic chromosome condensation in *Drosophila*. Curr. Biol *21*, 39-44.

Pek, J.W., and Kai, T. (2011b). DEAD-box RNA helicase Belle/DDX3 and the RNA interference pathway promote mitotic chromosome segregation. Proc Natl Acad Sci *108*, 12007-12012.

Pek, J.W., Patil, V.S., and Kai, T. (2012b). piRNA pathway and the potential processing site, the nuage, in the *Drosophila* germline. Dev. Growth Differ. *54*, 66-77.

Pelisson, A., Song, S.U., Prudhomme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila *flamenco* gene. Embo J. *13*, 4401-4411.

Pestova, T.V., Lomakin, I.B., Lee, J.H., Choi, S.K., Dever, T.E., and Hellen, C.U.T. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. Nature *403*, 332-335.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297-305.

Petit, I., Szyper-Kravitz, M., Nagler, A., Lahav, M., Peled, A., Habler, L., Ponomaryov, T., Taichman, R.S., Arenzana-Seisdedos, F., Fujii, N., *et al.* (2002). G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4. Nat. Immunol. *3*, 687-694.

Pho, D., Pennanec'h, M., and Jallon, J. (1996). Purification of adult *Drosophila melanogaster* lipophorin and its role in hydrocarbon transport. Arch. Insect Biochem. Physiol. *31*, 289-303.

Ponting, C.P. (1997). Tudor domains in proteins that interact with RNA. Trends in Biochem. Sci. *22*, 51-52.

Port, F., Chen, H.M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas9 tools for efficient germline and somatic genome engineering in *Drosophila*. Proc Natl Acad Sci *111*, E2967-E2976.

Poulin, F., and Sonenberg, N. (2003). Mechanism of translation initiation in eukaryotes. In Lapointe, J. and Brakier-Gingras, L. (Eds.), Translation mechanisms, 280–297 (Georgetown, Landes Bioscience).

Pray, L.A. (2008). Transposons: The jumping genes. Nat. Educ. *1*, 204

Printen, J.A., and Sprague, G.F. (1994). Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade. Genetics *138*, 609-619.

Rajyaguru, P., She, M., and Parker, R. (2012). Scd6 targets eIF4G to repress translation: RGG motif proteins as a class of eIF4G-binding proteins. Mol. Cell *45*, 244-254.

Ransick, A., Cameron, R.A., and Davidson, E.H. (1996). Postembryonic segregation of the germline in sea urchins in relation to indirect development. Proc Natl Acad Sci *93*, 6759-6763.

Renault, A.D. (2012). *vasa* is expressed in somatic cells of the embryonic gonad in a sex-specific manner in *Drosophila* melanogaster. Biol. Open *1*, 1043-1048.

Ribbeck, K., and Gorlich, D. (2002). The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. Embo J. *21*, 2664-2671.

Richardson, B.E., and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat. Rev. Mol. Cell Biol. *11*, 37-49.

Richter, J.D., and Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature *433*, 477-480.

Rivera-Pomar, R., Niessing, D., SchmidtOtt, U., Gehring, W.J., and Jackle, H. (1996). RNA binding and translational suppression by bicoid. Nature *379*, 746-749.

Robertson, S.E., Dockendorff, T.C., Leatherman, J.L., Faulkner, D.L., and Jongens, T.A. (1999). *germ cell-less* is required only during the establishment of the germ cell lineage of *Drosophila* and has activities which are dependent and independent of its localization to the nuclear envelope. Dev. Biol. *215*, 288-297.

Rocak, S., and Linder, P. (2004). Dead-box proteins: The driving forces behind RNA metabolism. Nat. Rev. Mol. Cell Biol. *5*, 232-241.

Rogers, G.W., Komar, A.A., and Merrick, W.C. (2002). ElF4A: The godfather of the DEAD box helicases. Prog. Nucleic Acid Res. Mol. Biol., *72*, 307-331.

Roth, S., and Lynch, J.A. (2009). Symmetry breaking during *Drosophila* oogenesis. Cold Spring Harb. Perspect. Biol. *1*, a001891.

Roth, S., Neumansilberberg, F.S., Barcelo, G., and Schupbach, T. (1995). Cornichon and the EGF receptor process are necessary for both anterior-posterior and dorsal-ventral pattern formation in *Drosophila*. Cell *81*, 967-978.

Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. Nature *467*, 1128-U1144.

Rouhana, L., Shibata, N., Nishimura, O., and Agata, K. (2010). Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. Deve. Biol. *341*, 429-443.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801-810.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-353.

Rudolph, M.G., and Klostermeier, D. (2015). When core competence is not enough: functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. Biol. Chem. *396*, 849-865.

Russell, L., and Frank, B. (1978). Ultrastructural characterization of nuage in spermatocytes of rat testis. Anat. Rec. *190*, 79-97.

Sachs, A. (2000). 10 physical and functional interactions between the mRNA cap structure and the poly (A) tail. Cold Spring Harbor Monograph Archive *39*, 447-465.

Saffman, E.E., Styhler, S., Rother, K., Li, W.H., Richard, S., and Lasko, P. (1998). Premature translation of *oskar* in oocytes lacking the RNA-binding protein Bicaudal-C. Mol. Cell. Biol. *18*, 4855-4862.

Saini, N., and Reichert, H. (2012). Neural stem cells in *Drosophila*: Molecular genetic mechanisms underlying normal neural proliferation and abnormal brain tumor formation. Stem Cells Int. *2012*, 486169.

Saito, K. (2013). The epigenetic regulation of transposable elements by Piwi-interacting RNAs in *Drosophila*. Genes Genet. Syst. *88*, 9-17.

Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293-300.

Salvetti, A., Rossi, L., Lena, A., Batistoni, R., Deri, P., Rainaldi, G., Locci, M.T., Evangelista, M., and Gremigni, V. (2005). DjPum, a homologue of *Drosophila* Pumilio, is essential to planarian stem cell maintenance. Development *132*, 1863-1874.

Santiveri, C.M., and Angeles Jimenez, M. (2010). Tryptophan residues: scarce in proteins but strong stabilizers of beta-hairpin peptides. Biopolymers *94*, 779-790.

Schaner, C.E., Deshpande, G., Schedl, P.D., and Kelly, W.G. (2003). A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. Deve. Cell *5*, 747-757.

Schuetz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: An RNA-helicase controlled by protein-protein interactions. Proc Natl Acad Sci *105*, 9564-9569.

Schupbach, T., and Wieschaus, E. (1986a). Germline autonomy of maternal-effect mutations altering the embryonic body pattern of *Drosophila*. Deve. Biol. *113*, 443-448.

Schupbach, T., and Wieschaus, E. (1986b). Maternal-effect mutations altering the anterior – posterior pattern of the *Drosophila* embryo. Roux Arch. Dev. Biol. *195*, 302-317.

Schwager, E.E., Meng, Y., and Extavour, C.G. (2015). *vasa* and *piwi* are required for mitotic integrity in early embryogenesis in the spider *Parasteatoda tepidariorum*. Deve. Biol. *402*, 276-290.

Schwer, B., and Meszaros, T. (2000). RNA helicase dynamics in pre-mRNA splicing. Embo J. *19*, 6582-6591.

Seleme, M.D., Busseau, I., Malinsky, S., Bucheton, A., and Teninges, D. (1999). High-frequency retrotransposition of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. Genetics *151*, 761-771.

Sengoku, T., Nureki, O., Nakamura, A., Satoru, K.I., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. Cell *125*, 287-300.

Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. Development *124*, 2191-2201.

Seydoux, G., Mello, C.C., Pettitt, J., Wood, W.B., Priess, J.R., and Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of *C. elegans*. Nature *382*, 713-716.

Shen, H. (2009). UAP56 a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. BMB Rep. *42*, 185-188.

Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., and Agata, K. (1999). Expression of *vasa*(*vas*)-related genes in germline cells and totipotent somatic stem cells of planarians. Dev. Biol. *206*, 73-87.

Shinomiya, A., Tanaka, M., Kobayashi, T., Nagahama, Y., and Hamaguchi, S. (2000). The *vasa*-like gene, *olvas*, identifies the migration path of primordial germ cells during embryonic body formation stage in the medaka, *Oryzias latipes*. Dev. Growth Differ. *42*, 317-326.

Silverman, I.M., Li, F., Alexander, A., Goff, L., Trapnell, C., Rinn, J.L., and Gregory, B.D. (2014). RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol. *15*, *R3*.

Singh, G., Ricci, E.P., and Moore, M.J. (2014). RIPiT-Seq: A high-throughput approach for footprinting RNA-protein complexes. Methods *65*, 320-332.

Sinsimer, K.S., Lee, J.J., Thiberge, S.Y., and Gavis, E.R. (2013). Germ plasm anchoring is a dynamic state that requires persistent trafficking. Cell Rep. *5*, 1169-1177.

Sisson, J.C., Field, C., Ventura, R., Royou, A., and Sullivan, W. (2000). Lava lamp, a novel peripheral golgi protein, is required for *Drosophila melanogaster* cellularization. J. cell Biol. *151*, 905-918.

Smibert, C.A., Wilson, J.E., Kerr, K., and Macdonald, P.M. (1996). Smaug protein represses translation of unlocalized *nanos* mRNA in the *Drosophila* embryo. Genes Dev. *10*, 2600-2609.

Smibert, P., Bejarano, F., Wang, D., Garaulet, D.L., Yang, J.-S., Martin, R., Bortolamiol-Becet, D., Robine, N., Hiesinger, P.R., and Lai, E.C. (2011). A *Drosophila* genetic screen yields allelic series of core microRNA biogenesis factors and reveals post-developmental roles for microRNAs. RNA *17*, 1997-2010.

Snee, M.J., and Macdonald, P.M. (2004). Live imaging of nuage and polar granules: evidence against a precursor-product relationship and a novel role for Oskar in stabilization of polar granule components. J. Cell Sci. *117*, 2109-2120.

Sobhanifar, S. (2003). Yeast two-hybrid assay: A fishing tale. BioTeach J. *1*, 81-87

Solana, J. (2013). Closing the circle of germline and stem cells: the Primordial Stem Cell hypothesis. Evodevo *4*, 2.

Solana, J., Lasko, P., and Romero, R. (2009). Spoltud-1 is a chromatoid body component required for planarian long-term stem cell self-renewal. Dev. Biol. *328*, 410-421.

Somma, M.P., Fasulo, B., Siriaco, G., and Cenci, G. (2003). Chromosome condensation defects in *barren* RNA-interfered *Drosophila* cells. Genetics *165*, 1607-1611.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Spike, C., Meyer, N., Racen, E., Orsborn, A., Kirchner, J., Kuznicki, K., Yee, C., Bennettt, K., and Strome, S. (2008). Genetic analysis of the *Caenorhabditis elegans* GLH family of P granule proteins. Genetics *178*, 1973-1987.

St Johnston, D. (2005). Moving messages: The intracellular localization of mRNAs. Nat. Rev. Mol. Cell Biol. *6*, 363-375.

St Johnston, D., Beuchle, D., and Nüsslein-Volhard, C. (1991). *staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. Cell *66*, 51-63.

Steimer, L., Wurm, J.P., Linden, M.H., Rudolph, M.G., Woehnert, J., and Klostermeier, D. (2013). Recognition of two distinct elements in the RNA substrate by the RNA-binding domain of the *T. thermophilus* DEAD-box helicase Hera. Nucleic Acids Res. *41*, 6259-6272.

Steinhauer, J., and Kalderon, D. (2006). Microtubule polarity and axis formation in the *Drosophila* oocyte. Dev. Dyn. *235*, 1455-1468.

Stolc, V., Katz, A., and Altman, S. (1998). Rpp2, an essential protein subunit of nuclear RNase P, is required for processing of precursor tRNAs and 35S precursor rRNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci *95*, 6716-6721.

Strome, S. (2005). Specification of the germline. In WormBook, the *C. elegans* Research Community, ed. 10.1895/wormbook.1.9.1.

Strome, S., and Updike, D. (2015). Specifying and protecting germ cell fate. Nat. Rev. Mol. Cell Biol. *16*, 406-416.

Stroschein-Stevenson, S.L., Foley, E., O'Farrell, P.H., and Johnson, A.D. (2005). Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. PLoS Biol *4*, e4.

Struebbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C., and Paro, R. (2011). Polycomb purification by *in vivo* biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. Proc Natl Acad Sci *108*, 5572-5577.

Struhl, G., Struhl, K., and Macdonald, P.M. (1989). The gradient morphogen Bicoid is a concentration-dependent transcriptional activator. Cell *57*, 1259-1273.

Styhler, S., Nakamura, A., and Lasko, P. (2002). Vasa localization requires the SPRY-domain and SOCS-box containing protein, Gustavus. Dev. Cell *3*, 865-876.

Styhler, S., Nakamura, A., Swan, A., Suter, B., and Lasko, P. (1998). Vasa is required for Gurken accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. Development *125*, 1569-1578.

Su, T.T., Campbell, S.D., and O'Farrell, P.H. (1998). The cell cycle program in germ cells of the *Drosophila* embryo. Dev. Biol. *196*, 160-170.

Subramaniam, K., and Seydoux, G. (1999). *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. Development *126*, 4861-4871.

Subramanya, H.S., Bird, L.E., Brannigan, J.A., and Wigley, D.B. (1996). Crystal structure of a DExx box DNA helicase. Nature *384*, 379-383.

Sugimoto, K., Koh, E., Sin, H.S., Maeda, Y., Narimoto, K., Izumi, K., Kobori, Y., Kitamura, E., Nagase, H., Yoshida, A., *et al.* (2009). Tissue-specific differentially methylated regions of the human *vasa* gene are potentially associated with maturation arrest phenotype in the testis. J. Hum. Genet. *54*, 450-456.

Suntharalingam, M., and Wente, S.R. (2003). Peering through the pore: Nuclear pore complex structure, assembly, and function. Dev. Cell *4*, 775-789.

Tada, H., Mochii, M., Orii, H., and Watanabe, K. (2012). Ectopic formation of primordial germ cells by transplantation of the germ plasm: Direct evidence for germ cell determinant in *Xenopus*. Dev. Biol. *371*, 86-93.

Tanaka, S., and Dan, K. (1990). Study of the lineage and cell cycle of small micromeres in embryos of the sea urchin, *Hemicentrotus pulcherrimus*. Dev. Growth Differ. *32*, 145-156.

Tanaka, S.S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., Yokoyama, M., and Noce, T. (2000). The mouse homolog of *Drosophila* Vasa is required for the development of male germ cells. Genes Dev. *14*, 841-853.

Tanaka, T., Kato, Y., Matsuda, K., Hanyu-Nakamura, K., and Nakamura, A. (2011). *Drosophila* Mon2 couples Oskar-induced endocytosis with actin remodeling for cortical anchorage of the germ plasm. Development *138*, 2523-2532.

Tarun, S.Z., and Sachs, A.B. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. Embo J. *15*, 7168-7177.

Tautz, D. (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by 2 maternal morphogenetic centers. Nature *332*, 281-284.

Teixeira, D., and Parker, R. (2007). Analysis of P-body assembly in *Saccharomyces cerevisiae*. Mol. Biol. Cell *18*, 2274-2287.

Thandapani, P., O'Connor, T.R., Bailey, T.L., and Richard, S. (2013). Defining the RGG/RG Motif. Mol. Cell *50*, 613-623.

Theurkauf, W.E., Smiley, S., Wong, M.L., and Alberts, B.M. (1992). Reorganization of the cytoskeleton during *Drosophila* oogenesis - implications for axis specification and intercellular transport. Development *115*, 923-936.

Thio, G.L., Ray, R.P., Barcelo, G., and Schupbach, T. (2000). Localization of *gurken* RNA in *Drosophila* oogenesis requires elements in the 5' and 3' regions of the transcript. Dev. Biol. *221*, 435-446.

Thomson, T., and Lasko, P. (2004). *Drosophila* Tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning. Genesis *40*, 164-170.

Timinszky, G., Bortfeld, M., and Ladurner, A.G. (2008). Repression of RNA polymerase II transcription by a *Drosophila* oligopeptide. Plos One *3*, e2506.

Tiveron, M.C., and Cremer, H. (2008). CXCL12/CXCR4 signalling in neuronal cell migration. Curr. Opin. Neurobiol. *18*, 237-244.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during *Drosophila* embryogenesis. Genome biol. *8*, R145.

Tomancak, P., Guichet, A., Zavorszky, P., and Ephrussi, A. (1998). Oocyte polarity depends on regulation of *gurken* by Vasa. Development *125*.

Torres, I.L., Lopez-Schier, H., and St Johnston, D. (2003). A notch/delta-dependent relay mechanism establishes anterior-posterior polarity in *Drosophila*. Dev. Cell *5*, 547-558.

Tran, E.J., Zhou, Y., Corbett, A.H., and Wente, S.R. (2007). The DEAD-box protein dbp5 controls mRNA export by triggering specific RNA: Protein remodeling events. Mol. Cell *28*, 850-859.

Tremblay, K.D., Dunn, N.R., and Robertson, E.J. (2001). Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development *128*, 3609-3621.

Tsang, T.E., Khoo, P.L., Jamieson, R.V., Zhou, S.X., Ang, S.L., Behringer, R., and Tam, P.P.L. (2001). The allocation and differentiation of mouse primordial germ cells. Int. J. Dev. Biol. *45*, 549-555.

Tseng, S.S.L., Weaver, P.L., Liu, Y., Hitomi, M., Tartakoff, A.M., and Chang, T.H. (1998). Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. Embo J. *17*, 2651-2662.

Ulvila, J., Parikka, M., Kleino, A., Sormunen, R., Ezekowitz, R.A., Kocks, C., and Ramet, M. (2006). Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in *Drosophila* S2 cells. J. Biol. Chem. *281*, 14370-14375.

Updike, D., and Strome, S. (2010). P granule assembly and function in *Caenorhabditis elegans* germ cells. J. Androl. *31*, 53-60.

Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear pore complex environment in the *C. elegans* germline. J. Cell Biol. *192*, 939-948.

Updike, D.L., Knutson, A.K.a., Egelhofer, T.A., Campbell, A.C., and Strome, S. (2014). Germ granule components prevent somatic development in the *C. elegans* Germline. Curr. Biol *24*, 970-975.

Vagin, V.V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G.J., and Aravin, A.A. (2009). Proteomic analysis of murine

Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes Dev. *23*, 1749-1762.

Van Criekinge, W., and Beyaert, R. (1999). Yeast Two-Hybrid: State of the Art. Biol. Proced. Online *2*, 1-38.

Van Doren, M., Williamson, A.L., and Lehmann, R. (1998). Regulation of zygotic gene expression in *Drosophila* primordial germ cells. Curr. Biol *8*, 243-246.

van Eeden, F.J.M., Palacios, I.M., Petronczki, M., Weston, M.J.D., and St Johnston, D. (2001). Barentsz is essential for the posterior localization of *oskar* mRNA and colocalizes with it to the posterior pole. J. Cell Biol. *154*, 511-523.

Vanzo, N.F., and Ephrussi, A. (2002). Oskar anchoring restricts pole plasm formation to the posterior of the *Drosophila* oocyte. Development *129*, 3705-3714.

Vasilescu, J., Guo, X., and Kast, J. (2004). Identification of protein-protein interactions using *in vivo* cross-linking and mass spectrometry. Proteomics *4*, 3845-3854.

Vogel, J., Drapkin, B., Oomen, J., Beach, D., Bloom, K., and Snyder, M. (2001). Phosphorylation of γ-tubulin regulates microtubule organization in budding yeast. Dev. cell *1*, 621-631.

Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., and Wessel, G. (2008). Vasa protein expression is restricted to the small micromeres of the sea urchin, but is inducible in other lineages early in development. Dev. Biol. *314*, 276-286.

Wagner, D.E., Ho, J.J., and Reddien, P.W. (2012). Genetic regulators of a pluripotent adult stem cell system in planarians identified by RNAi and clonal analysis. Cell Stem Cell *10*, 299-311.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha-subunits and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. Embo J. *1*, 945-951.

Wang, C., Dickinson, L.K., and Lehmann, R. (1994). Genetics of *nanos* localization in *Drosophila*. Dev. Dynam. *199*, 103-115.

Wang, D., Kennedy, S., Conte, D., Kim, J.K., Gabel, H.W., Kamath, R.S., Mello, C.C., and Ruvkun, G. (2005). Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. Nature *436*, 593-597.

Wang, J.T., and Seydoux, G. (2013). Germ cell specification. Adv. Exp. Med. Biol. *757*, 17-39

Wang, S.C., Hsu, H.J., Lin, G.W., Wang, T.F., Chang, C.C., and Lin, M.D. (2015). Germ plasm localisation of the HELICc of Vasa in *Drosophila*: analysis of domain sufficiency and amino acids critical for localisation. Sci. Rep. 5, doi:10.1038.

Wang, S.X., and Hazelrigg, T. (1994). Implications for *bcd* messenger RNA localization from spatial distribution of Exu protein in *Drosophila* oogenesis. Nature *369*, 400-403.

Wang, Z., and Lin, H.F. (2004). Nanos maintains germline stem cell self-renewal by preventing differentiation. Science *303*, 2016-2019.

Watson, F.L., Püttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., and Schmucker, D. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science *309*, 1874-1878.

Webster, A., Li, S., Hur, J.K., Wachsmuth, M., Bois, J.S., Perkins, E.M., Patel, D.J., and Aravin, A.A. (2015). Aub and Ago3 are recruited to nuage through two mechanisms to form a ping-pong complex assembled by Krimper. Mol. Cell *59*, 564-575.

Webster, P.J., Liang, L., Berg, C.A., Lasko, P., and Macdonald, P.M. (1997). Translational repressor Bruno plays multiple roles in development and is widely conserved. Genes Dev. *11*, 2510-2521.

Weick, E.M., and Miska, E.A. (2014). piRNAs: from biogenesis to function. Development *141*, 3458-3471.

Weidinger, G., Wolke, U., Koprunner, M., Klinger, M., and Raz, E. (1999). Identification of tissues and patterning events required for distinct steps in early migration of zebrafish primordial germ cells. Development *126*, 5295-5307.

Weidinger, G., Wolke, U., Koprunner, M., Thisse, C., Thisse, B., and Raz, E. (2002). Regulation of zebrafish primordial germ cell migration by attraction towards an intermediate target. Development *129*, 25-36.

Weil, T.T. (2014). mRNA localization in the *Drosophila* germline. RNA Biol. *11*, 1010-1018.

Weil, T.T., Xanthakis, D., Parton, R., Dobbie, I., Rabouille, C., Gavis, E.R., and Davis, I. (2010). Distinguishing direct from indirect roles for *bicoid* mRNA localization factors. Development *137*, 169-176.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Molecular Cell *2*, 135-140.

Weston, A., and Sommerville, J. (2006). Xp54 and related (DDX6-like) RNA helicases: roles in messenger RNP assembly, translation regulation and RNA degradation. Nucleic Acids Res. *34*, 3082-3094.

White, Y.A.R., Woods, D.C., Takai, Y., Ishihara, O., Seki, H., and Tilly, J.L. (2012). Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. Nat. Med. *18*, 413-U176.

Whitfield, M.L., George, L.K., Grant, G.D., and Perou, C.M. (2006). Common markers of proliferation. Nat. Rev. Cancer. *6*, 99-106.

Whitington, P.M., and Dixon, K.E. (1975). Quantitative studies of germ plasm and germ cells during early embryogenesis of *Xenopus laevis*. J. Embryol. Exp. Morphol. *33*, 57-74.

Wilhelm, J.E., Mansfield, J., Hom-Booher, N., Wang, S.X., Turck, C.W., Hazelrigg, T., and Vale, R.D. (2000). Isolation of a ribonucleoprotein complex involved in mRNA localization in *Drosophila* oocytes. J. Cell Biol. *148*, 427-439.

Wilhelm, J.E., and Smibert, C.A. (2005). Mechanisms of translational regulation in *Drosophila*. Biol. Cell *97*, 235-252.

Wilson, J.E., Connell, J.E., and Macdonald, P.M. (1996). Aubergine enhances *oskar* translation in the *Drosophila* ovary. Development *122*, 1631-1639.

Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell *118*, 619-633.

Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. (1997). Nanos and Pumilio establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of *hunchback* mRNA. Development *124*, 3015-3023.

Wylie, C.C., and Heasman, J. (1976). Formation of gonadal ridge in *Xenopus laevis* .1. Light and transmission microscope study. J. Embryol. Exp. Morphol. *35*, 125-138.

Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Coute, Y., Conn, S., Kadlec, J., Sachidanandam, R., *et al.* (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell *157*, 1698-1711.

Yajima, M., Gustafson, E.A., Song, J.L., and Wessel, G.M. (2014). Piwi regulates Vasa accumulation during embryogenesis in the sea urchin. Dev. Dynam. *243*, 451-458.

Yajima, M., and Wessel, G.M. (2011a). Small micromeres contribute to the germline in the sea urchin. Development *138*, 237-243.

Yajima, M., and Wessel, G.M. (2011b). The DEAD-box RNA helicase Vasa functions in embryonic mitotic progression in the sea urchin. Development *138*, 2217-2222.

Yajima, M., and Wessel, G.M. (2015). Essential elements for translation: the germline factor Vasa functions broadly in somatic cells. Development *142*, 1960-1970.

Yamaguchi, T., Kataoka, K., Watanabe, K., and Orii, H. (2014). Restriction of the *Xenopus DEADSouth* mRNA to the primordial germ cells is ensured by multiple mechanisms. Mech. Dev. *131*, 15-23.

Yamaguchi, T., Taguchi, A., Watanabe, K., and Orii, H. (2013). DEADSouth protein localizes to germ plasm and is required for the development of primordial germ cells in *Xenopus laevis*. Biol. Open *2*, 191-199.

Yan, R., Thomas, S.E., Tsai, J.-H., Yamada, Y., and McKee, B.D. (2010). SOLO: a meiotic protein required for centromere cohesion, coorientation, and SMC1 localization in *Drosophila melanogaster*. J. Cell Biol. *188*, 335-349.

Yang, J.X., Zhang, N., Wang, H.W., Gao, P., Yang, Q.P., and Wen, Q.P. (2015a). CXCR4 receptor overexpression in mesenchymal stem cells facilitates treatment of acute lung injury in rats. J. Biol. Chem. *290*, 1994-2006.

Yang, N., Yu, Z., Hu, M., Wang, M., Lehmann, R., and Xu, R.M. (2015b). Structure of *Drosophila* Oskar reveals a novel RNA binding protein. Proc Natl Acad Sci *112*, 11541-11546.

Yano, T., Lopez de Quinto, S., Matsui, Y., Shevchenko, A., and Ephrussi, A. (2004). Hrp48, a *Drosophila* hnRNPA/B homolog, binds and regulates translation of *oskar* mRNA. Dev. Cell *6*, 637-648.

Ying, Y., Liu, X.M., Marble, A., Lawson, K.A., and Zhao, G.Q. (2000). Requirement of Bmp8b for the generation of primordial germ cells in the mouse. Mol. Endocrinol. *14*, 1053-1063.

Ying, Y., and Zhao, G.Q. (2001). Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. Dev. Biol. *232*, 484-492.

Yoon, C., Kawakami, K., and Hopkins, N. (1997). Zebrafish *vasa* homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. Development *124*, 3157-3165.

Yoshida, S., Müller, H.A.J., Wodarz, A., and Ephrussi, A. (2004). PKA-R1 spatially restricts Oskar expression for *Drosophila* embryonic patterning. Development *131*, 1401-1410.

Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows *nanos* mRNA translation in *Drosophila* embryos by preventing its deadenylation by Smaug/CCR4. Development *133*, 4573-4583.

Zalokar, M. (1976). Autoradiographic study of protein and RNA formation during early development of *Drosophila* eggs. Dev. Biol. *49*, 425-437.

Zamudio, N., and Bourc'his, D. (2010). Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity *105*, 92-104.

Zapata, J.M., Martinez, M.A., and Sierra, J.M. (1994). Purification and characterization of eukaryotic polypeptide chain factor 4F from *Drosophila melanogaster* embryos. J. Biol. Chem. *269*, 18047-18052.

Zeeman, A.M., Stoop, H., Boter, M., Gillis, A.J.M., Castrillon, D.H., Oosterhuis, J.W., and Looijenga, L.H.J. (2002). Vasa is a specific marker for both normal and malignant human germ cells. Lab. Invest. *82*, 159-166.

Zhang, C., and Darnell, R.B. (2011). Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotech. *29*, 607-U686.

Zhang, F., Barboric, M., Blackwell, T.K., and Peterlin, B.M. (2003). A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. Genes Dev. *17*, 748-758.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., *et al.* (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871-884.

Zhang, H., Panula, S., Petropoulos, S., Edsgard, D., Busayavalasa, K., Liu, L., Li, X., Risal, S., Shen, Y., Shao, J., *et al.* (2015). Adult human and mouse ovaries lack DDX4-expressing functional oogonial stem cells. Nat. Med. *21*, 1116-1118.

Zhao, R., Shen, J.P., Green, M.R., MacMorris, M., and Blumenthal, T. (2004). Crystal structure of UAP56, a DExD/H-Box protein involved in pre-mRNA splicing and mRNA export. Structure *12*, 1373-1381.

Zhao, T., Graham, O.S., Raposo, A., and St Johnston, D. (2012). Growing microtubules push the oocyte nucleus to polarize the *Drosophila* dorsal-ventral axis. Science *336*, 999-1003.

Zhou, R., Hotta, I., Denli, A.M., Hong, P., Perrimon, N., and Hannon, G.J. (2008). Comparative analysis of argonaute-dependent small RNA pathways in *Drosophila*. Mol. Cell *32*, 592-599.

Zimyanin, V.L., Belaya, K., Pecreaux, J., Gilchrist, M.J., Clark, A., Davis, I., and St Johnston, D. (2008). *In vivo* imaging of *oskar* mRNA transport reveals the mechanism of posterior localization. Cell *134*, 843-853.