Dissertation

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Matteo Bordi, M.Sc.

born in Roma, Italy

Oral-examination: November 21th, 2017

Investigating the RNA-binding activity of the *Drosophila melanogaster* germline inducer Oskar

Referees: Prof. Georg Stoecklin

Dr. Alexander Aulehla

Acknowledgement

I want to thank Anne Ephrussi for giving me the opportunity to work in her lab on a very interesting project and for all the support that she provided during the four years of my PhD. I want to thank all members of Anne's lab for their help and the constructive discussions and advices. In particular, I would like to thank: Mandy Jeske for helping me when I arrived at EMBL and for all the help provided during my PhD work, and Ales Obrdlik for the scientific and technical help, discussions and all the cigarettes that we smoked together.

I'm grateful to my TAC members: Dr. Alexander Aulehla, Dr. Orsolya Bararbas and Dr. Prof. Georg Stoecklin for their constant help during my PhD time at EMBL.

I want to thank Dr. Vladimir Benes and all the people at Gene Core for the technical and scientific advice they provided at several points in my work.

I'm grateful to EMBL for funding my PhD thus giving me the opportunity to work at one of the best European research institutes. I want to thank the persons in the EMBL Graduate Office for the help provided in all the steps of my PhD.

Finally I want to thank my family and all my friends for all the support they provided during the past four years of my life.

Table of Contents

List of Abbreviations

3’UTR	3’ untranslated region
4EHP	4E homology protein
4-SU	4-thiouridine
6-SG	6-thioguanosine
Act42A	Actin42A
Ago2	Argonaute 2
aub	aubergine
AUF1	AU-rich element-binding factor 1
bcd	bicoid
bicD	bicaudal-D
BLE	bicoid localization element
bp	base pair
brat	brain tumor
BRE	Bruno recognition element
bru	bruno
btz	barentsz
cad	caudal
CB	coupling buffer
cdc25	cell division cycle 25
cdk1	cyclin-dependent kinase 1
CLIP	crosslinking and immunoprecipitation
CPEB	cytoplasmic polyadenylation element binding protein
CTD	C-terminal domain
cycB	cyclin B
DBE	dephosphorylation buffer for equilibration
DEPC	diethyl pyrocarbonate
DMSO	dimethyl sulfoxide
DRB	dephosphorylation reaction buffer
DTT	dithiothreitol
eCLIP	enhanced CLIP
EGFR	epidermal growth factor receptor
egl	egalitarian
EJC	exon junction complex
exu	exuperantia
gcl	germ cell-less
glo	glorund
hb	hunchback
HITS-CLIP	high throughput sequencing of RNA isolated by crosslinking and immunoprecipitation
hnRNP	heterogeneous ribonucleoprotein
Hsp83	Heat-shock protein 83
IBE	IMP-binding element
iCLIP	individual nucleotide resolution crosslinking and immunoprecipitation
IGF2BP1-3	insulin-like growth factor 2 mRNA-binding protein 1-3

IMP	insulin growth factor II mRNA-binding protein
IP	immunoprecipitation
LOTUS	Limkain, Oskar and Tudor containing proteins 5 and 7
LB1	lysis buffer 1
LB2	lysis buffer 2
LBE	ligation buffer for equilibration
LBU	lysis buffer urea
mago	mago nashi
MTOC	microtubule organizing center
MPF	maturation promoting factor
mRNA	messenger RNA
Na deoxycholate	sodium deoxycholate
NGS	next generation sequencing
nos	nanos
OES	oocyte entry signal
orb	oo18 RNA-binding protein
osk	oskar
otd	orthodenticle
PABPC	cytoplasmic poly(A)-binding protein
PAP	poly(A) polymerase
PAR-CLIP	photoactivable ribonucleoside-enhanced crosslinking and immunoprecipitation
PCR	polymerase chain reaction
PEG	polyethylene glycol
pgc	polar granule component
PGC	primordial germ cell
PKB	proteinase K buffer
PNK	polynucleotide kinase
PTB	polypyrimidine tract binding protein
P-TEFb	positive transcription elongation factor b
pum	pumilio
PUM2	Pumilio homologue 2
QKI	Quaking
qPCR	quantitative real-time PCR
RBD	RNA-binding domain
RBP	RNA-binding protein
RIP	RNA immunoprecipitation
RIP-ChiP	RNA immunoprecipitation coupled to microarray
RNAi	RNA interference
RNP	ribonucleoprotein particle
RRM	RNA recognition motif
rRNA	ribosomal RNA
rump	rumpelstiltskin
SAFB1	Scaffold Attachment Factor B1
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
Ser2	Serine 2
Ser5	Serine 5
smg	smaug
SOLE	spliced oskar localization element

SRE	Smaug recognition element
stau	staufen
TCE	translational control element
TDP-43	TAR DNA-binding protein 43
TIA1	T-cell intracellular antigen 1
TIAL1	TIA-like 1
tm1	tropomyosin 1
tud	tudor
vas	vasa
vls	valois
WB1	washing buffer 1
WB2	washing buffer 2
WB3	washing buffer 3
WB4	washing buffer 4
WBU1	washing buffer urea 1
WBU2	washing buffer urea 2

List of Figures and Tables

Summary

In many animals, a specialized cytoplasm forms within the oocyte that harbors all the molecular factors required for germ cell fate specification and is defined as the germ plasm. In *Drosophila melanogaster*, the germ plasm (or pole plasm), is assembled at the posterior pole of the oocyte in a stepwise process triggered by Oskar protein. *oskar* mRNA transcribed in the nuclei of nurse cells is actively transported into, and to the posterior pole of, the oocyte. At the posterior pole, *oskar* mRNA is translated into Oskar protein, which recruits the other pole plasm components required for germline specification and posterior patterning in the embryo.

Recently, it was shown that Oskar binds polyadenylated mRNAs *in vivo* and that the C-terminal domain of the protein binds RNA *in vitro*. In that study, by using UV crosslinking and immunoprecipitation experiments, I showed that Oskar associates *in vivo* with three mRNAs involved in posterior patterning and germ cell fate specification: *nanos*, *polar granule component* and *germ cell-less*. In order to identify Oskar's binding site(s) on its target transcripts I applied the iCLIP method to early *Drosophila* embryos. To this end, I analyzed the efficiency of three key steps of the iCLIP protocol: RNAse digestion, 3' end dephosphorylation, and adapter ligation. I found that, while the 3' end dephosphorylation was generally efficient, the ligation of an adapter to RNA was a limiting step in the protocol. By performing a series of optimization experiments, I established new reaction conditions for adapter ligation that increased the efficiency of the reaction significantly.

Even after optimization of the iCLIP protocol, the data produced in the Oskar iCLIP were inconclusive, due at least in part to the low affinity of Oskar for RNA. Hence, positional information regarding Oskar's interaction with specific RNAs *in vivo* is still lacking. Such information would enable validation and further characterization of the RNA-binding activity of Oskar, providing important new insight into the molecular mechanisms underlying Oskar's unique pole plasm inducing activity.

The optimized iCLIP protocol I developed should be applicable to other RNA-binding proteins in *Drosophila* and in other model organisms.

Zusammenfassung

In der Oozyte zahlreicher Tiere wird ein spezialisiertes Zytoplasma gebildet, das alle für die Spezifikation der Keimzellen benötigten molekularen Faktoren beherbergt und als Keimplasma definiert ist. In *Drosophila melanogaster* wird das Keimplasma (oder Polplasma) am posterioren Pol der Oozyte in einem durch Oskar-Protein ausgelösten schrittweisen Prozess assembliert. *Oskar* mRNA, die in den Kernen der sogenannten Nurse Zellen (Nährzellen) transkribiert wird, wird aktiv zu dem posterioren Pol der Oozyte transportiert. Am posterioren Pol wird *oskar* mRNA in Oskar-Protein translatiert, das die für die Spezifikation der Keimzellen und frühe embryonale Musterbildung erforderlichen Polplasmakomponenten rekrutiert.

Vor kurzem wurde gezeigt, dass Oskar polyadenylierte mRNAs *in vivo* und seine C-terminale Domäne RNA *in vitro* binden kann. In Rahmen dieser Arbeit wurde mit Hilfe der UV-Crosslinking und Immunpräzipitationsexperimenten demonstriert, dass Oskar *in vivo* mit drei mRNAs, die an der embryonale Musterbildung und der Spezifikation der Keimzellen beteiligt sind, assoziiert: *nanos*, *polar granule component* and *germ cell-less*. Um Oskars Bindungsstelle(n) auf seinen Zieltranskripten zu identifizieren, wurde in dieser Studie die iCLIP-Methode auf frühe *Drosophila*-Embryonen angewendet. Dafür wurde die Effizienz von drei Schlüsselschritten des iCLIP-Protokolls analysiert: RNAse-Verdauung, 3'-End-Dephosphorylierung und Adapter-Ligation. Es konnte gezeigt werden, dass, während die 3'-End-Dephosphorylierung im Allgemeinen effizient war, die Adapter-Ligation ein limitierender Schritt in dem Protokoll war. Durch eine Reihe von Optimierungsexperimenten wurden neue Reaktionsbedingungen für die Adapter-Ligation etabliert, die die Effizienz der Reaktion signifikant erhöhten.

Trotz der Optimierung des iCLIP-Protokolls war die RNA-Sequenzierung nach Oskar iCLIP ergebnislos, was, zumindest teilweise, auf die geringen Affinität von Oskar für RNA zurückzuführen ist. Demzufolge bleibt die Positionsinformation über Oskars *in vivo* Interaktion mit spezifischen RNAs unbekannt. Solche Informationen würden die Validierung und weitere Charakterisierung der RNA-bindenden Aktivität von Oskar ermöglichen und einen wichtigen neuen Einblick in die molekularen Mechanismen liefern, die der einzigartigen polplasmainduzierenden Aktivität von Oskar zugrunde liegen.

Das in dieser Arbeit optimierte iCLIP-Protokoll sollte auf andere RNA-bindende Proteine in *Drosophila* und in anderen Modellorganismen anwendbar sein.

1. Introduction

1.1. *Drosophila melanogaster* oogenesis and early embryogenesis

An adult *Drosophila melanogaster* female possesses one pair of ovaries, in which the process of egg production, oogenesis, occurs. Each ovary is composed of approximately 16 ovarioles, each of which can be considered as an egg production line. Indeed, a single ovariole consists of a sequence of progressively older egg chambers separated by a few somatic cells. At the anterior end of each ovariole is a structure called the germarium, in which germline stem cells are located and where the process of oogenesis is initiated. From the germarium, the newly formed egg chambers progressively move towards the distal (posterior) end of the ovariole as they develop into eggs competent for fertilization. The entire process, based on morphological criteria, has been divided into 14 stages, from the newly formed egg chamber (stage 1) to the mature egg (stage 14) (Fig. 1) [1] [2].

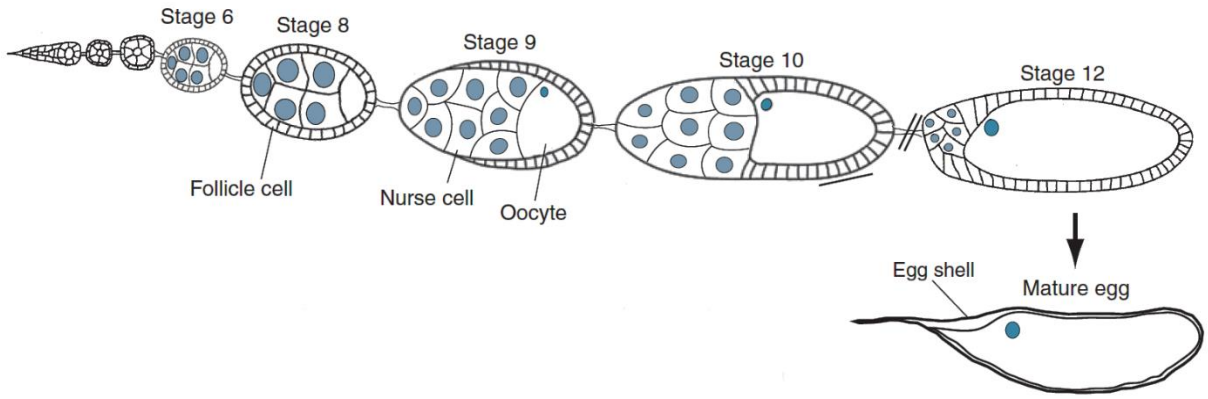

Figure 1: *Drosophila* oogenesis. A single *Drosophila* ovariole contains a series of egg chambers at different developmental stages. In the germarium (on the left), new egg chambers are formed. Asymmetric divisions of germline stem cells in the germarium produce a new stem cell and a cystoblast. The cystoblast divides four times with incomplete cytokinesis thus producing a cyst composed of 16 germline cells interconnected by cytoplasmic bridges. This group of cells is surrounded by somatic follicular cells, together constituting an egg chamber. Inside the egg chamber, one cell is selected to become the oocyte, while the remaining 15 cells become nurse cells and provide metabolic support to the growing oocyte. The development of an egg chamber has been divided into 14 morphological stages and the newly formed egg chamber arising from the germarium is defined as stage 1. While moving toward the end of the ovariole, the egg chamber matures. The oocyte increases its volume thanks to the material produced by the nurse cells. At stage 11, the nurse cells dump their cytoplasm into the

oocyte and undergo apoptosis. Toward the end of oogenesis follicular cells surrounding the oocyte produce the egg shell and other specialized structures. Stage 14 egg chambers are mature eggs ready to be fertilized.

Adapted from [3].

The germarium, the site of egg chamber production, contains two cell types: germline stem cells and somatic stem cells. At the beginning of oogenesis, a germline stem cell undergoes an asymmetric division, producing a new germline stem cell and its sibling daughter cell, the cystoblast. Subsequently, the cystoblast divides four times with incomplete cytokinesis, thus producing a cyst of 16 cells interconnected by cytoplasmic bridges called ring canals. As the cyst travels through the germarium, one of the 16 cells is specified as the oocyte, and the cyst is surrounded by a single layer of somatic follicle cells. At this point the fully formed egg chamber exits the germarium and continues to develop while moving to the posterior end of the ovariole. [1] [2].

The process of oocyte selection restricts the oocyte fate to a single of the 16 cells resulting from the four rounds of division of the cystoblast. At the beginning of the process, several cells display oocyte features, which become progressively restricted to one cell, the future oocyte; the remaining 15 cells develop into so-called nurse cells [4]. A key role in oocyte selection is played by the fusome, a vesicular organelle associated with microtubules that links the 16 cells in the cyst via the ring canals. During the cystoblast divisions, the fusome is unequally distributed among the cells, with one of the cells retaining more fusome material than the others. This particular cell will develop as the future oocyte [5]. The fusome is required for the organization of a polarized microtubule network that allows the transport of molecular components toward the single microtubule organizing center (MTOC) of the cyst, located in the future oocyte [6]. The microtubule-dependent transport of oocyte factors requires the molecular motor Dynein. Mutations in *egalitarian* (*egl*) and *bicaudal-D* (*bicD*), genes that encode an RNA-binding protein (RBP) and a Dynein cargo adapter, respectively, cause a failure to select the future oocyte [7] [8]. Shortly after its fate has been specified the oocyte

enters meiosis, but the process is arrested at Prophase I until the mature oocyte reaches the tip of the ovary [4] [9].

The nurse cells provide metabolic support to the growing oocyte. They are highly active in both transcription and translation, producing the materials required by the growing oocyte, which is transcriptionally quiescent during oogenesis, as well as by the young embryo whose first steps of development after fertilization, occur before activation of the zygotic genome. These “maternally provided components” include messenger RNAs (mRNA), proteins and ribosomes, as well as organelles such as mitochondria, all of which are transported along microtubules, through the ring canals and into the oocyte [4] [9]. Once in the oocyte, some of these mRNAs and proteins are actively transported to or otherwise accumulate at specific subcellular locations. The process of asymmetric protein and RNA localization is key for specification of the anterior-posterior and dorsal-ventral axes of the future embryo [4] [10].

The amount of material that is transported from nurse cells into the growing oocyte is also responsible for the increase in volume of the oocyte itself. During the first stages of oogenesis (1 to 10) the slow, active and regulated transport of mainly proteins and mRNAs from the nurse cells, together with uptake of yolk proteins produced by the follicle cells, cause a progressive increase in the size of the oocyte. At the beginning of stage 10, roughly half of the egg chamber is occupied by the oocyte. Toward the end of oogenesis (stage 11), the nurse cells dump their cytoplasmic content into the oocyte and undergo apoptosis. At the end of these processes the oocyte reached its final volume (Fig. 1) [3].

Follicle cells that envelop the egg chamber also play an important role during oogenesis. Interactions that occur between the oocyte and follicle cells are critical for body axis specification. As the mature egg chamber leaves the germarium, the oocyte has been already polarized by the action of Par-1 together with BicD, Egl and the microtubule minus end directed motor protein Dynein [11]. Outside the germarium two signaling events that involve Gurken are required for both anterior-posterior and dorsal-ventral axis specification. Gurken is an epidermal growth factor receptor (EGFR) ligand that is received by the EGFR expressing

follicle cells. During stage 6/7 the first Gurken signal induces the follicle cells at the posterior of the egg chamber to differentiate into posterior follicle cells. An unidentified signal is sent back to the oocyte by the posterior follicle cells and as a consequence the microtubule cytoskeleton of the oocyte is repolarized [12] [13]. This process involves the disassembly of the MTOC at the posterior of the oocyte and formation of new arrays of microtubules that now nucleate from the anterior and lateral cortex of the oocyte, as well as from the nuclear membrane [14] [15]. At stage 7, the nucleus migrates to the anterior lateral cortex of the oocyte. Once it reaches this position at stage 9, a second round of Gurken signaling induces dorsal-ventral axis specification [12] [13] [16]. Finally, toward the end of oogenesis the follicle cells produce the chorion proteins forming the eggshell, as well as other specialized eggshell structures, such as the dorsal appendages, the operculum and the micropyle [17].

As previously mentioned, the meiosis begins shortly after the oocyte has been specified, then arrests in Prophase I. This arrest is relieved upon oocyte maturation, a process that takes place during stage 12/13 of oogenesis [9]. The process of oocyte maturation has been extensively studied in some animal species, but is much less well understood in *Drosophila*. In amphibians and other animals, high levels of Cyclin-dependent kinase 1 (Cdk1) and Cyclin B (CycB) activity are required for oocyte maturation. In addition, the phosphatase Cell division cycle 25 (Cdc25) is another important factor required during this process. By removing inhibitory phosphate groups from Cdk1, Cdc25 stimulates the maturation process. Cdk1 in complex with CycB forms the maturation promoting factor (MPF) that triggers nuclear envelope breakdown, chromosome condensation and spindle assembly [18]. Mutations in Cdk1 delay the oocyte maturation process in *Drosophila* [19]. Twine, the *Drosophila* homolog of Cdc25, is also required for proper oocyte maturation [20]. Another factor that plays an important role in this process is Polo kinase, which acts as an activator of Twine [21] [22]. At the end of the maturation process meiosis is arrested once again in Metaphase I, until egg activation occurs [9].

The process of egg activation in *Drosophila* occurs during the passage of the mature oocyte (stage 14) from the ovary into the oviduct. During egg activation, the meiotic process is completed and the egg becomes ready for fertilization. The exact molecular mechanisms underlying the egg activation process are still not completely understood but specific factors act as initial triggers: the mechanical stress to which the oocyte is subjected in the oviduct and fluid uptake that lead to an increase of the intracellular levels of calcium [23] [24].

After completion of egg activation, the fertilization process occurs in the uterus, the egg is deposited in the environment and embryogenesis starts. The zygotic nucleus undergoes a series of 13 rapid and synchronous mitotic divisions without cytokinesis. This process happens in 2-2.5 hours and can be so rapid because it relies on maternally deposited gene products thus permitting to skip S-phases in the mitotic divisions [25]. At the end of this process the early embryo consists of a syncytium containing approximately 6000 nuclei that migrate to the periphery of the embryo, where cellularization occurs. This is the first developmental process that requires transcription of zygotic genes and marks the beginning of gastrulation [9] [26].

1.2. Anterior-posterior axis specification

Anterior-posterior or head to tail axis specification in *Drosophila* is one of the best understood examples of pattern formation in animals. This process is achieved through the asymmetric localization within the oocyte of three key regulators: *bicoid* (*bcd*) mRNA at the anterior pole, *oskar* (*osk*) and *nanos* (*nos*) mRNAs at the posterior pole (Fig. 2). The accumulation of these mRNAs at specific subcellular locations is required to establish proper morphogen gradients along the anterior-posterior axis thus leading to proper patterning of the embryo [4] [10].

bcd mRNA is transcribed in the nurse cells and transferred into the oocyte, where it can be detected already in early stages of oogenesis. Within the oocyte *bcd* mRNA localizes at the anterior end (Fig. 2) [27]. This process is an example of mRNA localization achieved through

active transport. The signal responsible for the anterior localization of *bcd* was first identified as a 625 nucleotide stretch in the 3' untranslated region (3'UTR) of the mRNA [28]. This region contains shorter sequence elements, called *bcd* localization elements (BLE) that are necessary for different steps of *bcd* localization [28]. BLE1, a 50 nucleotide long sequence element able to form a stem loop structure, is required for transport of *bcd* mRNA from nurse cells to the oocyte [29]. Additional sequence elements are required during other steps of *bcd* mRNA localization. In the same manner as BLE1, these additional elements can also fold into stem loop structures. Mutations in the primary sequence that do not impair the secondary structure of BLEs still allow proper *bcd* mRNA localization [30] [31]. It has also been shown that *bcd* mRNA is able to dimerize both *in vitro* and *in vivo*, through a stem loop structure in the 3'UTR, and that the dimerization is crucial for the interaction with the protein Staufen (Stau) [31].

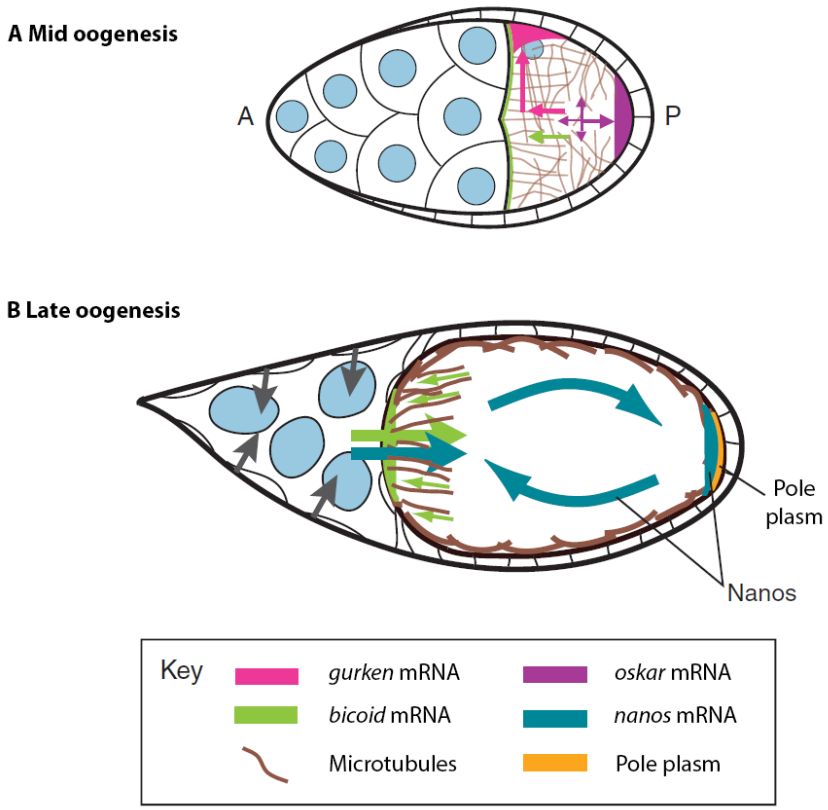

Figure 2: *bcd*, *osk*, *nos* and *gurken* mRNA localization during *Drosophila* oogenesis. All four mRNAs are transcribed in the nurse cells and transported into the oocyte by a microtubule- and Dynein-dependent process. (**A**). Maternal mRNA localization during mid oogenesis. Once in the oocyte,

gurken mRNA accumulates next to the oocyte nucleus, at an anterior-lateral position of the oocyte cortex. Once in the oocyte *bcd* mRNA accumulates at the anterior cortex. *osk* mRNA accumulation at the posterior pole is a microtubule- and Kinesin-dependent process. (**B**). mRNA localization during late oogenesis. During nurse cell cytoplasmic dumping into the oocyte, *bcd* is continuously transported to the anterior cortex by Dynein. *nos* mRNA accumulation at the posterior depends on cytoplasmic flows and Osk protein. When nurse cell dumping occurs, cytoplasmic streaming distributes *nos* RNPs throughout the ooplasm and those which reach the posterior pole are trapped by the pole plasm. Arrows represent the movement of mRNAs and other cytoplasmic components within the oocyte during streaming. A: anterior. P: posterior.

Adapted from [3].

The anterior localization of *bcd* mRNA is a multi-step process during which redundant mechanisms act to ensure proper *bcd* localization. The initial transport of *bcd* mRNA into the oocyte is a microtubule- and Dynein-dependent process. *bcd* mRNAs travel through the nurse cell cytoplasm along microtubules towards the ring canals, transported by the microtubule motor Dynein [32] [33] [34]. During mid oogenesis accumulation of *bcd* mRNA at the anterior of the oocyte requires Exuperantia (Exu) protein activity [27] [32]. Exu and *bcd* mRNAs associate in large ribonucleoprotein particles (RNP) in the nurse cells and travel together into the oocyte (Fig. 2A). In Exu mutant egg chambers, *bcd* mRNA is able to reach the oocyte, but instead of accumulating at the anterior diffuses throughout the ooplasm [32]. It has been hypothesized that after entering into the oocyte *bcd*-Exu RNPs are remodeled in order to incorporate those factors that are required for anterior accumulation [33]. Exu is phosphorylated by the kinase Par-1 and this modification is required for anterior localization of *bcd* mRNA [35].

During the later stages of oogenesis, from nurse cell dumping onwards, *bcd* mRNAs continue to accumulate at the anterior of the oocyte. The process is microtubule- and Dynein-dependent but appears to be mechanistically different from the one active during mid oogenesis [36]. Indeed, mutations in two genes: *swallow* and *stau*, specifically compromise this late stage in *bcd* mRNA localization [27] [36]. In particular, Stau, a double stranded RBP, colocalizes with *bcd* RNPs during late stages of oogenesis [36]. It has been hypothesized that in this process, *bcd* RNPs move on a particular subpopulation of microtubules nucleated from

the anterior cortex, different from the microtubules present on the rest of the oocyte cortex (Fig. 2B) [36] [37].

Finally, at the end of oogenesis, *bcd* mRNAs become anchored at the anterior cortex through an actin-dependent mechanism [38]. This mechanism is important for maintaining *bcd* at the anterior in mature oocytes prior to fertilization, which might only occur several days later. Swallow protein plays an important role in anchoring *bcd*. Indeed, Swallow localizes near the oocyte plasma membrane, where it seems to have a role in regulating the actin cytoskeleton that is required for *bcd* mRNA anchoring at the anterior [39]. The final step of *bcd* localization is triggered by egg activation, as a consequence of which the actin cytoskeleton at the cortex of the egg undergoes a rapid rearrangement that releases *bcd* RNPs from the anterior cortex, thus generating an anterior to posterior gradient [38] [40].

bcd mRNA encodes a transcription factor and is translated upon egg fertilization. As a consequence of the graded distribution of *bcd* mRNA and of the movement of Bcd protein itself, Bcd forms a morphogen gradient that controls anterior-posterior patterning in a concentration-dependent fashion (Fig. 3A) [40] [41]. In early embryogenesis, during the blastoderm phase, following the 13 rapid nuclear divisions the majority of the nuclei move to the periphery of the embryo where cellularization occurs. As a consequence of the Bcd gradient, depending upon their position cells incorporate different amounts of Bcd protein, resulting in differential gene expression along the anterior-posterior axis [10].

Bcd regulates the expression of different genes involved in the formation of the anterior structures of the future animal (head and thorax). Among these targets an important role is played by *hunchback* (*hb*). Bcd activates *hb* transcription at the anterior of the embryo directly binding regulatory DNA elements upstream of the *hb* gene [42]. Hb protein is also a transcription factor involved in the regulation of genes required for embryonic patterning and development. In addition to the zygotic expression of *hb* resulting from Bcd transcriptional activation, *hb* mRNA is produced maternally during oogenesis and is distributed uniformly in the oocyte [43]. Maternal *hb* mRNA is translated at the anterior of the embryo and repressed

towards the posterior by the activity of a complex formed by Nos, Pumilio (Pum) and Brain tumor (Brat), which restricts the activity of maternal *hb* anteriorly (Fig. 3C) [44]. Another gene activated by Bcd is *orthodenticle* (*otd*). Sequences upstream of the *otd* gene are recognized by Bcd, promoting *otd* transcription and thus formation of head structures [45].

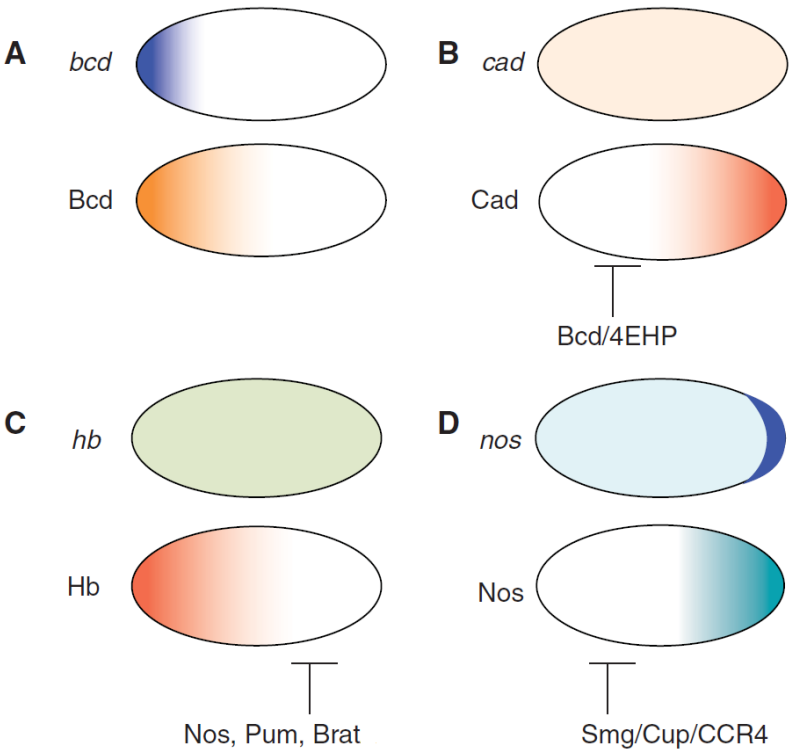

Figure 3: Bcd, Cad, Hb and Nos morphogen gradients in *Drosophila* embryo. (**A**). *bcd* mRNA accumulates at the anterior pole. *bcd* translation generates an anterior to posterior gradient of Bcd morphogen. (**B**). *cad* mRNA is evenly distributed in the embryo. Bcd together with 4EHP blocks *cad* translation at the anterior, thus generating an opposing gradient of Cad protein. (**C**). *hb* mRNA is evenly distributed in the embryo. *hb* translation is repressed at the posterior by Nos acting in complex with Pum and Brat. (**D**). *nos* mRNA accumulates at the posterior pole, but is also present throughout the embryo. Only posteriorly localized *nos* mRNAs are translated. Unlocalized *nos* mRNA is translationally repressed by Smg, which recruits Cup and the CCR4/NOT complex.

Adapted from [10].

In addition to acting as a transcription factor, Bcd also acts as a repressor of translation of specific mRNAs in the embryo, for example of *caudal* (*cad*) mRNA. Cad is a transcription factor involved in posterior patterning. *cad* mRNA is uniformly distributed in the embryo and its translation at the anterior is repressed by the binding of Bcd to the 3'UTR: Bcd binds the 4E homology protein (4EHP) and recruits it to the 5' end of *cad* mRNA; 4EHP then interacts with

the mRNA cap structure directly, inhibiting the assembly of an active cap-binding complex and thus repressing translation. This mechanism results in a posterior to anterior Cad protein gradient opposite to the Bcd gradient (Fig. 3B) [10] [46]. In this way, by regulating different steps of gene expression during early embryogenesis, the Bcd morphogen orchestrates the patterning of anterior structures of the future animal.

The posterior localization of *osk* and *nos* mRNAs is crucial for proper patterning of the posterior structures of *Drosophila* embryo. Posterior localization of *osk* mRNA restricts Osk protein activity to the posterior pole, where directs the assembly of a specialized cytoplasm called the pole plasm (Fig. 2) [47]. Here I will describe the mechanism of *nos* mRNA localization during oogenesis and the function of Nos protein in the embryo. The features of *osk* mRNA localization and Osk protein activity will be discussed in the next paragraphs.

Nos protein is the major player in patterning the posterior structures of the *Drosophila* embryo [48]. *nos* mRNAs localize at the posterior pole of the oocyte during late oogenesis and, as opposed to *bcd* mRNA transport, this is a passive process. The process requires several factors, including the previous localization of Osk, Vasa (Vas) and Tudor (Tud) proteins at the posterior pole [49]. The localization of *nos* mRNA is an inefficient process. Indeed, at the end of the process the majority of *nos* mRNA remains spread throughout the ooplasm and only 4% of the mRNA localizes at the poster pole of the oocyte [50]. The mechanism that drives *nos* posterior localization is initiated when nurse cell dumping generates cytoplasmic flows in the oocyte. As a consequence of cytoplasmic streaming, *nos* particles move throughout the oocyte and those that reach the posterior pole are trapped and accumulate there [51]. It has been shown that *nos* RNPs that reach the posterior pole are incorporated into the pole plasm and anchored by the actin cytoskeleton (Fig. 2B) [51].

As in the case of *bcd* and other localizing mRNAs, *nos* transcripts contain the information required to achieve their ultimate posterior localization. In its 3'UTR, *nos* mRNA contains a large localization element that can be subdivided into four smaller localization elements [52]. These elements might be recognized by specific factors. It has been shown that the protein

Rumpelstiltskin (Rump), the *Drosophila* homolog of the heterogeneous nuclear ribonucleoprotein (hnRNP) M, binds to one of the localization elements in the *nos* 3'UTR. This interaction is required for achieving proper *nos* localization [53]. An additional protein that has been recently identified as *nos* mRNA interactor *in vivo* is Aubergine (Aub). Aub affects *nos* localization and interacts *in vivo* with *nos* 3'UTR. Moreover, Aub interacts *in vivo* with Rump in an RNA-dependent manner [54].

Since the mechanism underlying *nos* localization is inefficient and the majority of the mRNA remains unlocalized, restriction of the source of Nos protein to the posterior pole is crucial and this is achieved by repression of translation and possibly degradation of unlocalized transcripts (Fig. 3D). Different mechanisms act in *nos* translational repression during different developmental stages. The *nos* 3'UTR contains a 90 nucleotide long region that is involved in translational regulation. Defined as the translational control element (TCE), the TCE partially overlaps with some of the localization elements [55]. During late oogenesis, translation of unlocalized *nos* transcripts is repressed by the activity of Glorund (Glo) protein. Glo is the *Drosophila* homolog of hnRNP F/H proteins and, by interacting with the TCE, prevents translation of unlocalized *nos* mRNAs [56]. During early embryogenesis another mechanism of repression is active that involves the protein Smaug (Smg). The *nos* TCE contains two Smg recognition elements (SRE) that can be bound by Smg [57]. *nos* mRNA bound by Smg can be repressed in two different ways. On one hand Smg can recruit the protein Cup by direct interaction. Cup binds the cap-binding protein eIF4E and by doing so interferes with the recruitment of the scaffold protein eIF4G by eIF4E, thus inhibiting the assembly of the 48S pre-initiation complex [58]. A second mechanism of Smg-mediated translational repression involves deadenylation. After binding the SRE in the *nos* 3'UTR Smg recruits the CCR4/NOT deadenylation complex, resulting in *nos* mRNA degradation (Fig. 3D) [59].

Only *nos* mRNAs localized at the posterior are translated and translational activation depends on Osk protein function. Osk can interact directly with Smg and relieve the CCR4/NOT-dependent deadenylation, allowing Nos protein to accumulate at the posterior pole

[59] [60]. Therefore *nos* RNPs that reach the posterior pole escape translational repression and degradation. Local *nos* translation at the posterior starts after fertilization and generates a gradient of Nos protein that mirrors the Bcd protein gradient (Fig. 3D) [61]. During early embryogenesis Nos (together with Pum and Brat proteins) controls abdomen formation by inhibiting *hb* translation at the posterior (Fig. 3C) [44]. This repressive mechanism allows the expression of genes required for abdomen formation.

The opposing gradients of Bcd and Nos regulate anterior-posterior patterning of *Drosophila* by acting on *hb* at different levels. This process in turn affects the activation and repression of other patterning genes that control formation of the head and the abdomen of the future animal [62].

1.3. *oskar* mRNA localization during oogenesis

osk mRNA localization at the posterior pole triggers the assembly of a specialized cytoplasm called pole plasm that contains all the molecular factors required for both abdomen and germ cell formation [47]. In the next paragraphs I will discuss the various aspects involved in *osk* mRNA localization and Osk protein function in the context of pole plasm assembly.

During early stages of oogenesis, up to stages 6 and 7, *osk* mRNA is detected in all the cells of the germline cyst, but appears to be more concentrated in the oocyte. Initially *osk* mRNA is evenly distributed in the ooplasm. During stages 8 and 9 the mRNA shows a slight and transient accumulation at the anterior of the oocyte and starts to accumulate at the posterior pole. From stage 10 onward *osk* mRNA is restricted at the posterior pole and the localization pattern does not change until the end of oogenesis. During early embryogenesis *osk* is detected at the posterior pole and begins to disappear during nuclear division 6 and 7 [63] [64]. Different mechanisms mediate the accumulation of *osk* mRNA at the posterior pole of the oocyte.

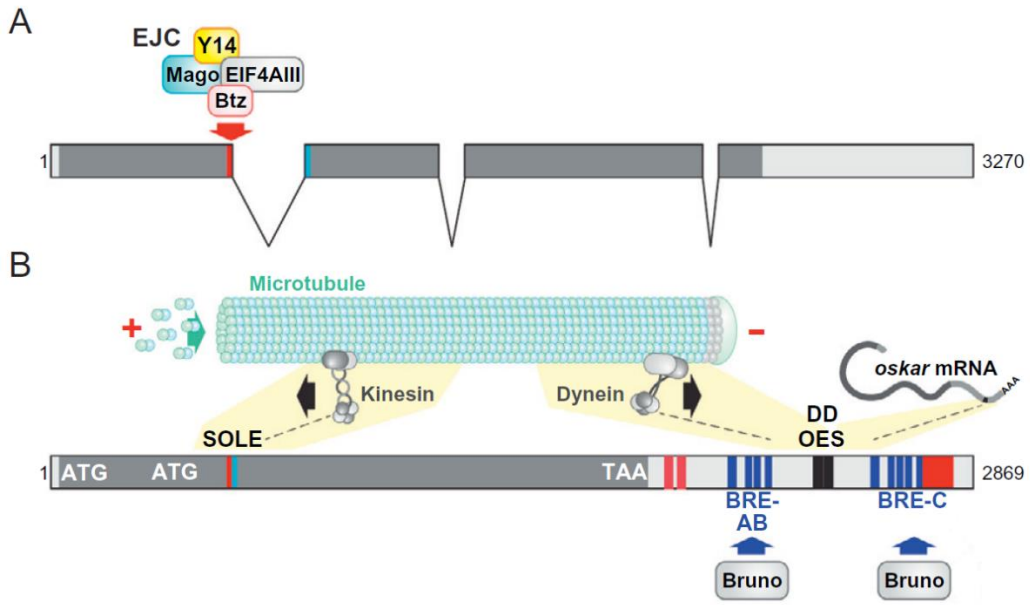

Figure 4: *oskar* pre-mRNA and mRNA sequences. (**A**). Structure of the *osk* pre-mRNA. *osk* pre-mRNA contains 3 introns that are removed by the splicing machinery. Splicing of the first intron leads to formation of the SOLE element, otherwise split between the first (red line) and second (blue line) exons. The EJC is deposited upon splicing 20-24 nucleotides upstream of exon-exon junctions. (**B**). *osk* mRNA structure. *osk* mRNA contains two in-frame start codons (ATGs). The SOLE element is required for Kinesin-mediated transport of the mRNA towards the posterior pole of the oocyte. Several RNA regulatory elements are present in the 3'UTR. The OES (black line), a stem-loop structure, is required for Dynein-mediated transport of the mRNA from the nurse cells into the oocyte. Within the OES, the loop promotes *osk* mRNA dimerization (DD). The *osk* 3'UTR contains several BREs (blue lines) that mediate translational repression and activation by Bru protein. IBEs (red lines) are required for *osk* translation at the posterior pole.

Adapted from [65].

osk mRNAs are transcribed in nurse cells, then transported into the oocyte - the first step in *osk* localization. A 67 nucleotide long element in the *osk* 3'UTR, the *osk* oocyte entry signal (OES), mediates this transport process. The OES folds into a stem loop structure and mediates the Dynein- and microtubule-dependent transport of *osk* mRNAs from the nurse cells into the oocyte (Fig. 4B) [66]. Different protein factors are involved in this transport process, among them Egl and BicD. Egl does not possess a classical RNA-binding domain (RBD) yet is an RBP that has been shown to be able to recognize specific localization signals within RNAs [67]. Furthermore a recent study suggests that Egl directly interacts with *osk* mRNA *in vivo* [68]. Egl directly interacts with BicD and both proteins interact with components of the Dynein

motor complex [69] [70]. Altogether these data show that *osk* mRNA transport into the oocyte is mediated by a cis-acting element in the 3'UTR that mediates the recruitment of Egl-BicD complex and Dynein, thus promoting the active transport of *osk* RNPs into the oocyte.

After *osk* mRNA transport into the oocyte during early oogenesis, a second mechanism of active transport mediated by Kinesin motor proteins is responsible for the posterior localization of *osk* mRNA during mid oogenesis (Fig. 2A). As previously mentioned, beginning from stage 7 the oocyte microtubule cytoskeleton disassembles and new microtubules nucleate from the anterior and lateral cortex of the oocyte with their plus ends extending into the ooplasm [14]. At this stage *osk* mRNA transport to the posterior pole requires the activity of the microtubule plus end motor protein Kinesin-1 [71] [72]. By *in vivo* observation of *osk* RNPs movement, it has been shown that the posterior accumulation is achieved through a series of random walks on a poorly polarized cytoskeleton. These walks are only slightly biased toward the posterior pole but they nevertheless suffice to ensure efficient posterior localization of *osk* [72].

Several sequence elements in *osk* mRNA are required for the Kinesin-1-dependent localization step. One of these RNA elements is assembled upon splicing of the *osk* pre-mRNA in the nurse cell nuclei. Indeed, it has been shown that splicing of the first intron in *osk* pre-mRNA is strictly required for posterior localization [73]. Upon removal of the first intron, the last 18 nucleotides of exon 1 3' end and the first 10 nucleotides of exon 2 5' end form the spliced *osk* localization element (SOLE), which folds into a stem loop structure (Fig. 4A) [74] [75]. As a consequence of splicing, the exon junction complex (EJC) is loaded on mRNAs, including *osk* mRNA, 20-24 nucleotides upstream of the splice junction. EJCs mark exon-exon boundaries and the complex is involved in several processes of mRNA post-transcriptional regulation [76]. In *Drosophila*, the four core components of the EJC, eIF4AIII, Mago nashi (Mago), Y14 and Barentsz (Btz), are required for *osk* mRNA transport to the posterior of the oocyte [77] [78] [79] [80]. Although splicing of the first intron is crucial for posterior localization of *osk* mRNA, intron-less transcripts that contain the *osk* 3'UTR can still localize correctly in wild-type oocytes, as a result of "hitch-hiking" with endogenous, spliced *osk* transcripts. This

mechanism involves RNA-RNA interactions between the stem loop structure of the OES element of two *osk* 3'UTRs, reinforced by proteins that have the capacity to dimerize or oligomerize. By this mechanism RNAs that lack the SOLE element can localize to the posterior together with *osk* mRNAs that are fully competent for posterior localization (Fig. 4B) [81]. Therefore, RNA-RNA interaction appears to be an additional mechanism that contributes to *osk* mRNA accumulation at the posterior pole.

After nuclear pre-mRNA processing occurs, *osk* is exported to the nurse cell cytoplasm where a Dynein-mediated mechanism transports *osk* RNPs into the oocyte, where a Kinesin-1-mediated mechanism takes over and mediates posterior localization of the mRNAs. How *osk* RNPs switch between opposite microtubule polarity motors during the different stages of localization is not completely understood. A recent study sheds new light on this mechanism. It has been shown that *osk* mRNA is mislocalized in *tropomyosin* mutant flies [82] and that Kinesin-1 recruitment to *osk* RNPs is mediated by the Tropomyosin 1 (Tm1) I/C isoform, that is able to bind the *osk* mRNA 3'UTR [83]. Tm1-mediated recruitment of Kinesin-1 seems to occur right after nuclear export of mature *osk* mRNAs to the perinuclear cytoplasm of nurse cells. The process of recruitment appears to be inefficient, as only a fraction of *osk* RNPs are loaded with Kinesin-1. However, the recruitment of Kinesin-1 by Tm1 is dynamic and ultimately ensures the posterior localization of the bulk of *osk* mRNAs. In the model proposed, Tm1 bound to the 3'UTR recruits inactive Kinesin-1, which once the RNPs are in the oocyte is activated by the SOLE element associated with the EJC on spliced *osk* mRNAs [83].

Two additional proteins that are involved in *osk* posterior localization are Hrp48 and Stau. Hrp48, the *Drosophila* homolog of hnRNP A/B, is required for *osk* localization, as in oocytes produced from *hrp48* mutant females, *osk* mRNAs fail to accumulate at the posterior pole [84]. Hrp48 directly interacts with the 5' region and 3'UTR of *osk* mRNA and regulates *osk* translation in addition to localization [84] [85]. *osk* mRNA also fails to localize at the posterior in *stau* mutants [63]. Stau protein contains five RBDs and specifically interacts with double

stranded RNA. It has been proposed that Stau RBD2 might be involved in regulating *osk* localization while RBD5 is required for *osk* translational activation [86].

During the late stages of oogenesis, when cytoplasmic streaming from the nurse cells occurs, an additional mechanism of *osk* posterior localization becomes active that involves the proteins Rump and Lost. Rump binds directly to the *osk* 3'UTR and recruits Lost by direct interaction. This mechanism specifically boosts the amount of *osk* mRNAs localized at the posterior pole, ensuring the accumulation of an amount of *osk* mRNA and protein sufficient for germ plasm assembly [87].

The anchoring of *osk* RNPs after they have been transported to the posterior pole is also a crucial step in germ plasm assembly. This mechanism requires the activity of Osk protein itself, as well as the endocytic pathway and rearrangements in the F-actin cytoskeleton. Posteriorly localized *osk* mRNAs are translated into two protein isoforms Long Osk and Short Osk (Fig. 5) [88]. The features of the two Osk isoforms will be discussed in detail in the next paragraphs, but a short introduction is required here in order to explain the mechanisms underlying *osk* mRNA anchoring at the posterior pole. The Long Osk isoform is required for anchoring both *osk* mRNA and Short Osk at the posterior pole during late oogenesis. In the absence of Long Osk, pole plasm components detach from the posterior, resulting in posterior patterning defects and absence of germ cells [89]. Osk protein somehow stimulates endocytosis, and the Long Osk isoform has been shown to recruit several endosomal proteins [90]. Stimulation of endocytosis by Osk is required for anchoring the pole plasm at the posterior pole. Moreover Osk promotes rearrangement of the F-actin cytoskeleton. Long F-actin filaments project from the posterior cortex into the ooplasm and similar structures are not observed in other parts of the oocyte cortex [90] [91] [92]. These processes appear to act cooperatively to stably anchor *osk* mRNA and Osk protein at the posterior pole, spatially restricting Osk activity during oogenesis and early embryogenesis.

1.4. *oskar* mRNA translational regulation

Translation of unlocalized *osk* mRNAs leads to severe patterning defects [47] [93]. In order to block ectopic Osk synthesis and restrict Osk protein accumulation to the posterior pole, translation of *osk* mRNA must be strictly regulated.

Several mechanisms are active during oogenesis that repress *osk* mRNA translation. Some data suggest that during early oogenesis *osk* translation is inhibited by the RNA interference (RNAi) pathway. Indeed, mutations in several genes involved in RNAi as *armitage*, *aub*, *spindle-E* and *maelstrom*, lead to premature *osk* mRNA translation [94] [95].

An important protein required for *osk* translational repression in the ooplasm is Bruno (Bru). Bru is an RBP containing three RNA recognition motifs (RRM) that recognize sequence elements, called Bru response elements (BRE), in the *osk* 3'UTR (Fig. 4B). The direct interaction of Bru with *osk* mRNA represses its translation [96] [97]. Experimental evidence indicates that different mechanisms mediate Bru-dependent repression of *osk* translation. One of these mechanisms requires Bru interaction with the protein Cup, an eIF4E-binding protein. By interacting with the 5' cap-binding protein eIF4E, Cup prevents recruitment of eIF4G, and hence of the small ribosomal subunit to unlocalized *osk* mRNAs [98]. An additional Bru-mediated mechanism that might be involved in *osk* translational regulation requires Cup activity and involves deadenylation. Indeed Cup might promote *osk* poly(A) tail shortening by the direct recruitment of the CCR4/NOT complex and at the same time inhibit decapping and decay processes thus maintaining the mRNA in a repressed state [99]. This mechanism is supported by data showing that *osk* translational activation at the posterior pole requires polyadenylation [100].

osk mRNA translation is repressed by an additional Bru-mediated mechanism that does not involve Cup. Bru alone can promote *osk* mRNA oligomerization into large silencing particles inaccessible to ribosomes [101]. Formation of these silent *osk* RNPs, coupled with active transport, might represent an efficient mechanism for localizing a large number of mRNA

molecules at the posterior pole in a translational repressed state. An additional protein involved in *osk* oligomerization is hnRNP I, also known as polypyrimidine tract binding protein (PTB). PTB is required for *osk* translational repression and binds directly to multiple binding sites in the 3'UTR of *osk* transcripts. PTB contains four RRMs and by binding to different RNA molecules can promote the formation of large *osk* RNPs [102].

Only those *osk* mRNAs that localize at the posterior pole are translated into Osk protein; hence, starting from mid oogenesis, Osk accumulates at the posterior pole of the oocyte where it promotes pole plasm formation. The mechanism required for *osk* translational activation is not fully understood but studies over the last 20 years provide some information regarding this process.

Interestingly it has been shown that Bru, in addition to its activity as a translational repressor, is involved in the activation of *osk* translation. The *osk* 3'UTR contains two clusters of BREs, both of which promote transcriptional repression of unlocalized *osk* mRNAs (Fig. 4B). The BRE cluster located close to the 3' end of the 3'UTR is additionally required to promote translation. Both repression and activation are mediated by the different BREs and can occur in trans on different mRNA molecules, such that mRNAs concentrated in the same RNP can be regulated in a coordinated manner [103]. Additional sequence elements in the *osk* 3'UTR are important for translational activation. These sequences are recognized by the insulin growth factor II mRNA-binding protein (IMP) and are referred to as IMP-binding elements (IBE). *osk* mRNA contains 13 IBEs that are required for translation at the posterior pole (Fig. 4B). Interestingly, IMP protein is not required for *osk* translation [104].

An important function in activating *osk* mRNA translation at the posterior pole is exerted by oo18 RNA-binding protein (Orb). Orb is the *Drosophila* homolog of the cytoplasmic polyadenylation element binding protein (CPEB), whose activity is involved in regulation of *osk* mRNA poly(A) tail length. Orb colocalizes with *osk* at the posterior pole, and if Orb is lacking *osk* mRNAs bear shorter poly(A) tails. This reduces *osk* translation and Osk protein accumulation, leading to patterning defects [100]. Orb's effect on the length of the poly(A) tail

on *osk* transcripts is exerted through the direct interaction of Orb with two poly(A) polymerases (PAP), PAP and Wispy, the first being required during mid oogenesis and the second during late oogenesis [105] [106].

Finally, *osk* mRNA itself is endowed with a function as a regulatory RNA. This is a non-coding function of *osk* mRNA and is linked to the presence of BREs in the 3'UTR. *osk* mutant alleles that strongly reduce or completely abolish *osk* mRNA synthesis cause oogenesis arrest much earlier than the classical *osk* mutant alleles, which impair Osk protein function or production, although many support accumulation of the mutant *osk* mRNA at the posterior pole. The *osk* 3'UTR alone is sufficient to rescue the early oogenesis arrest phenotype of the so-called RNA null alleles [107]. A model has been proposed to explain the non-coding function of *osk* mRNA. In this model *osk* mRNA, via one BRE cluster in the 3'UTR, sequesters Bru molecules reducing the likelihood of Bru binding to other targets. Supporting this model, reducing Bru levels genetically in the oocyte partially rescues the *osk* mRNA null phenotype [108].

1.5. Two Oskar protein isoforms accumulate at the posterior pole

The *osk* mRNA contains two in-frame start codons, both of which are used for translational initiation (Fig. 4B), leading to production of two protein isoforms: Long Osk (606 amino acids) and Short Osk (467 amino acids). The sole difference between the isoforms is the N-terminal extension that is present in Long Osk and absent in Short Osk (Fig. 5) [88].

Both Osk isoforms contain two globular domains separated by an intrinsically disordered region; the crystal structures of the two domains have been solved [109] [110]. The domain at the N-terminus (residues 139 to 240 in Long Osk) is a so-called LOTUS (Limkain, Oskar and Tudor containing proteins 5 and 7) domain (Fig. 5). This particular domain was identified in bioinformatic studies and proposed to act as an RBD [111] [112]. However, *in vitro* the LOTUS domain of Osk shows no RNA-binding activity. On the other hand it was shown that the LOTUS

domain forms dimers and interacts with the DEAD-box helicase Vas, an important component of the pole plasm [109] [110]. In a recent study the crystal structure of Osk LOTUS domain in complex with the C-terminal moiety of Vas was solved [113]. This study further revealed that the RNA helicase activity of Vas is stimulated by the interaction with LOTUS domain. In addition, Vas mutations that abolish the interaction with the LOTUS domain *in vitro* impair the recruitment of Vas to the pole plasm. These data suggest that Osk-dependent recruitment of Vas to the pole plasm is mediated by the N-terminal LOTUS domain.

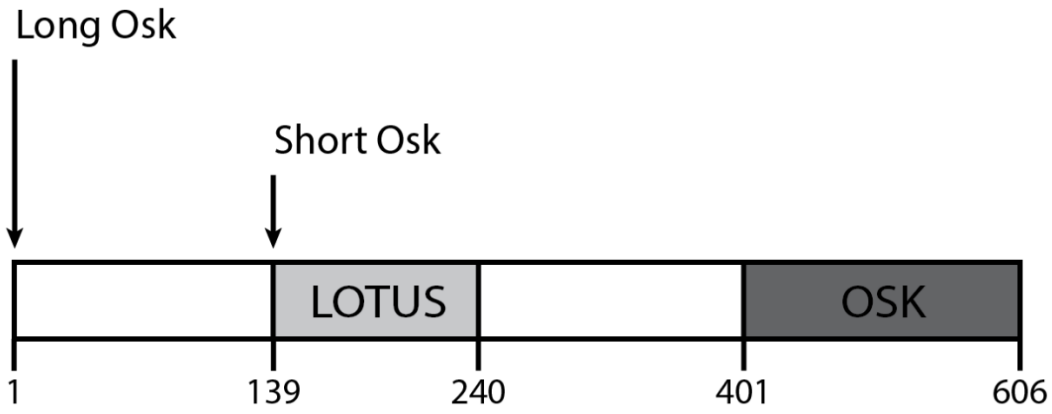

Figure 5: Oskar protein isoforms. The presence of two in-frame start codons in *osk* mRNA results in its translation into two Osk protein isoforms: Long Osk and Short Osk. Long Osk contains an N-terminal extension that is absent in the Short isoform. Osk contains two globular domains: a LOTUS domain at the N-terminus and an OSK domain at the C-terminus. The LOTUS domain is required for Osk dimerization and interacts with Vas. The OSK domain binds RNAs *in vitro*.

The domain at the C-terminus (from residues 401 to 606 in the Long Osk) of Osk protein has named the OSK domain [109]. The crystal structure of the OSK domain revealed a folding similar to SGNH hydrolase enzymes. Nevertheless, the OSK domain lacks several residues required for the catalytic activity of the SGNH class of enzymes, suggesting that is catalytically inactive. In addition, it was shown that the OSK domain is able to bind RNAs *in vitro* [109] [110].

Despite the fact that the only difference between the two isoforms is the presence of the N-terminal extension, Short Osk and Long Osk possess different biological activities. The Short Osk protein isoform (the more abundant of the two isoforms) is the one possessing pole plasm

inducing activity, therefore its function is critical for both posterior patterning of the embryo and germ cell formation [88] [89] [114]. The pole plasm inducing activity of Short Osk relies on the ability of this specific isoform to recruit other pole plasm factors, such as Vas, Valois (Vls) (which is required to recruit Tud protein) and Lasp [115] [116] [117]. At the posterior pole, Short Osk induces formation of and localizes within polar granules, electron-dense and membrane-less structures that are large RNPs containing other molecular components, both proteins and RNAs, of the pole plasm [89].

In contrast to Short Osk, the Long isoform is not required for abdomen and germ cell formation, but its activity is essential for anchoring the pole plasm to the posterior cortex of the oocyte and early embryo. As discussed above, in the absence of Long Osk protein, Short Osk detaches from the posterior cortex and disperses in the ooplasm together with other pole plasm components such as Vas, causing defective germ cell and abdomen formation [89]. At the posterior pole, Long Osk is associated with endosomes, and this isoform has been shown to stimulate endocytosis at the posterior pole of the oocyte. Nevertheless, oocytes that produce only the Long Osk isoform at the posterior pole form endocytic structures that are strikingly different from those in wild-type oocytes. Therefore, although Long Osk stimulates endocytosis, the activity of both isoforms is required for correctly regulating the endocytic pathway at the posterior pole [92]. A recent study has shown that Long Osk is required for the transmission of mitochondria to the germline. Mitochondria localize at the posterior pole during the late stages of oogenesis independently on germ plasm assembly. Long Osk function is required to maintain mitochondria at the posterior pole, thus promoting their transmission to the germline. In the absence of Long Osk fewer mitochondria are inherited by the germ cells, causing a reduction in the number of germ cells that form during embryogenesis [118].

It has been proposed that the different biological activity of the two Osk protein isoforms might be explained by an uncharacterized inhibitory effect of the N-terminal extension on the activity of the protein. According to this hypothesis, the presence of this extension in the Long Osk isoform blocks the pole plasm inducing activity of the rest of the protein. Another

hypothesis is that the differential biological activities of the two might be due to their different subcellular localization. Being localized in the polar granules or on the endosomes might expose the two isoforms to different arrays of interacting partners [109]. Gaining more knowledge regarding the isoform-specific interactors of Osk might help understand the molecular mechanisms underlying the pole plasm inducing activity of the protein.

1.6. Oskar triggers pole plasm assembly

Germline specification is a crucial event in embryonic development. The germline plays a special role compared to the soma, as germ cells transmit the genetic information of the organism from one generation to the next. In metazoans, the germline is specified early during embryonic development by one or the other of two distinct mechanisms, preformation and epigenesis. In the case of preformation, the germline is specified through the localization of factors within the oocyte that assemble into a specialized cytoplasm called germ plasm. Those cells that will retain the germ plasm will adopt a germ cell fate. In the case of preformation, the germ cells form in response to specific signals from the surrounding environment [119].

Drosophila melanogaster is one of the best studied model organisms in which preformation is the mechanism that segregates the germline. As previously mentioned, the molecular factors required for germline specification start accumulating during mid oogenesis at the posterior pole of the oocyte where they assemble the germ plasm. Because the assembly occurs at the posterior of the oocyte, the *Drosophila* germ plasm is also called the pole plasm. During early embryogenesis, when cellularization occurs, the pole plasm is incorporated by a few cells forming at the posterior pole. These cells, the so-called “pole cells”, are the primordial germ cells (PGC) of *Drosophila* that later in development will form the germline [119].

Several experiments performed during the ‘60s and ‘70s were aimed at understanding the role of the pole plasm in *Drosophila* development. Irradiation of the posterior pole with UV light

was used to destroy pole plasm of embryos before cellularization. Flies generated from these embryos were sterile as a consequence of the reduced number of PGCs that formed [120]. On the other hand, transplantation of pole plasm from a non-irradiated embryo to the posterior of an UV irradiated one, restored PGC formation [121]. In addition, in now classical developmental biology experiments, it was shown that transplantation of pole plasm from a donor embryo to an ectopic position in a host embryo induced formation of cells that displayed characteristics of germ cells at the ectopic site [122]. These experiments demonstrated that pole plasm is sufficient for germline specification in *Drosophila* and also defined the cytological features of the pole plasm. Besides being assembled early during oogenesis, pole plasm contains mitochondria and large electron-dense structures called polar granules [122]. Today we know that polar granules are large RNPs.

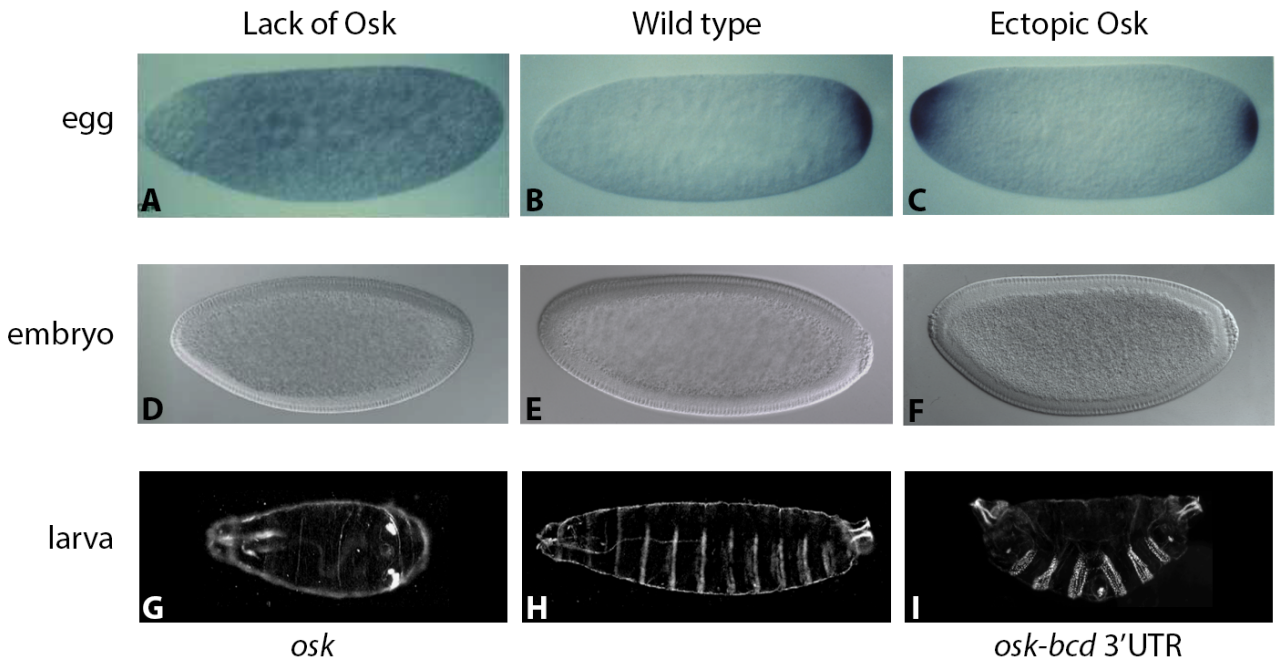

Figure 6. Oskar is both necessary and sufficient for abdomen and germ cell formation in *Drosophila*. In wild-type embryos, Osk accumulates at the posterior pole (**B**). As a consequence, germ cells form at the posterior (**E**) and abdominal structures are formed (**H**). When Osk is absent (**A**) no germ cells form in the embryo (**D**) and patterning of the abdomen fails (**G**). By using a transgene, it is possible to trigger Osk accumulation at the anterior pole, in addition to the canonical posterior accumulation (**C**). As a consequence, germ cells form both at the anterior and at the posterior (**F**) and an additional abdomen forms instead of the head (**I**).

From Dr. Anne Ephrussi

After more than 20 years the germline determinant in *Drosophila* was shown to be encoded by *osk*. Embryos that lack *osk* fail to assemble a functional pole plasm at the posterior pole and as a consequence lack pole cells and abdominal structures (Fig. 6A, D and G) [123]. By using a transgenic construct in which the *osk* 3'UTR was replaced by the *bcd* 3'UTR, it was possible to engineer the accumulation of Osk protein at the anterior pole of the embryo (in addition to the normal posterior accumulation) (Fig. 6C). In this particular condition pole cells formed at both poles and embryos developed two abdomens in a perfect mirror image (Fig. 6F and I) [47]. This experiment demonstrated that *osk* is both necessary and sufficient for the assembly of a fully functional pole plasm, which in *Drosophila* is required for both germline specification and abdomen formation (Fig. 6).

The mechanisms involved in anterior-posterior axis specification have been discussed in the previous paragraphs, therefore in the following paragraphs I will describe the mechanisms underlying germ cell fate specification. The pole plasm assembly activity of Osk relies on the ability of the protein to recruit other pole plasm components. One of these factors is the RNA helicase Vas, which is directly recruited to the posterior pole by Short Osk [109] [113] [115]. Several *vas* mutant alleles were identified in the last 30 years with strikingly different phenotypes. Some *vas* alleles lead to embryos that lack PGCs and abdominal structures, while others cause defects during early oogenesis linked to a failure in *gurken* translational activation [124] [125]. Vas function is crucial for posterior patterning and germline specification and in this context Vas might act as translational regulator of mRNAs such *osk* and *nos*. Indeed, in *vas* mutant ovaries Osk and Nos protein levels are reduced [55] [88]. The molecular mechanism underlying Vas activity in translational regulation of posterior localized mRNAs is not clear. It was proposed that a direct interaction between Vas and eIF5B is crucial for Gurken accumulation in the oocyte and might underlie a regulatory mechanism for other mRNAs [126]. However, a recent study did not detect a direct interaction between Vas and eIF5B, suggesting that Vas might recruit eIF5B indirectly [113].

An additional important component of the pole plasm is the protein Tud. Embryos from *tud* mutant flies show posterior patterning defects and lack PGCs [127]. Tud recruitment to the pole plasm is mediated by Vls, a protein that interacts with both Tud and Osk [116]. Tud is a large protein that contains 11 copies of the Tud domain. Tud domains are involved in protein-protein interactions and for binding require methylated arginine and lysine residues on the target proteins [128] [129]. Different Tud domains within the Tud protein sequence have been shown to be required for proper abdomen and germ cell formation, and it has been proposed that Tud functions as a scaffold protein, interacting with several targets within the pole plasm [128] [130]. Protein methylation at the posterior pole required for efficient Tud binding might be provided by Capsuleen, an arginine methyltransferase that interacts with Vls and is required for posterior patterning and germ cell formation [131] [132].

The mechanisms underlying the localization of *nos* mRNA and the activity of Nos as a morphogen during embryonic development were discussed above. However, in addition to its role in posterior patterning, *nos* is also required for germ cell formation. PGCs that lack Nos fail to migrate and to colonize the future gonads of the animal, possibly because gene expression that usually starts in the gonads, starts prematurely while the cells are migrating [133]. In the absence of Nos, PGCs express somatic genes, indicating that Nos function is required to maintain a germline fate [134]. Nos is also required to block mitosis of the PGCs before colonization of the gonads occurs. In this regard it has been proposed that Nos together with Pum blocks *CycB* mRNA translation in PGCs, thus downregulating cell proliferation [135] [136].

polar granule component (*pgc*) is another pole plasm component that acts in germline specification. *pgc* mRNA starts accumulating at the posterior pole during stage 11 of oogenesis and its localization depends on Osk [137]. *pgc* mutant embryos form PGCs normally but during embryogenesis these cells degenerate without reaching the future gonads of the animal causing sterility [138]. Initially identified as a non-coding RNA it has been shown that *pgc* transcripts encode a short peptide (71 amino acids) that acts to repress gene expression in

PGCs [138] [139]. RNA polymerase II transcriptional activity is linked to the phosphorylation state of the C-terminal domain (CTD) of the largest subunit. In *Drosophila*, the CTD of RNA polymerase II consists of 42 repeats of a heptapeptide motif that can be phosphorylated at different positions. Serine 5 (Ser5) phosphorylation is a hallmark of transcriptional initiation while phosphorylation of Serine 2 (Ser2) marks elongation. In *Drosophila*, high levels of phosphorylated Ser2 and Ser5 can be detected in somatic nuclei, while low levels of the same markers are present in nuclei of the PGCs. Transcriptional quiescence of PGCs appears to be linked to downregulation of RNA polymerase II-dependent transcription [140]. It has been shown that Pgc plays a fundamental role in inhibiting Ser2 phosphorylation in PGCs [138] [139]. In *Drosophila* Ser2 phosphorylation is catalyzed by the positive transcription elongation factor b (P-TEFb), a complex formed by Cdk9 and Cyclin T proteins. The fact that both Cdk9 and Cyclin T directly interact with Pgc led to the proposal of a mechanism for Pgc activity, according to which Pgc by binding P-TEFb interferes with its recruitment to chromatin, thus suppressing zygotic transcription in PGCs [138].

Another mRNA that localizes at the posterior pole in an Osk-dependent manner is *germ cell-less* (*gcl*). *gcl* appears to be exclusively required for PGC formation and viability, as *gcl* mutant embryos have no patterning defects but show a strong reduction in the number of PGCs [141] [142]. Gcl protein can be detected already in the germarium, is present in the ooplasm, and associates with nuclear membranes in the nurse cells [143]. Starting in embryogenesis, Gcl accumulates at the posterior pole and remains restricted to the PGCs until their migration into the gonads [141]. It has been proposed that Bru activity is required for restricting Gcl activity to the posterior pole and that, by binding the *gcl* 3'UTR, Bru represses *gcl* translation in the ooplasm thus restricting the expression of Gcl protein [143]. Gcl exerts two different functions in PGC formation and maintenance: when PGCs form, Gcl activity controls centrosomal dynamics, which is essential for correct formation of the future germline [144]; additionally, Gcl acts to maintain the transcriptional quiescence of PGCs. Indeed, PGCs that lack Gcl activate transcription of several somatic genes [145].

In conclusion, pole plasm assembly is crucial for posterior patterning and for establishing the future germline in *Drosophila*. Osk is the key factor involved in pole plasm assembly and most of the proteins and RNAs that accumulate at the posterior pole depend on the initial accumulation and activity of Osk protein. Even if we have accumulated considerable experimental information regarding Osk, the molecular mechanisms underlying its pole plasm inducing activity remains quite unclear. In this regard, it is interesting to note that many maternal mRNAs have been shown to localize at the posterior pole during oogenesis and early embryogenesis, and most of these localize either in an *osk* RNA- or Osk protein-dependent manner [146] [147]. Only recently it has been shown that Osk protein possesses an RNA-binding activity. Further analysis of this feature might allow us to better understand the biological function of the germ plasm inducer Osk.

1.7. Oskar associates with polyadenylated mRNAs *in vivo*

It was recently demonstrated that Osk associates with polyadenylated mRNAs in *Drosophila* embryos. The first experimental evidence came from a work of Jeske and colleagues in 2015 [109]. In this study 0 to 2 hour old *Drosophila* embryos were subjected to an mRNA interactome capture strategy [148]. Using this experimental approach, wild-type embryos were irradiated with UV light, which induces the formation of covalent bonds between RNAs and proteins in direct contact. Lysates prepared from these embryos were incubated with magnetic beads coupled with oligo(dT) stretches that can anneal with the poly(A) tails of the mRNAs. In this way protein-mRNA complexes were purified and analyzed to identify protein factors associated with the transcriptome of the early embryo (Fig. 7A). Upon UV crosslinking and oligo(dT) capture, Osk protein, as well as the cytoplasmic poly(A)-binding protein (PABPC), a well-known RBP, was specifically recovered from the embryo lysates (Fig. 7B). This experiment showed that Osk is able to associate with mRNAs *in* vivo*,* confirming in a physiological context the RNA-binding activity detected *in vitro* for the OSK domain [109].

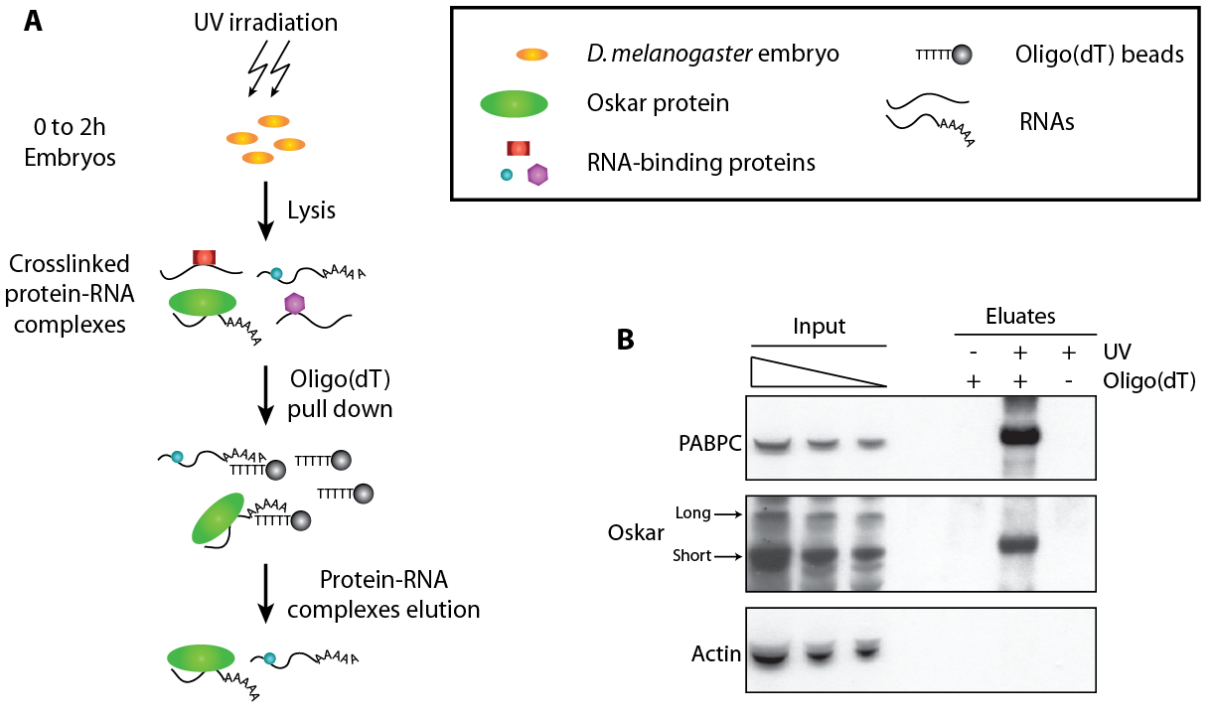

Figure 7. Oskar associates with polyadenylated mRNAs *in vivo*. (**A**). Interactome capture experimental workflow. 0 to 2 hour old *Drosophila* embryos are irradiated with UV light. After lysis, embryo extracts are incubated with magnetic beads conjugated to oligo(dT) stretches that anneal to the poly(A) tail of mRNAs. After stringent washing steps, the bound protein-mRNA complexes are eluted. (**B**). Western blot analysis of an interactome capture experiment on *Drosophila* embryos. Osk signal is specifically detected only after UV irradiation and oligo(dT) pull-down, as well as PABPC (used as a positive control). Actin, an abundant protein that does not bind RNA (used as a negative control), is not detected in the eluates. No Osk or PABPC signals are detected in the 2 negative control conditions: non-irradiated embryo extracts incubated with oligo(dT) beads and UV irradiated embryo extracts incubated with beads lacking oligo(dT) stretches.

Adapted from [109].

Subsequently, in 2016 the association of Osk with mRNAs in 0 to 2 hour old embryos was confirmed by another research group. In this case, two different mRNA interactome capture methods were applied to identify on a proteome-wide level *Drosophila* RBPs. Among the numerous targets identified in that study was Osk protein [149].

Osk's ability to bind polyadenylated RNAs *in vivo* suggests that the protein is a genuine RBP, even if it does not contain a canonical RBD. In this regard, several studies conducted in different biological contexts, in cell lines and in several organisms, indicated that proteins lacking canonical RBDs can bind RNAs *in vivo* [149] [150] [151] [152] [153] [154]. In the case

of Osk, RNA-binding might be crucial during pole plasm assembly. Osk might be involved in direct recruitment of some mRNAs to the posterior pole, it might regulate their translation, or it might be involved in regulating the stability of target transcripts. The identification of the transcripts bound by Osk *in vivo* and of the binding site(s) on the target mRNAs might provide useful information to understand the molecular mechanism by which Osk promotes pole plasm assembly and function.

1.8. Methods for the study of protein-RNA interactions *in vivo*

Gene expression processes can be regulated at different time points during the life of an mRNA, from transcription and processing in the nucleus, to translation and degradation in the cytoplasm. All the different steps and processes in the mRNA life cycle are precisely regulated by a large number of RBPs that dynamically associate with mRNAs at different locations in the cell [76].

A first step towards understanding the physiological role of a given RBP is the identification of the transcripts that are bound by the protein. Several methods both *in vitro* and *in vivo* are available as tools for answering this question. Although *in vitro* methodologies can provide useful information, techniques that allow the study of protein-RNA interactions *in vivo* are more informative. Indeed the ability to capture and detect an interaction that occurs in a specific biological context allows one to study protein-RNA interactions in their native physiological context.

High throughput sequencing technologies have become more advanced and cheaper in the last twenty years [155]. These methods generate a large amount of information and have been used for studying different aspects of RNA biology. High throughput sequencing technologies have also been used to analyze protein-RNA interactions *in vivo*. It is possible with these methods to identify on a transcriptome-wide level the binding targets of a given RBP, and is possible to obtain positional information regarding where on the RNA the specific

interaction occurs [156]. Data produced in recent years have expanded our knowledge regarding numerous RBPs. Today we know that a specific factor within a cell can bind to a huge number of target mRNAs and regulate diverse aspects of their biology, ranging from nuclear events, such as splicing and transcriptional termination, to cytoplasmic events such as translation and decay.

1.8.1. UV crosslinking of protein-RNA complexes

The first experimental attempts to identify target mRNAs of a specific RBP were based on RNA immunoprecipitation (RIP) approaches. In this type of experiment, an RBP of interest bound to its target mRNAs was isolated using a specific antibody and the transcripts were identified by hybridization to a microarray (RIP-ChiP) [157]. The RIP-ChiP method presents several disadvantages. Indeed the method is prone to contaminations. Other proteins that can bind RNAs as well can co-precipitate with the RBP of interest, thus producing false positive results [158]. In addition, RNA targets might not reflect an *in vivo* association but might result from reassociation of the RBP with unspecific targets in the lysate. Furthermore, RIP-ChiP does not yield information regarding the binding site of the protein.

In order to improve the specificity of these methodologies, protein-RNA crosslinking strategies were developed. Inducing the formation of covalent bonds between proteins and RNAs *in vivo* allows one to stabilize transient interactions that occur within a cell and facilitates the identification of the binding targets of RBPs in a physiological context. More than 30 years ago it was shown that UV light (from 250 to 270nm in wavelength) induces the formation of covalent bonds between proteins and RNAs. UV light is a zero-length crosslinking agent, and proteins and RNAs must be in direct contact in order to be crosslinked [159] [160]. These features make UV light the most useful crosslinking agent for protein-RNA interaction studies. Indeed, based on UV crosslinking, several methods have been developed for studying RBPs *in vivo*.

1.8.2. Transcriptome-wide approaches for studying protein-RNA interactions

The first protocol that used UV crosslinking to identify the RNA targets of an RBP was developed in 2003 in the Darnell lab. The methodology is called crosslinking and immunoprecipitation (CLIP) and was successfully applied for the identification of RNAs bound by Nova protein in the mouse brain. In the protocol, protein-RNA complexes stabilized by UV light irradiation, are immunoprecipitated. The RNAs crosslinked to the protein of interest are extracted and reverse-transcribed, cDNAs are generated and amplified by polymerase chain reaction (PCR), and finally analyzed by Sanger sequencing [161]. In 2008, a new version of the protocol was developed to exploit next generation sequencing (NGS) technologies: the high throughput sequencing of RNA isolated by crosslinking and immunoprecipitation (HITS-CLIP). In this method, RNA fragments crosslinked to the protein of interest are used as substrate for generating cDNA libraries suitable for NGS [162]. Coupling CLIP protocols with NGS allows one to generate a great amount of data that can be used to study protein-RNA interactions that occur *in vivo*, in detail on a transcriptome-wide level. Among the RBPs that have been studied by HITS-CLIP method there are: SRSF1, PTB, Fox2, Nova1, Nova2 and Argonaute 2 (Ago2) [163] [164] [165] [166] [167]. These methodologies also allow the identification of the binding site of the protein of interest on the target RNAs. This is achieved by analyzing deletions and mutations in the sequencing data. Indeed, it has been shown that the nucleotide(s) located at the sites of crosslinking are deleted or mutated during cDNA synthesis [156] [168].

CLIP methodologies use 254nm UV light to induce crosslinking. The crosslinking efficiency using this method is low, somewhere between 1% and 5% depending on the specific RBP [169]. In order to increase the efficiency of crosslinking, the Tuschl lab developed the photoactivable ribonucleoside-enhanced CLIP (PAR-CLIP) protocol. In this method cells are fed with the photoreactive nucleoside analog 4-thiouridine (4-SU) or 6-thioguanosine (6-SG), which are incorporated during transcription into RNA molecules. Upon 365nm UV light

irradiation, crosslinking events between proteins and RNAs occur; after immunoprecipitation of the RBP of interest, cDNA libraries are produced and subjected to NGS as described before. Is interesting to note that, upon crosslinking, the photoreactive nucleoside used in the PAR-CLIP can lead to specific mutations in the cDNA that can later be used for the identification with high resolution of the binding site of a specific RBP. This method, which involves the use of nucleoside analogs and results in a higher crosslinking efficiency, allowing one to map the binding site of an RBP, is particularly suited to cell culture experiments. PAR-CLIP has been used to identify the RNA targets of several proteins including: Pumilio homologue 2 (PUM2), Quaking (QKI), the insulin-like growth factor 2 mRNA-binding protein 1-3 (IGF2BP1-3), Ago2, hnRNP LL and AU-rich element-binding factor 1 (AUF1) [156] [170] [171] [172] [173].

The finding that in CLIP experiments, reverse-transcription stops before the nucleotide crosslinked to the RBP, resulting in truncated cDNA libraries was the basis of a new CLIP methodology published in 2010. This method, developed in the Ule lab, is called individual nucleotide resolution CLIP (iCLIP) [174]. The iCLIP protocol will be described in detail in the following chapters. This methodology allows on one hand to identify on a transcriptome-wide level all the RNA targets of a specific RBP and in addition allows the identification of the binding sites on the target transcripts. In order to study RBP binding specificity and function *in vivo*, iCLIP represents one of the most powerful experimental approaches available so far. Several proteins have been studied by iCLIP including hnRNP C, T-cell intracellular antigen 1 (TIA1), TIA-like 1 (TIAL1), TAR DNA-binding protein 43 (TDP-43), hnRNP L, Aub, hnRNP A1 and Scaffold Attachment Factor B1 (SAFB1) [156] [174] [175] [176] [177] [178] [179] [180].

Crosslinking and immunoprecipitation strategies coupled with NGS are the state of the art methodologies available to study protein-RNA interactions *in vivo*. The successful application of these methods to the study of Osk-RNA interactions might allow to identify both the RNA interactome of Osk and the binding site(s) of the protein on its target transcripts. These information would allow me to investigate both the molecular mechanism and the physiological

relevance of Osk RNA-binding activity, in the context of pole plasm assembly during *Drosophila* development.

2. Aims of the project

It was recently discovered that Osk protein associates with polyadenylated mRNAs in early *Drosophila* embryos (Fig. 7B) [109] [149]. Of the two globular domains harbored within Osk protein sequence, the C-terminal OSK domain shows RNA-binding activity *in vitro* [109]. Interestingly among the *osk* alleles that have been characterized over the years and that cause posterior patterning defects are two main classes of mutations: nonsense mutations harboring premature stop codons that lack the OSK domains, or missense point mutations that lie in the OSK domain itself [63] [114] [123]. Considering the RNA-binding activity of Osk, which most likely involves the C-terminal OSK domain, one can speculate that mutations that interfere with RNA binding strongly impair the biological function of Osk.

The pole plasm inducing activity of Osk relies on its ability to recruit other factors, proteins and RNAs, that are required for both germ cell specification and posterior patterning. More than 20 years of studies have revealed that Osk colocalizes with several mRNAs in the pole plasm [146] [147]. In addition, the Long Osk protein isoform is required to ensure the correct posterior anchoring of *osk* mRNA [89]. However, we still lack experimental evidence of a direct interaction between Osk and any mRNA in the pole plasm.

In order to understand in greater detail the molecular mechanism of action of Osk *in vivo*, it is important to identify which RNAs are directly bound by Osk and to study how and to what extent Osk RNA-binding activity affects its mRNA targets. The RNA-binding activity of Osk might underlie a possible regulatory function of different aspects of RNA biology, like localization and translational control, fundamental processes orchestrating pole plasm assembly and activity.

In conclusion it is important to answer some key questions regarding the putative RNA-binding activity of Osk protein *in vivo*:

- Identify the Osk RNA bound interactome *in vivo*.
- Identify Osk binding site(s) on its target RNAs.

- Generate transgenic flies harboring specific mutations in the target RNAs that interfere with Osk binding.
- Analyze which biological aspects of the targets RNAs are controlled by Osk *in vivo*.
- Study the physiological role of Osk RNA-binding activity in the context of posterior patterning and germ cell formation during early embryogenesis.

3. Results

3.1. Oskar CLIP

In order to identify the target mRNAs bound by Osk *in vivo*, I decided to apply the CLIP method [161] to study Osk-RNA interaction in *Drosophila* embryos. As described in the Introduction (Paragraph 1.8.1), UV light has been extensively used as a crosslinking agent for stabilizing protein-RNA interactions. UV light is a zero-length crosslinking agent that promotes the formation of covalent bonds between proteins and RNAs in direct contact [159] [160]. This feature is extremely useful for the identification of the direct RNA targets of an RBP because it reduces the likelihood of false positive results. These can be generated when using less stringent crosslinking procedures that allow covalent bond formation between the protein of interest and RNAs in close proximity even if not directly bound to one another. It has been shown that numerous mRNAs localize at the posterior pole of the *Drosophila* oocyte within the pole plasm where Osk protein accumulation occurs [146] [147]. It is therefore reasonable to speculate that some posteriorly localized mRNAs might be directly bound by Osk, while others might be indirectly recruited to the pole plasm by other factors. For this reason, UV crosslinking appeared to be the best strategy for capturing transient interactions between Osk and its target RNAs and avoiding capture of unspecific interactions that might merely reflect the accumulation of posteriorly localized mRNAs in the same subcellular compartment as Osk.

A UV crosslinking protocol was previously established in our laboratory to perform interactome capture experiments on 0 to 2 hour old *Drosophila* embryos [154]. As previously discussed (Paragraph 1.7) this approach was used to identify proteins bound to polyadenylated mRNAs *in vivo*, and Osk was detected among the *Drosophila*'s RBPs (Fig. 7B) [109] [149]. 0 to 2 hour old embryos contain Osk protein and possess a functional pole plasm. Therefore this particular developmental stage represents a good starting point for trying to identify which mRNAs are bound by Osk *in vivo*. The first step required in order to perform an Osk CLIP experiment was to set up the conditions for a clean Osk immunoprecipitation (IP).

3.1.1. Optimization of Oskar Immunoprecipitation conditions

During my first attempts to define optimal conditions for performing Osk IPs, I faced a major issue when trying to prepare embryo extracts under physiological conditions: after the centrifugation step necessary to remove debris and other particulate material from the soluble fraction, Osk protein was almost entirely present in the pellet fraction, rendering it inaccessible to IP (data not shown). This is most likely due to the fact that Osk is present in heavy RNPs that tend to sediment with the pellet fraction when lysates are prepared using conditions that do not promote RNP solubilization.

The lysis conditions used in the interactome capture method were stringent enough to allow the recovery of a good amount of Osk protein in the soluble fraction (Fig. 7B). Indeed, this lysis buffer contains high salt (500mM LiCl) and high detergent (0.5% LiDS) concentrations [109] [148]. I used these conditions as a starting point for the Osk IP protocol optimization. Initially I tested several anti-Osk antibodies available in the lab and found that only a rabbit polyclonal antibody directed against the C-terminal half of Osk protein worked in the IPs [89]. By using this specific antibody and the lysis conditions from the interactome capture protocol, it was possible to IP Osk from embryo extracts, although the amount of protein recovered was low (Fig. 8). I tested several conditions, changing types and amounts of detergents in the lysis buffer, trying to find conditions that would yield the highest amount of Osk protein in solution, while allowing good antigen-antibody interaction. In this regard the addition of sodium deoxycholate (Na deoxycholate) to the embryo lysates (after preparation of the extracts) strongly improved the amount of protein collected in the IP. This was very striking in the case of the Long Osk isoform: indeed, Long Osk was barely detectable in the input and IP samples in the absence of Na deoxycholate, but was strongly enriched after addition of the detergent. The protocol was tested on both UV crosslinked and non-crosslinked extracts; this revealed that the treatment did not interfere with protein recovery (Fig. 8).

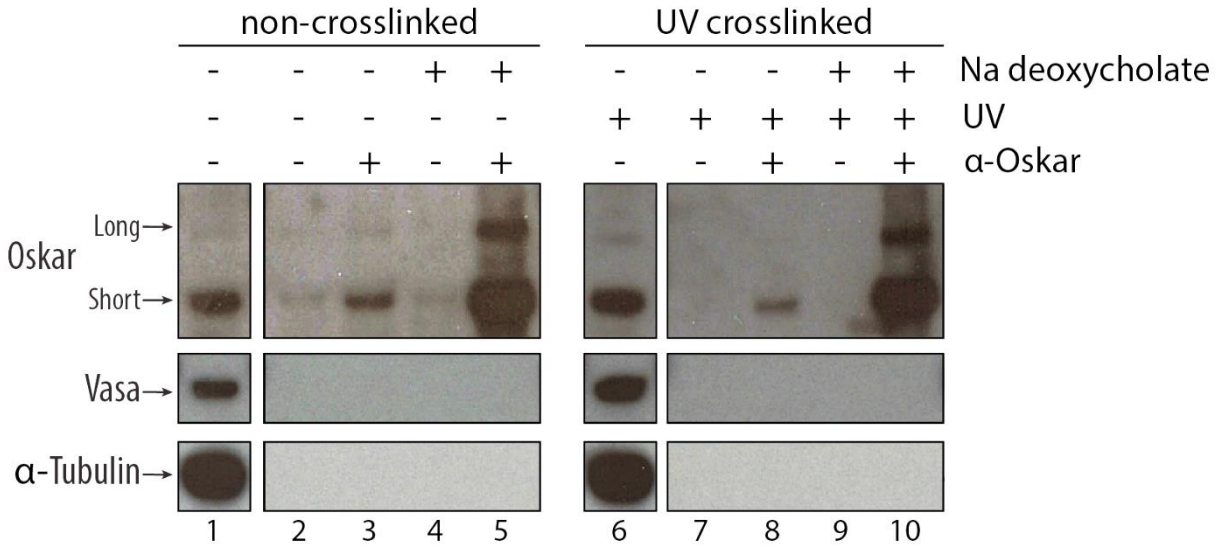

Figure 8: Comparison of Oskar IP conditions in presence or absence of Na deoxycholate. *Drosophila* embryos collected between 0 and 2 hours after egg laying were either treated (UV +) or not (UV -) with UV light before lysis. The extracts were diluted using lysis buffer containing (Na deoxycholate +) or not (Na deoxycholate -) 0.5% Na deoxycholate and subsequently incubated with protein G sepharose beads conjugated with an anti-Osk antibody (α-Oskar +) or no antibody (α-Oskar -). After washings, proteins were eluted in 2X NuPage LDS sample buffer and analyzed on western blot for: Osk, Vas and α-Tubulin. Lane 1: non-crosslinked input. Lane 6: UV crosslinked input.

The IP conditions I established are stringent enough to avoid the co-IP of Vas, a well-known interacting partner of Osk *in vivo* [113] [115] and α-Tubulin, a highly abundant protein. Vas is an RNA helicase and as such might represent a source of contamination in the CLIP experiment. No Vas signal was detected in either the UV crosslinked or non-crosslinked IP samples, confirming the stringency of the conditions used for the IP. For these experiments, a large amount of starting material was involved: indeed, for each sample I used 5ml of embryo lysate prepared from 0.5g of 0 to 2 hour old embryos.

The possibility to IP Osk from embryo extracts in a clean manner represented a good starting point for the identification of RNAs bound *in vivo* using CLIP based methodologies. The fact that using Na deoxycholate increased Osk IP efficiency is interesting, as this particular detergent is used for solubilizing and isolating membrane associated proteins [181]. It has been reported that Long Osk associates with endosomes and other membranous structures within the pole plasm, and that Long Osk is involved in anchoring Short Osk to the posterior

pole of the oocyte [89]. In this regard Na deoxycholate, by solubilizing membranes, might help get Osk protein into solution, making it more accessible to antibodies in the IP.

3.1.2. Oskar associates with *nos*, *pgc* and *gcl* mRNAs *in vivo*

In order to identify RNAs bound by Osk *in vivo*, I adapted the CLIP method to *Drosophila* early embryos. In this experiment 0 to 2 hour old wild-type embryos were irradiated with UV light (254nm); as a control I used embryos that were not UV irradiated. After lysis and clarification of the lysates, 5ml of embryo extract (prepared from 0.5g of embryos) were incubated with beads conjugated with a specific anti-Osk antibody. After stringent washing steps, the proteins retained on the beads were digested by proteinase K, the RNAs extracted, reverse-transcribed and analyzed by quantitative real-time PCR (qPCR) (Fig. 9A). For both the UV crosslinked and non-crosslinked Osk IP, I performed a parallel IP in which I omitted the anti-Osk antibody (beads-only). The beads-only condition would allow me to measure background levels of RNA that co-IP with the beads in both UV crosslinked and non-crosslinked conditions.

Since Osk protein is produced and accumulates exclusively at the posterior pole, I hypothesized that Osk binding targets might include known posteriorly localized mRNAs. Therefore I took a candidate approach and tested the samples for enrichment of a set of mRNAs by performing qPCR analysis. I chose 7 posteriorly localized mRNAs: *nos*, *orb*, *cycB*, *pgc*, *gcl*, *osk* and *Heat-shock protein 83* (*Hsp83*) mRNAs. As negative controls, I chose *bcd*, an anterior localized mRNA, and *Actin42A* (*Act42A*) mRNA, which is evenly distributed in the embryo [146] (Fig. 9B). Among the posteriorly localized mRNAs chosen as candidate targets, several genetic studies have shown that *nos*, *cycB*, *pgc*, *gcl*, *osk* and *Hsp83* are dependent on Osk function for their localization at the posterior pole of the oocyte and embryo [49] [89] [137] [141] [182] [183].

The results of the experiment are shown in Figure 9B. For each target mRNA tested, the level of enrichment in the Osk IP samples was normalized with respect to that of the beads-only controls, in which the specific anti-Osk antibody was omitted. Among the mRNAs tested, I detected a significant UV- and Osk-dependent enrichment of *nos*, *pgc* and *gcl* mRNAs. The other posteriorly localized mRNAs, *orb*, *cycB*, *osk* and *Hsp83*, as well as the *bcd* and *Act42A* control mRNAs, were not significantly enriched upon UV crosslinking (Fig. 9B).

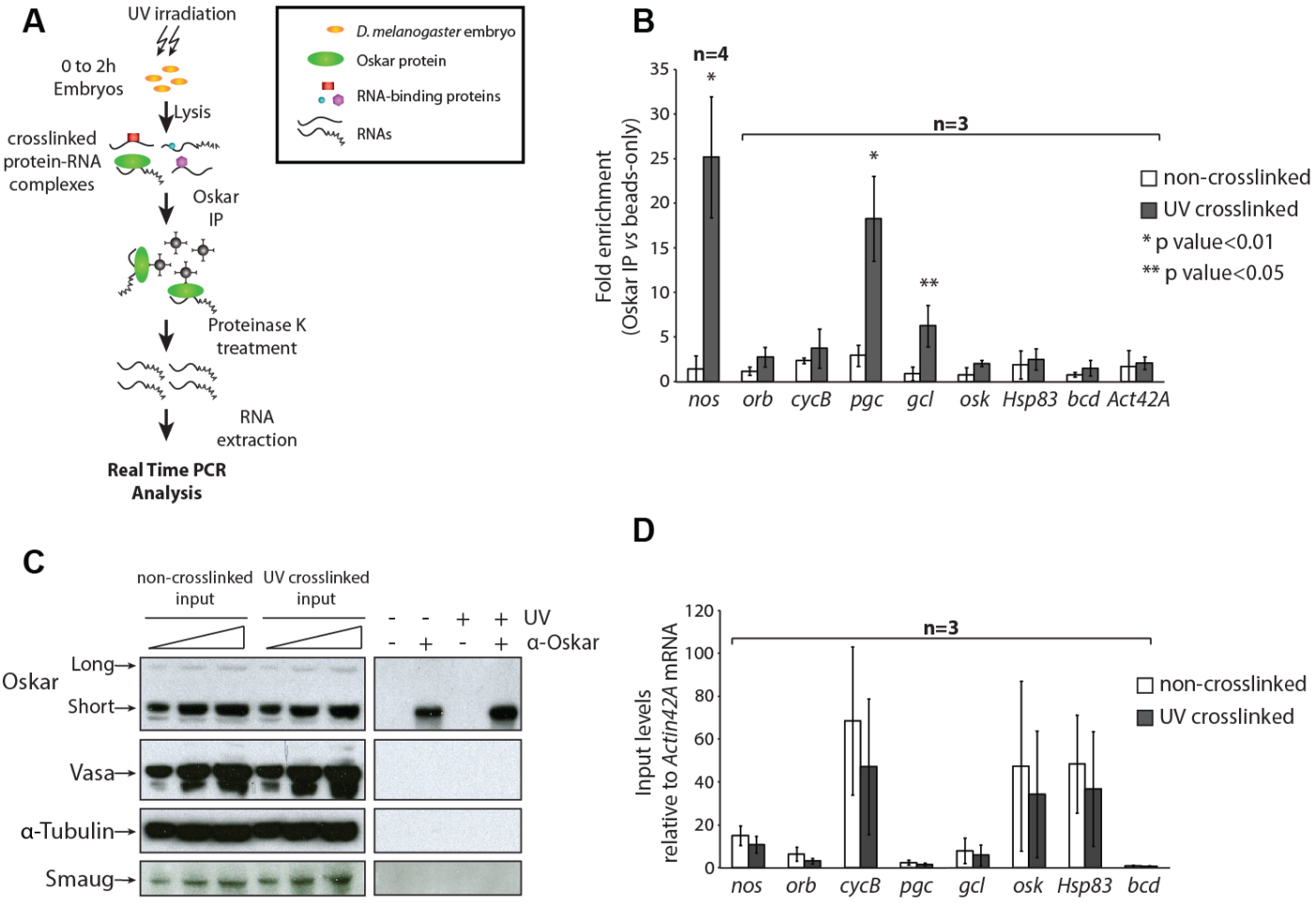

Figure 9: Oskar associates with *nos*, *pgc* and *gcl* mRNAs *in vivo*. (**A**). Schematic representation of the Osk CLIP experimental procedure. Lysates are prepared from 0 to 2 hour old *Drosophila* embryos treated with UV light and Osk IP is performed. After stringent washing steps the proteins are digested by proteinase K treatment and RNAs are extracted. After reverse-transcription the cDNAs are analyzed by qPCR. (**B**). Enrichment levels of a set of mRNAs posteriorly localized in *Drosophila* embryos: *nos*, *orb*, *cycB*, *pgc*, *gcl*, *Hsp83*. *bcd*, an anteriorly localized mRNA and *Act42A*, an evenly distributed mRNA, were used as negative controls. For both UV crosslinked (UV +) and non-crosslinked (UV -) conditions the enrichment levels in the Osk IP samples (α-Oskar +) were normalized respect to those in control samples in which the anti-Osk antibody was omitted (α-Oskar -). Bars

represent mean values ± standard deviation. p values were calculated using Student's t test. (**C**). Western blot analysis of the Osk CLIP experiment. Proteins were eluted from 1% of the beads used in 2X NuPage LDS sample buffer and analyzed on western blot for: Osk, Vas, α-Tubulin and Smg. (**D**). Relative expression levels of the mRNAs tested in B measured in the input embryo extracts. For the different transcripts the levels were normalized with respect to *Actin42A* mRNA. Bars represent mean values ± standard deviation.

Adapted from [109].

The stringency of the conditions used in the IP prevented the co-IP of known RBPs such as Vas and Smg (Fig. 9C). Both proteins co-purify with Osk under physiological conditions and could have been a source of contamination in the experiment due to their RNA-binding activity [60] [113] [115]. In addition, I analyzed the relative expression level of each candidate mRNA in the input material, normalizing its amount with respect to the housekeeping *Actin42A* mRNA. This allowed me to compare the relative abundance of each transcript in the inputs with the enrichment measured for each of them in the CLIP experiment (Fig. 9D). This analysis showed that there was no correlation between the relative abundance of the mRNAs in the input and their enrichment in the CLIP. Altogether these data revealed that endogenous Osk protein associates with *nos*, *pgc* and *gcl* mRNAs in *Drosophila* early embryos [109].

The three mRNAs that were specifically enriched after Osk CLIP, *nos*, *pgc* and *gcl,* are involved in germline specification. As previously mentioned, in addition to encoding the abdominal determinant of the fly, *nos* mRNA is also required for germ cell specification. Specifically, Nos activity is required in PGCs, where it blocks mitosis and silences zygotic gene expression before the PGCs colonize the future gonads of the animal [48] [133] [134] [135] [136]. *pgc* and *gcl* gene products are required for proper germ cell formation in *Drosophila* [137] [138] [141]. Proteins produced by both genes were shown to be involved in transcriptional repression in PGCs, which is required to prevent their premature differentiation prior to reaching the somatic portion of the gonads in the early embryo [138] [145]. The molecular function of the three mRNAs identified in the Osk CLIP experiment is in line with the molecular function of Osk. Osk is required for both abdomen and germline formation and consistent with this function it can be UV crosslinked *in vivo* with mRNAs required for such biological functions.

In addition, it has been shown that all three mRNAs (*nos*, *pgc* and *gcl* mRNAs) require Osk activity for their accumulation at the posterior pole [49] [137] [141] and that Osk activity is required for translational activation of *nos* mRNA [59] [60]. In contrast, *cycB*, *osk* and *Hsp83* mRNAs, which accumulate at the posterior in an Osk-dependent manner [89] [182] [183], did not associate with Osk. Similarly, *orb* mRNA, whose posterior localization in the oocyte is *osk*-independent (embryo data not available) [184] [185], was not enriched in the IPs (Fig. 9B).

Altogether these data indicate that Osk binds *nos*, *pgc* and *gcl* transcripts *in vivo*. Although the ability to detect these three particular mRNAs in Osk CLIP experiments could be explained by their localization in the same subcellular compartment as Osk, this seems unlikely as other posteriorly localized mRNAs were not UV crosslinked to Osk. These experiments suggest that, within the crowded pole plasm, Osk associates specifically with mRNAs required for abdomen and germline formation, and not with other pole plasm mRNAs.

It is possible that Osk might interact with additional mRNAs not tested in the experiment. Furthermore, Osk might be involved in regulating some of its target mRNAs, for instance promoting their translation or their posterior accumulation. By using a transcriptome-wide unbiased approach it would be possible to identify all the binding targets of Osk *in vivo* and in addition the binding site(s) of the protein on the target transcripts. This information would allow me to design perturbation experiments to address the biological significance of Osk RNA-binding activity *in vivo*.

3.2. Oskar iCLIP

To identify and analyze in an unbiased manner the Osk RNA interactome *in vivo* I chose to apply the iCLIP methodology. As previously mentioned, iCLIP allows one to identify all the target mRNAs of a given RBP on a transcriptome-wide level and, additionally, yields information regarding the specific sites to which the protein of interest is bound [174]. In particular, the identification of the precise sequences or structures in target RNAs to which Osk

binds, would guide the design of site-specific mutations that would allow me to test whether Osk is a genuine RBP, as well as mechanistic analyses regarding the role of Osk RNA-binding activity in e.g. localization, translation and stability of the target mRNAs.

3.2.1. Overview of the Oskar iCLIP protocol

The iCLIP protocol consists of the following sequence of complex biochemical reactions [186].

- ***in vivo* UV crosslinking**

The iCLIP method relies on the ability to induce the formation of covalent bonds between proteins and RNAs upon UV light irradiation [159] [160]. I identified *nos*, *pgc* and *gcl* as Osk associated mRNAs in CLIP experiments performed on UV irradiated *Drosophila* embryos collected between 0 to 2 hours after egg laying (Fig. 9). For this reason I decided to perform the Osk iCLIP experiment using the same approach (Fig. 10, steps 1 and 2).

- **Osk IP**

The next step of the iCLIP protocol consists of the immunoprecipitation of Osk using a specific anti-Osk antibody (Fig. 10, step 3). Osk-RNA complexes are purified using the same conditions developed for the previously described Osk CLIP experiment (Fig. 9). The stringency of the IP protocol I developed prevents co-IP of known RBPs that interact with Osk such as Vas and Smg (Fig. 9C).

- **Partial RNA digestion**

A partial RNase I-mediated digestion step is performed to reduce the length of crosslinked RNAs, ideally in the range between 50 and 300 nucleotides (Fig. 10, step 4).

- **3’ end dephosphorylation**

After the RNase treatment, a residual phosphate group at the 3’ end of the RNA fragments must be removed, as L3App adapter ligation requires that the RNA substrates bear a terminal

3'-OH group. The dephosphorylation step is catalyzed by polynucleotide kinase (PNK) (Fig. 10, step 5).

- **L3App adapter ligation**

A short (22 nucleotide long) DNA adapter (L3App adapter) is ligated at the 3' end of the RNA fragments in a reaction catalyzed by a truncated version of T4 RNA ligase (T4 RNA ligase 2 truncated) (Fig. 10, step 6). The L3App adapter bears two nucleotide modifications. At the 5' end the adapter is pre-adenylated. Since T4 RNA ligase 2 truncated enzyme cannot catalyze the adenylation step required for the ligation [187], pre-adenylated L3App adapters are used as ligation substrates. A second modification at the 3' end of the L3App adapter bears a dideoxycytidine nucleotide that lacks a 3'-OH group. This feature prevents the ligation of multiple adapters on the same RNA molecule.

- **RNA-L3App extraction**

After L3App adapter ligation, the RNPs are treated with proteinase K to remove the protein components (Fig. 10, step 7), and RNA-L3App products are extracted (Fig 10, step 8). It is important to note that proteinase K digestion always leaves a short peptide at the UV crosslinking site [188].

- **Reverse-transcription**

Extracted RNAs are reverse-transcribed using primers complementary to the L3App adapter sequence (Fig. 10, step 9). The primers also contain short barcodes that allow one to distinguish the products of the different experimental conditions. The residual peptides attached to RNAs act as road blocks for enzymes (such as reverse-transcriptases) that move processively on the RNA templates. Consequently, cDNA synthesis stops one nucleotide before the site of protein crosslinking in the vast majority of the cases and this feature allows one to map the binding site of an RBP on its target transcript [174] [188].

- **Circularization**

cDNA fragments are circularized by using a specific single stranded DNA ligase (CircLigase II) (Fig. 10, step 10).

- **Linearization**

Following circularization, cDNAs are re-linearized using a DNA oligo that contains a BamHI cleavable site (Fig. 10, step 11).

- **PCR amplification and sequencing**

The resulting linear cDNAs are amplified by PCR to obtain libraries suitable for high throughput Solexa sequencing (Fig. 10, step 12). In the reads, the first nucleotide sequenced after the experiment-specific barcode corresponds to the nucleotide that precedes a UV crosslink on the RNA (Fig. 10), allowing one to identify with precision the site where an RBP is bound to the target RNA [174] [186].

- **Final considerations**

A revised and simplified version of the iCLIP method originally developed by Dr. Julian Konig (Institute of Molecular Biology, Mainz), was recently established and applied with success in our laboratory by Dr. Ales Obrdlik (unpublished data). The major differences between the two iCLIP protocols lie in the procedure for size selection during library preparation.

In the classical iCLIP method, after L3App adapter ligation (Fig. 10, step 6), the RNA components of the RNP complexes are labeled with radioactive phosphate groups, separated on a polyacrylamide gel and blotted onto a nitrocellulose membrane. The radioactive signal allows one to detect the protein crosslinked to RNA fragments and cut out the portion of membrane bearing those of the desired length (between 50 and 300 nucleotides). These fragments are extracted and used as substrates for reverse-transcription. An additional extraction step is performed after reverse-transcription to select cDNA fragments of the desired size to be PCR amplified [186].

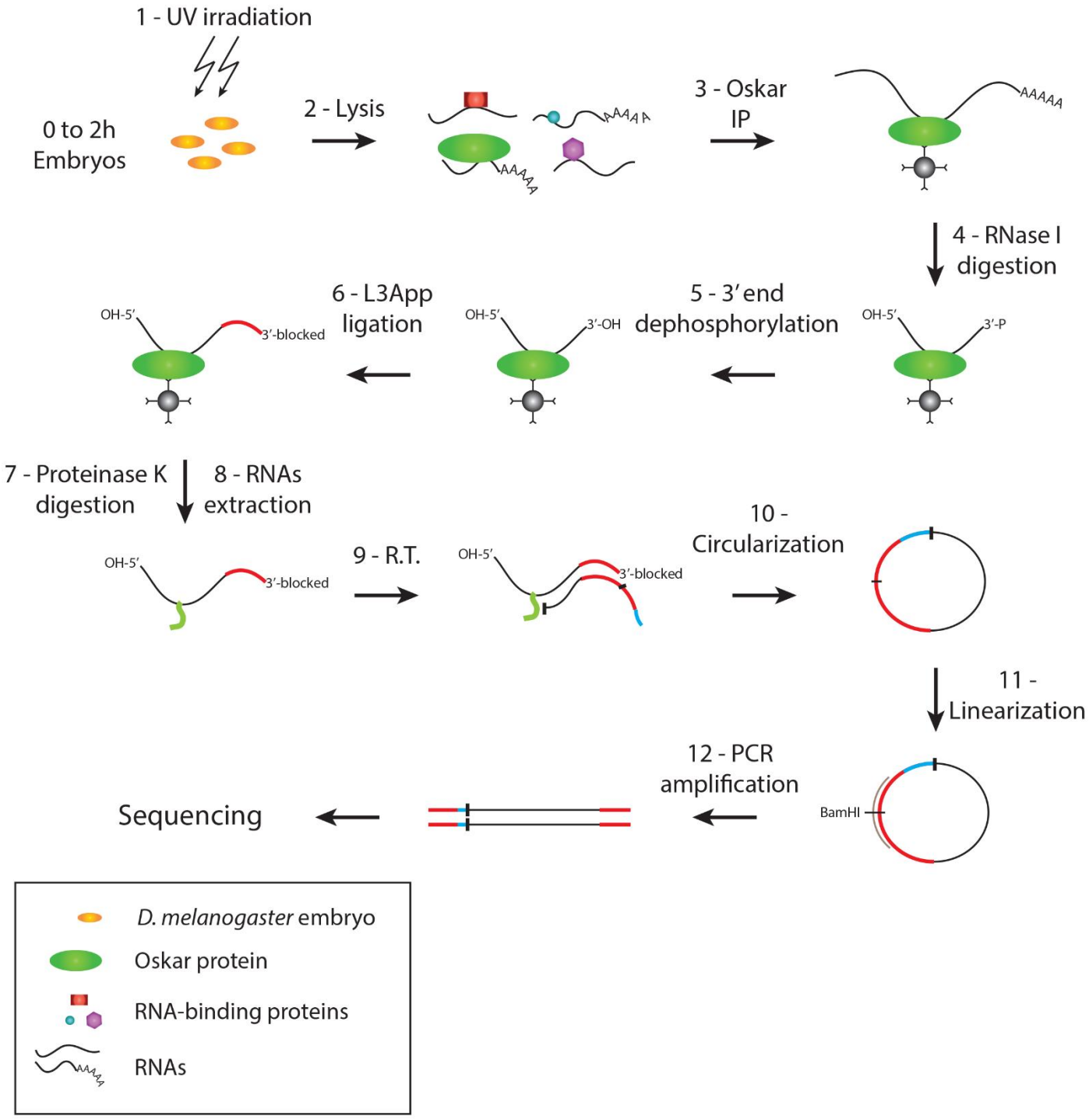

Figure 10: Oskar iCLIP protocol overview. After UV crosslinking of *Drosophila* embryos, Osk protein is immunoprecipitated using anti-Osk specific antibodies. After stringent washing steps, a partial RNase I-mediated digestion is performed and a DNA adapter (L3App adapter) is ligated at the 3' end of the RNA fragments covalently bound to the protein. Proteins are digested with proteinase K and RNAs extracted. The recovered RNAs are reverse-transcribed (R.T.), circularized, and then re-linearized by hybridizing with a DNA oligo encoding a BamHI recognition and cleavable site. Linear cDNA fragments are amplified by PCR. The obtained libraries are analyzed via Illumina sequencing.

In the protocol that was established in our lab, these extractions steps are omitted thus shortening considerably the duration of the experiment. Size selection of cDNA libraries is performed at the very end of the protocol: after PCR amplification, the libraries are run on a 4% agarose gel and DNA fragments of the desired size are extracted and purified prior to sequencing. Considering the simplifications introduced in the iCLIP protocol by Dr. Obrdlik and the opportunity that I had to take advantage of his experience, I decided to apply the iCLIP method developed in our laboratory for studying Osk-RNA interaction *in vivo*.

3.2.2. Oskar iCLIP preparative experiments

The first step before starting with the library preparation process was to check the amount and quality of RNAs extracted after Osk CLIP. The experiment consisted of four samples. For both UV crosslinked and non-crosslinked conditions I performed an Osk-specific IP and a beads-only control without the anti-Osk antibody. The beads-only conditions would allow me to control for RNA stickiness to the matrix used in the experiment (protein G sepharose). The non-crosslinked Osk IP would allow me to control for RNAs that co-precipitate with Osk in absence of crosslinking, while the UV crosslinked Osk IP would allow me to measure which RNA species, upon UV light irradiation, are enriched compared to the control conditions.

I performed the experimental procedure described in Figure 9A. To analyze the quality of the extracted material, after extraction and heat denaturation (70ºC for 2'), I analyzed the RNAs on an Agilent Bioanalyzer 2100 instrument (Fig. 11). The profile of total RNA extracted from non-crosslinked and UV crosslinked embryo extracts respectively is shown in Figure 11 (panels A and B). The two major peaks, which cluster at around 2000 nucleotides, represent 18S and 28S ribosomal RNAs (rRNA).

The profiles of the RNAs present in the input samples (Fig. 11A and B) appear different from total RNA extracted from other eukaryotic organisms, for instance human cell lines. In the latter, 18S rRNA migrates around 2000 nucleotides while 28S rRNA migrates around 4000

nucleotides. In *Drosophila* and in other insect species, the 28S rRNA is processed into two fragments of a size similar to that of 18S rRNA [189] [190]. Knowing this intrinsic feature of insect rRNAs allowed me to judge the quality of the total RNA extracted from the input material: in both cases (non-crosslinked and UV crosslinked), the overall quality of the RNA looked acceptable, with no evident signs of degradation (Fig. 11A and B). Therefore I considered that the protein-RNA complexes in the input material were protected from degradation, and that the extract preparation procedure was suitable for an Osk iCLIP experiment.

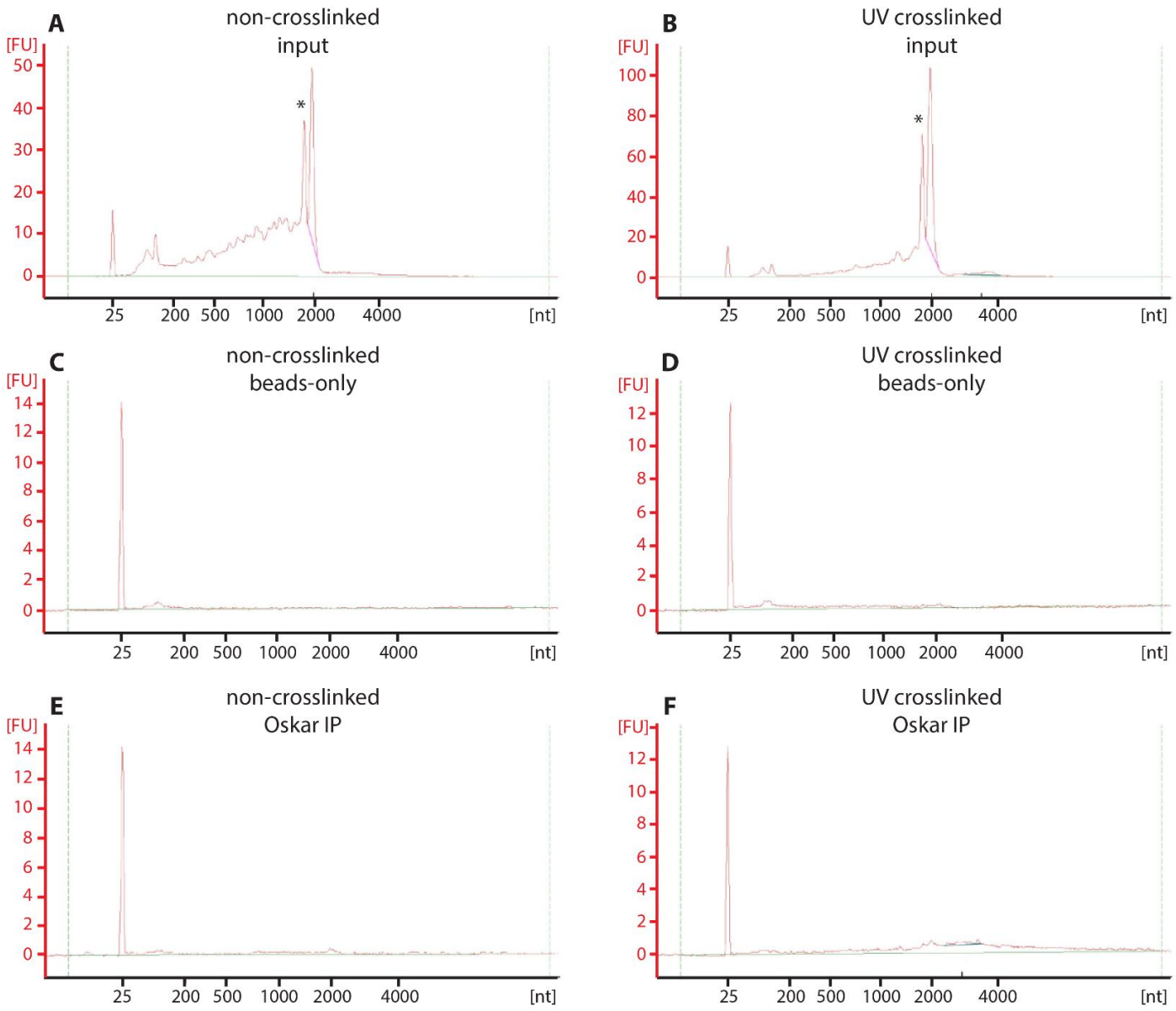

Figure 11: Bioanalyzer profiles of RNA extracted from an Oskar CLIP experiment. (**A** and **B**). Total RNA extracted from non-crosslinked and UV crosslinked input embryo extracts. For both conditions 5ng of total RNA were loaded into the machine. The asterisks (*) indicate the 18S and 28S

rRNA species. (**C**, **D**, **E** and **F**). Total RNA extracted from: non-crosslinked beads-only, UV crosslinked beads-only, non-crosslinked Osk IP and UV crosslinked Osk IP samples respectively. After extraction, RNA pellets were resuspended in 10µl of H_2O, 1µl has been loaded into the machine. On the Y axis, FU represents arbitrary fluorescent units measured by the instrument; on the X axis all samples were aligned respect to an RNA ladder (nt: nucleotides).

I also analyzed the quality of RNA in the three Osk iCLIP experimental controls (Fig. 11C, D and E): non-crosslinked beads-only, UV crosslinked beads-only, and non-crosslinked Osk IP. All three samples from the controls appeared extremely clean: little or no rRNA was detected, nor were other RNA species (Fig. 11C, D and E). In contrast, the profile of RNA extracted under the UV crosslinked Osk IP condition showed a slight enrichment for RNA species ranging from 1500 to more than 4000 nucleotides in length (Fig. 11F). According to the measurements of the Bioanalyzer instrument, the amount of RNA recovered after Osk IP from extracts of UV crosslinked embryos, was ~1ng, approximately 10 times more than what was recovered in the three controls.

The analysis of these results revealed both positive and negative aspects of the experiment. Among the positive aspects was the good quality of the input material, which showed no sign of RNA degradation. Also the cleanliness of the three control conditions represented a promising aspect of the experiment, as less background signal would help the identification of the binding targets of Osk. Of great concern, however, was the very small amount of RNA recovered after Osk IP from UV crosslinked embryo extracts, as iCLIP library preparation from ~1ng of total RNA is considered sub-optimal and at the limit of feasibility. I therefore attempted to scale up the IP protocol by a factor of 10. This, on one hand, helped to increase the amount of RNA recovered, but on the other hand increased the rRNA contamination in all four experimental conditions (data not shown). Therefore, all things considered, especially the reproducible enrichment of *nos*, *pgc* and *gcl* mRNAs I had measured by qPCR in my Osk CLIP pilot experiments (Fig. 9), and considering the high sensitivity of NGS detection methods that might circumvent the limitations of the experimental setup, I decided to apply those conditions and proceed to Osk iCLIP library preparation, in spite of the small amount of RNA in my samples.

A critical step in the preparation of iCLIP libraries is the definition of optimal conditions for the RNase I digestion step. As mentioned in the Paragraph 3.2.1, this is done by radioactive labeling of protein-RNA complexes following the IP and RNase digestion steps [186] [191]. Since the iCLIP protocol established in the lab does not include performing this step during the library preparation, it was important to include a preparative experiment to determine the amount of RNase needed to produce RNA fragments of the desired length.

In this preparative experiment, I performed the Osk IP as previously described, on both UV crosslinked and non-crosslinked conditions. At the end of the IP, the beads were resuspended in 1ml of Washing Buffer 3 (see Materials and Methods), and 10µl of RNase I at increasing dilutions were added to different Osk IPs from UV crosslinked extracts. Samples were incubated 3' at 37ºC followed by three additional washing steps. Then protein-RNA complexes were labeled by PNK-mediated addition of radioactive ^{32}P containing phosphate groups at the 5' end of RNA fragments, separated by sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. The membrane was first exposed to a phosphor imaging plate and subsequently used for Osk protein detection by western blot (Fig. 12). A particular pattern of radioactive signal was expected as a consequence of the decreasing RNase I amounts used in the experiment. When high concentrations of RNase I are used, RNA fragments crosslinked to Osk should be short and most of the radioactivity should concentrate at the same position as the protein, thus overlapping with the Osk signal on the western blot. Decreasing concentrations of RNase I result in spreading of the radioactive signal toward the upper part of the membrane, reflecting the increasing length of RNAs attached to Osk.

Figure 12 shows the result of such an experiment. Comparison of the radioactive signal and the Osk protein signal on the western blot shows that, in the absence of UV crosslinking, Osk was not radioactively labeled (lanes 4 and 5). Conversely a faint radioactive signal that overlaps perfectly with Osk is detected in a crosslinking-dependent manner (lanes 8, 9, 10, 11 and 12). In all 5 lanes, most of the radioactive signal is equally detected in the upper part of

the membrane, above the faint band overlapping with Osk, regardless of the amount of RNase I used in the assay (Fig. 12). This is unfortunately due to the incompatibility of RNase I with the Washing Buffer 3: in this buffer the activity of the enzyme was completely blocked. Nevertheless, the UV-dependent labeling of Osk suggested that the protein can be crosslinked to RNAs *in vivo*. I repeated this experiment several times using an RNase I compatible buffer, yet I was never able to detect radioactively labeled Osk-RNA complexes and as a consequence I was not able to identify the exact amount of RNase I to be used for achieving the partial RNA digestion required for iCLIP library preparation.

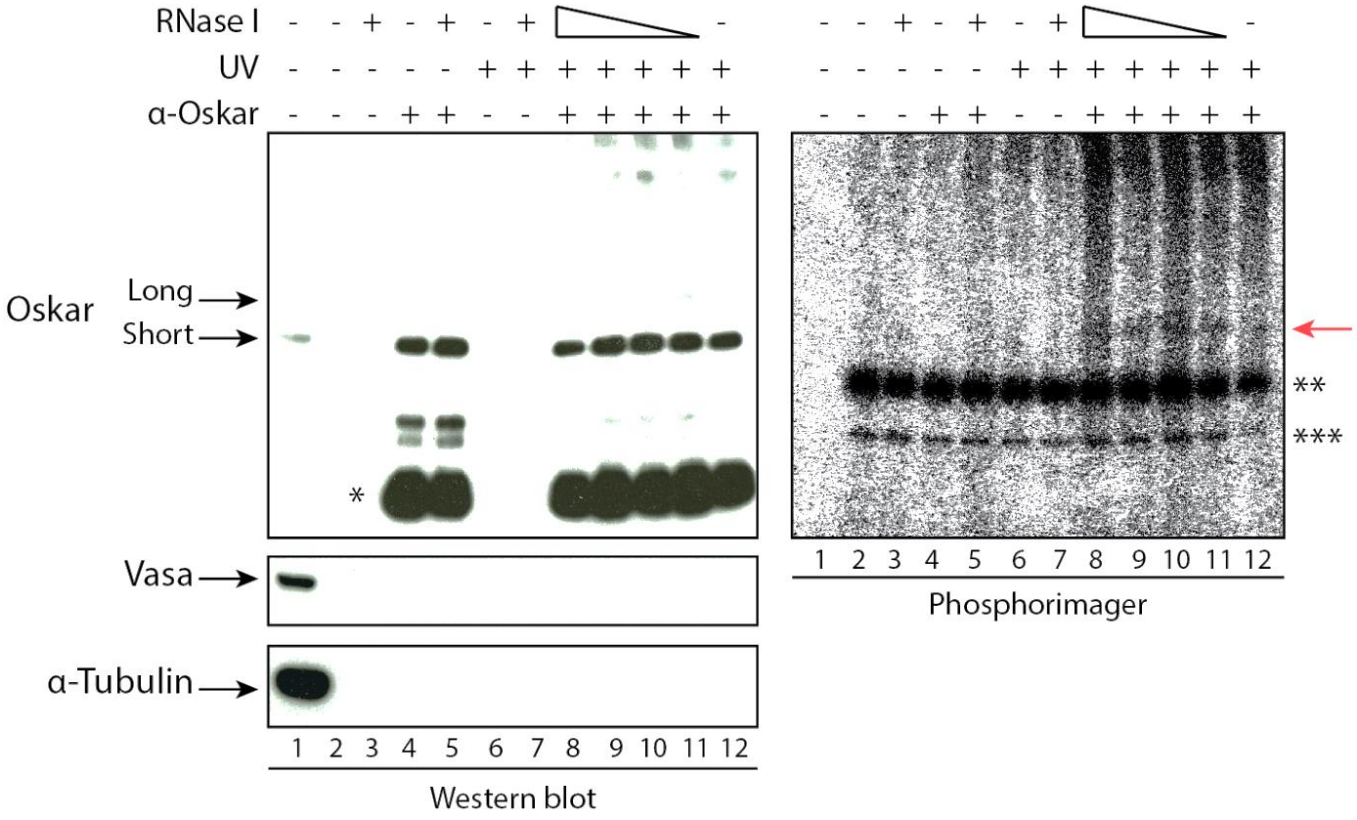

Figure 12: Oskar IP and radioactive-dependent detection of Oskar-RNA complexes. Osk IP (α-Oskar +) was performed using extracts prepared from 0 to 2 hour old embryos either UV irradiated (UV +) or not (UV -). For both conditions a parallel IP omitting the anti-Osk antibody (α-Oskar -) has been performed. At the end of the IP, beads were resuspended in 1ml of Washing Buffer 3 and 10µl of RNase I (RNase +) at different dilutions were added. For samples 3, 5, 7 and 8 a 1:50 dilution was used, for samples 9, 10 and 11 1:500, 1:1000 and 1:2000 dilutions were added, respectively. Samples were incubated for 3' at 37ºC. In parallel the same samples were incubated without RNase I (RNase I -). RNA fragments were labeled with radioactive phosphate groups in a PNK-mediated reaction. Protein-RNA complexes were separated by SDS-PAGE and transferred to nitrocellulose membrane. The membrane was exposed to a phosphor imaging plate and subsequently analyzed by western blot for: Osk, Vas and

α-Tubulin. (*) IgG light chain. (**) RiboLock RNase inhibitor. (***) PNK. The red arrow indicates radioactively labeled Osk protein.

The inability to reproducibly detect Osk-RNA complexes using the RNA labeling assay prevented me from defining the conditions to be used for the partial RNase I-mediated digestion in the iCLIP protocol. In addition, these negative results led me to question the ability of Osk to bind RNAs *in vivo*. Nevertheless, the results I had reproducibly obtained from the Osk CLIP experiment (Fig. 9) convinced me to attempt iCLIP library preparation.

As a last preparatory experiment, I checked the stability of the interaction between Osk and the anti-Osk antibody throughout the entire protocol, up to RNA extraction. Indeed, it was possible that the different enzymatic reactions that have to be performed directly on beads, in the presence of relatively high concentrations of dithiothreitol (DTT) (10mM), polyethylene glycol (PEG) 400 (20%) and at relatively high temperature (37°C) [186], might interfere with the antigen-antibody interaction. By western blot analysis I showed that the Osk protein is retained on the beads until the end of the protocol (data not shown).

3.2.3. Oskar iCLIP libraries preparation

Knowing that the experimental conditions that I set up for the Osk CLIP (Fig. 9) work throughout the different steps of the iCLIP method, I proceeded to perform the entire Osk iCLIP experiment. As I was not able to identify the optimal RNase I amount required for correct trimming of RNA fragments crosslinked to Osk, I prepared libraries from samples treated with different RNase I concentrations. I was able to generate Osk iCLIP libraries by using an RNase I amount of $1*10^{-3}$U/µl. Figure 13 shows the results of the Osk iCLIP library preparation and analysis.

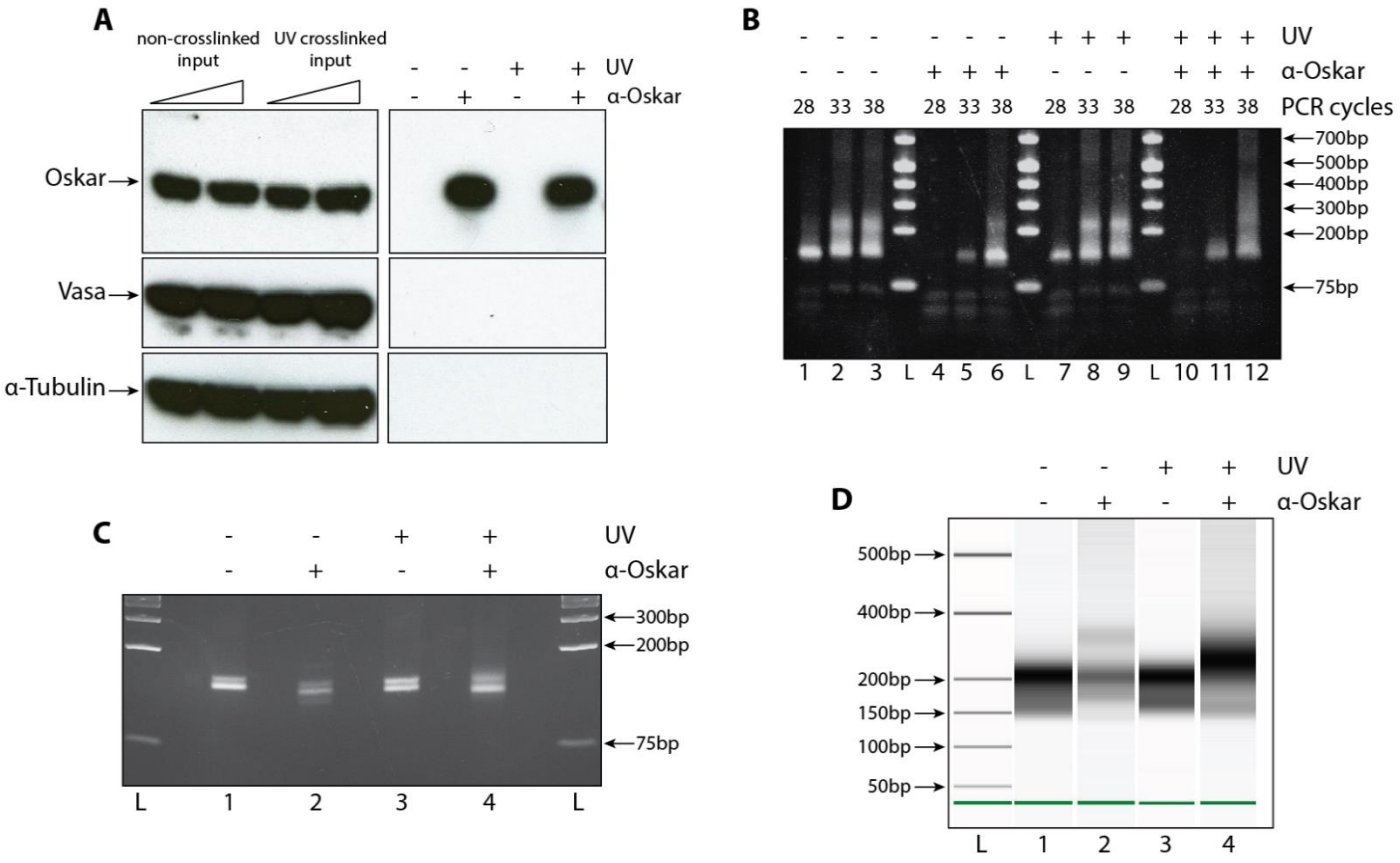

Figure 13: Oskar iCLIP libraries preparation. (**A**). 0 to 2 hour old embryos were either UV irradiated (UV +) or not (UV -) before lysis and extract preparation. For both conditions, an Osk specific IP (α-Oskar +) and a parallel IP omitting the anti-Osk antibody (α-Oskar -) were performed. After stringent washing steps 1% of the beads were subjected to protein elution. Protein samples were separated by SDS-PAGE and analyzed by western blot for: Osk, Vas and α-Tubulin. (**B**). iCLIP libraries prepared for all four experimental conditions were PCR amplified and analyzed on a 6% polyacrylamide native gel in TBE buffer. L: DNA ladder (**C**). iCLIP libraries amplified with 38 PCR cycles in B (lanes: 3, 6, 9 and 12) were separated on a 4% agarose gel for extraction. L: DNA ladder (**D**). Bioanalyzer profiles of the PCR fragments extracted from C. L: DNA ladder.

As previously described, at the end of the iCLIP protocol the cDNA libraries are amplified by PCR using specific primers that allow insertion of specific sequences required for Illumina sequencing. In order to assess the quality of the libraries, the PCR products were separated on a 6% native polyacrylamide gel. Panel B in Figure 13 shows the test PCR for the four experimental samples, each amplified using different PCR cycles (28, 33 and 38). A single sharp band migrating around 150 base pair (bp) (precisely 127bp) is present in all samples. In the two beads-only samples, the band is already detectable at 28 PCR cycles (Fig. 13B, lanes 1, 2, 3, 7, 8, and 9) while in the two Osk IPs it starts to accumulate at 33 cycles (Fig. 13B, lanes

4, 5, 6, 10, 11 and 12). This band is produced by the unligated L3App adapter, as a consequence of the high amount of the oligo used during the ligation step (60 pmoles) that is not completely removed before RNA extraction. This product is commonly present in the libraries prepared using Dr. Obrdlik's protocol and does not represent a major contamination, as it can be easily removed by gel extraction. An additional band can be detected in all samples running slightly above the 200bp band of the DNA ladder, starting from 33 PCR cycles in the beads-only samples (Fig. 13B, lanes 2, 3, 8 and 9) and from 38 cycles in the Osk IP samples (Fig. 13B, lanes 6 and 12). It is possible that this band is the result of multiple ligation events between different L3App adapter molecules. Although unlikely, this might happen when high amounts of free adapter molecules during the ligation step are present together with very small amounts of RNA substrate. This is likely to occur, especially in the three control samples, from which only a very low amount of RNA can be extracted (Fig. 11C, D and E), but it might also apply to the UV crosslinked Osk IP sample (Fig. 11F). Indeed, not all the L3App adapter molecules contain the two modifications described in the Paragraph 3.2.1: only 20% of them according to information provided by the manufacturer. As a consequence, those oligo that harbor a pre-adenylated 5' end and do not contain a dideoxycytidine at the 3' end might form concatamers as result of multiple ligation events.

Interestingly, in the UV crosslinked Osk IP sample, after 38 PCR cycles one can detect a smeary signal starting above the band produced by the L3App adapter alone, up to the 400bp band of the DNA ladder (Fig. 13B, lane 12). This particular signal is absent in the control samples, possibly indicating the existence of cDNA produced from RNA fragments crosslinked to Osk. Even if this finding represented a positive indication that the experiment might be working, some major issues arose from the analysis of the experiment.

The first concern regards the libraries produced from the control samples, which showed considerable background, putting the entire experiment at risk. The second problem was the high number of cycles required for the library amplification (38); this indicates that the starting amount of RNA was very low and therefore that the libraries were likely to be strongly PCR

biased, in other words, that many reads from the sequencing would be PCR duplicates generated by over-amplification of the same substrate. The third issue concerned the band running around 200bp. In principle, all the fragments running above the 127bp free adapter band might be the result of ligation of the L3App adapter to RNA fragments, and thus might be informative. The 127bp band can be removed during gel extraction and the rest of the library could be sequenced. But the presence of material corresponding to the second, 200bp band, which represented a ligation artifact, would most likely overload the sequencing instrument and reduce the number of reads from other fragments.

In order to perform a deeper analysis of the libraries, I decide to proceed with gel extraction of the material. The PCR products were separated on a 4% agarose gel, so I cut out the PCR products running above the 127bp product, up to the 200bp band of the DNA ladder. It is interesting to see how differently the same libraries run on different gels (Fig. 13C). After extraction, the libraries were analyzed on an Agilent Bioanalyzer 2100 instrument that allows the quantification of the size of the products with high accuracy (Fig. 13D). From the Bioanalyzer traces, one can see that, in the three control samples, the average size of fragments was roughly 200bp (Fig. 13D, lanes 1, 2 and 3). In the UV crosslinked Osk IP sample, the average size of the library was ~230bp (Fig. 13D, lane 4). The difference in size of the libraries might reflect the presence of genuine ligation products in the UV crosslinked Osk IP condition that were absent in the controls. To understand the content of the libraries generated in the Osk iCLIP experiment, I decided to proceed with their sequencing. Although the quality of the libraries was clearly suboptimal, the sequencing would at least allow me to determine whether the three mRNAs I identified in the Osk CLIP using the candidate approach were present, which would constitute a starting point to further improve the protocol.

3.2.4. Oskar iCLIP libraries analysis

The four Osk iCLIP libraries were sequenced on a MySeq Illumina platform (single-end sequencing). First, I demultiplexed the raw data to separate the reads generated under the

different experimental conditions [192]. Each library contained between 3 and 4 million reads. I aligned the data against *Drosophila* genome using the TopHat2 tool [193], and PCR duplicates were removed [194]. Only a minor portion (~0.01%) of the reads was mapped to the *Drosophila* genome and all the reads mapped to two rRNA clusters on chromosomes 2 and X. The great majority of the reads in all four conditions (~99.99%) contained sequences from the L3App adapter and PCR primers used for the final amplification of the libraries. These data suggest that the bands extracted from the four different experimental conditions most likely represent undesired products generated during preparation of the libraries. I was not able to confirm any of the target mRNAs identified in the candidate approach among the reads from any of the iCLIP samples prepared under any of the conditions; indeed no reads could be mapped to *nos*, *pgc* or *gcl* mRNAs. The analysis of the libraries showed a high level of PCR duplication, higher than 95% in all four conditions; this drastically reduced the number of unique mappable reads, to between 20 and 40 in each of the four libraries.

These data confirmed that the Osk iCLIP experiment was not successful. The iCLIP method is one of the gold standard experimental approaches for studying protein-RNA interactions. Several explanations for the outcome of the experiment could be envisioned. One might be that Osk is simply not an RBP, and that the enrichment of *nos*, *pgc* and *gcl* mRNAs measured in the Osk CLIP experiment (Fig. 9) was due to the co-IP of other RBPs together with Osk. Even if, by western blot analysis, the stringent washing conditions of the method destroyed the interaction of Osk with Vas and Smg, I cannot exclude that the interaction of Osk with other RBPs was similarly affected. Alternatively, it is possible that although present in residual amounts below the sensitivity threshold of western blot detection, some Vas and Smg crosslinked to their target RNAs might have co-IPed with Osk.

3.3. Oskar CLIP from denatured embryo extracts

If Osk is not a genuine RBP, then the UV-dependent enrichment of *nos*, *pgc* and *gcl* upon Osk IP might be due to residual interactions of Osk with some proteins that directly contact

these mRNAs. To see if this might be the case, I sought to increase the stringency of the Osk IP protocol and established a method for Osk IP from denatured embryo extracts. This protocol consists of lysing 0 to 2 hour old embryos, UV irradiated or not, in a buffer containing 8M urea. After lysis, the extracts are diluted 10 fold, to reduce the urea concentration to 0.8M. Afterwards, an Osk specific IP is performed overnight (16 hours) at 8ºC. A 0.8M concentration of urea and high salt and detergent concentrations are maintained during the first washing steps after the IP. The denaturing conditions used during embryo lysis should denature Osk and thus prevent its association with factors that might otherwise mediate the co-IP of specific mRNAs; blocking even residual interactions with other factors might allow me to IP Osk protein exclusively. Under such conditions, I should not be able to measure any specific RNA enrichment upon UV crosslinking if Osk was not directly crosslinked to mRNAs.

Fortunately, the Osk specific antibody I was using functions in denatured embryo extracts. This allowed me to specifically IP Osk from both UV crosslinked and non-crosslinked extracts, as can be seen by the specific depletion of Osk protein from the input material (Fig. 14A). Furthermore, Vas did not co-purify with Osk, and its level did not change after IP in the input material used in the experiment. As previously described, after RNA extraction from the IPs and after reverse transcription, a qPCR analysis of candidate RNAs was performed as a first measure of the success of the experiment. Therefore, I tested some of the targets tested in the first Osk CLIP experiment: *nos*, *pgc*, *gcl*, *osk*, *bcd* and *Act42A* mRNAs.

The results of the qPCR analysis are shown in Figure 14B. With the new experimental conditions, *nos*, *pgc* and *gcl* mRNAs were again significantly enriched in a UV- and Osk-dependent manner. *osk* mRNA, as well as the negative control mRNAs *bcd* and *Act42A* were not enriched in the precipitated material. The fold enrichment of *pgc* (~18 fold) was similar to that in the Osk CLIP experiment shown in Figure 9B. However, *gcl* was only ~3 fold enriched, compared with the ~6 fold I had obtained previously (Fig. 9B). Strikingly, the level of *nos* enrichment dropped from ~25 fold (Fig. 9B) to ~4 fold using the new experimental conditions (Fig. 14B).

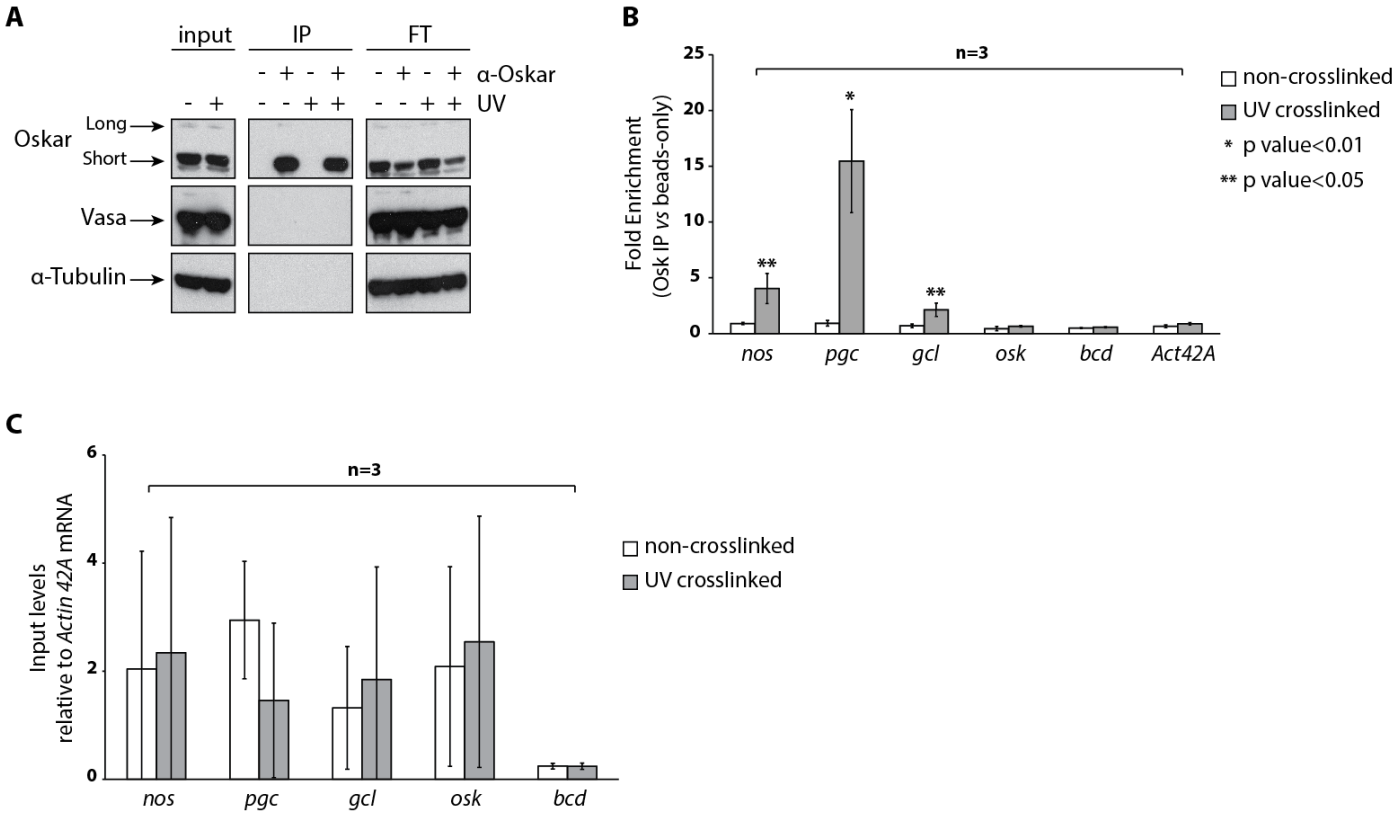

Figure 14: Oskar CLIP in denatured embryo extracts. (**A**). 0 to 2 hour old *Drosophila* embryos were treated (UV +) or not (UV -) with UV light. The lysis was performed in denaturing lysis buffer containing 8M urea. After dilution and clarification, 5ml aliquots (0.5g of embryos) of the extracts from both conditions were incubated with beads conjugated with an anti-Osk antibody (α-Oskar +) or with unconjugated beads (α-Oskar -). IP incubation was performed overnight (16 hours) at 8ºC. After washing, 1% of the total protein content of the beads (IP) was analyzed by western blot for: Osk, Vas and α-Tubulin. The four lanes from the right show the protein content of the flow through (FT) after IP incubation. (**B**). qPCR analysis of the RNAs extracted after Osk CLIP. *nos*, *pgc gcl*, *osk*, *bcd* and *Act42A* mRNAs were analyzed. For both UV crosslinked and non-crosslinked conditions, the level of enrichment of the different targets measured in the Osk IP (α-Oskar +) was normalized to the level measured in the beads-only condition (α-Oskar -). Bars represent mean values ± standard deviation. p values were calculated using Student's t test. (**C**). Relative abundance of the target mRNAs in B measured in the input extracts. The levels were normalized to that of the housekeeping *Act42A* mRNA. Bars represent mean values ± standard deviation.

These data demonstrate that Osk can be UV crosslinked to *nos*, *pgc* and *gcl* mRNAs *in vivo*. *pgc* seems potentially a direct target of Osk, since the level of *pgc* RNA enrichment upon UV crosslinking and Osk IP didn't change when protein-RNA complexes were denatured after UV irradiation. *gcl* was just slightly enriched in the experiment in Figure 9B and in the new experimental conditions its enrichment was decreased. Although the enrichment of *gcl* is

statistically significant, the interaction of Osk with this mRNA appears weak, suggesting that it might be unspecific. The results obtained for *nos* mRNA opens new questions regarding its possible direct interaction with Osk. If *nos* was the most enriched target in the previous Osk CLIP experiment (Fig. 9B), in the new conditions its level of enrichment was very close to that of *gcl*. The drop in enrichment might indicate that, in the previous experimental conditions, the majority of *nos* mRNA was not directly crosslinked to Osk, but that an additional protein crosslinked to *nos* was co-IPed together with Osk. Protein denaturation during extract preparation might prevent Osk from interacting with the protein(s) that directly bind *nos* transcripts.

3.3.1. Conclusions and next steps towards Oskar target RNAs identification

The Osk CLIP experiment performed in denatured embryo extracts showed that Osk can be UV crosslinked to specific mRNAs *in vivo*. *pgc* appears to be a genuine target of the protein *in vivo*, while in case of *nos* other factors might be involved in regulating or stabilizing the Osk-*nos* interaction. Unfortunately, because the UV crosslinking step is performed on intact embryos, these data did not allow me to conclude that Osk is a genuine RBP. Knowing where Osk is crosslinked within its putative target mRNAs would provide key information for understanding if Osk is an RBP. If Osk is a genuine RBP, it might recognize specific motifs or secondary structure elements in mRNAs. This would allow me to design mutations in these elements that might disrupt the interaction with Osk. In the event that the Osk RNA-binding activity is totally unspecific, the UV crosslinking of some posteriorly localized mRNAs to Osk might simply reflect the close proximity of the protein with these particular transcripts within the pole plasm.

The fact that Osk can be UV crosslinked *in vivo* to *nos*, *pgc* and *gcl* mRNAs (Fig. 14) suggests that iCLIP method can be applied to Osk. It appears that the lack of information obtained from the Osk iCLIP libraries shown in Figure 13 was not due to the absence of RNAs

crosslinked to the protein. Other steps of the protocol might have caused the poor outcome of the experiment. In particular, the three enzymatic steps performed before RNA extraction, partial RNA digestion, 3' end dephosphorylation and L3App adapter ligation, appear to be crucial steps in the preparation of libraries of good quality. In fact, when performing the experiment, I was not able to truly control the partial RNA digestion step, and it is possible that over-digestion of the RNA in the samples might have caused the poor results. Moreover. I did not assess the efficiency of the dephosphorylation and the L3App ligation steps. Considering the small amount of starting material in the experiment (Fig. 11), a low efficiency of any or all these steps would compromise the overall quality of the libraries and of the whole experiment.

After discussion with colleagues, with this in mind I decided to analyze in detail each of the three enzymatic steps in the iCLIP protocol, in order to check their efficiency and optimize them. The following steps in the protocol, reverse transcription, circularization and linearization, were all to be performed using standard conditions provided by the manufacturers of the different enzymes and commercial kits utilized, hence they were not the primary focus of the following optimizations.

3.4. Oskar iCLIP optimization

The Osk CLIP experiment performed in denatured embryo extracts showed that Osk can be directly crosslinked to *pgc* mRNA, and to a lesser extent to *nos* and *gcl* mRNAs *in vivo* (Fig. 14). This finding represent the starting point for trying to apply the iCLIP method to Osk protein. The iCLIP protocol is complex and involves several key enzymatic steps (Fig. 10), therefore it was possible that the unsuccessful results I had obtained so far might be due to improperly optimized reaction conditions. I therefore decided to focus on the optimization of the three enzymatic steps performed before RNA extraction: partial RNA digestion, 3' end dephosphorylation and L3App adapter ligation. The analysis of these reactions and their optimization would ideally lead to an optimized protocol, appropriate for studying the interactions of Osk with RNAs *in vivo*.

3.4.1. Partial digestion by RNases

The iCLIP protocol involves partial digestion of RNAs crosslinked to the protein of interest using RNases, to obtain fragments ranging between 50 and 300 nucleotides in size. The amount of RNases that is required to obtain such fragments can be determined in preparative experiments. In such experiments, after treating the samples with different amounts of a specific RNase, RNA fragments are radioactively labeled. Protein-RNA complexes are then separated by SDS-PAGE, blotted on a nitrocellulose membrane and, by analyzing the radioactive signal from the blot, it is possible to identify the right amount of enzyme to be used in the reaction [186] [191]. I performed the RNA labeling assay after Osk CLIP performed in denatured embryo extracts but I was not able to detect any signal from Osk-RNA complexes (data not shown). The negative outcome of the experiment prevented me from defining the amount of RNase I to be used during partial RNA digestion step in the iCLIP protocol.

As an alternative approach, I decided to perform the partial RNase-mediated digestion directly on the input extracts, during the Osk IP. My aim was to find conditions that would allow me to partially digest the total RNA content of the extracts and produce fragments of an average length around 200-300 nucleotides. Since the RNase I used in my first Osk iCLIP experiment (Fig. 13) is inhibited in the lysis buffer used for embryo extract preparation (data not shown), I used a mix of RNase A and T1. In the test experiment, 1ml aliquots of embryo extracts were incubated in the presence of increasing amounts of RNase A and T1. As ultimately I wanted to perform the RNase-mediated digestion step while the Osk IP was ongoing, the RNase optimizations involved incubating the reactions for 3 hours at 8ºC (the conditions under which the IPs are performed). At the end of the incubation, I extracted the RNA and analyzed it on an Agilent Bioanalyzer 2100 (Fig. 15A).

The results of the test experiment are shown in Figure 15A. In the untreated sample (lane 1), the total RNA appears intact, with no sign of degradation. The two prominent bands running slightly below the 2000 nucleotide band of the RNA ladder are the 28S and 18S rRNAs [189] [190]. Addition of increasing amounts of RNase A and T1 to the extracts resulted in digestion

of the total RNA into progressively smaller fragments (lanes from 2 to 11). Examination of the different samples undergoing the controlled digestion revealed that, as the rRNA bands lost intensity, additional bands appeared lower on the gel. Whereas the rRNAs bands were totally gone in the presence of RNase A and T1 at a concentration of $4*10^{-5}$U/µl (lane 5), with $4*10^{-3}$U/µl or higher, the total RNA extracted was fully degraded (lanes 9, 10 and 11). Reactions performed at a concentration of $8*10^{-5}$U/µl of RNase A and T1 (lane 6) seemed to produce RNA fragments of the most optimal size: the RNA fragments spread from the bottom of the lane to the 1000 nucleotide band of the RNA ladder, and the majority of the signal accumulated below the 500-nucleotide band of the ladder, reflecting fragments in a size-range optimal for an iCLIP experiment. It is important to realize that the fact that the only bands clearly visible in the RNA trace in lane 1 (Fig. 15A) correspond to rRNAs does not necessarily mean that the digestion products accumulating after RNase A- and T1-mediated digestion are solely derived from rRNAs. Bioanalyzer traces are produced by the instrument and represented as images resembling a gel. The white background color of each lane is arbitrarily set by the instrument after sample reading. In this case, the rRNA signal is particularly intense and therefore masks any additional RNA traces running above and/or below these bands. In order to control for possible mRNA over-digestion, I reverse-transcribed 1µg of total RNA from samples 1 and 6. Then I measured by qPCR the relative abundance of *nos*, *pgc* and *gcl* mRNAs, comparing the untreated and RNase treated samples. All three mRNAs could be detected after partial RNase A and T1 digestion, but at lower levels compared with the control (data not shown). This indicated that the RNase treatment didn't over-digest the mRNAs in the embryo extracts.

Altogether, these data suggested that performing the RNase A- and T1-mediated RNA digestion step during Osk IP from the input extracts might allow better control of this particular step in the iCLIP protocol. The RNA fragments generated appear to be of an optimal size for iCLIP library preparation, the samples do not look over-digested, and the target mRNAs identified by candidate approach, *nos*, *pgc* and *gcl*, are detectable in the extracts after partial RNA digestion. An additional positive aspect of this method is that it allows easy checking the degree of RNA fragmentation prior to library preparation. Indeed, at the end of the IP and

RNase treatment, total RNA can be extracted from the different aliquots of embryo extract used and analyzed to assess the outcome of the RNase-mediated digestion step before proceeding with the experiment.

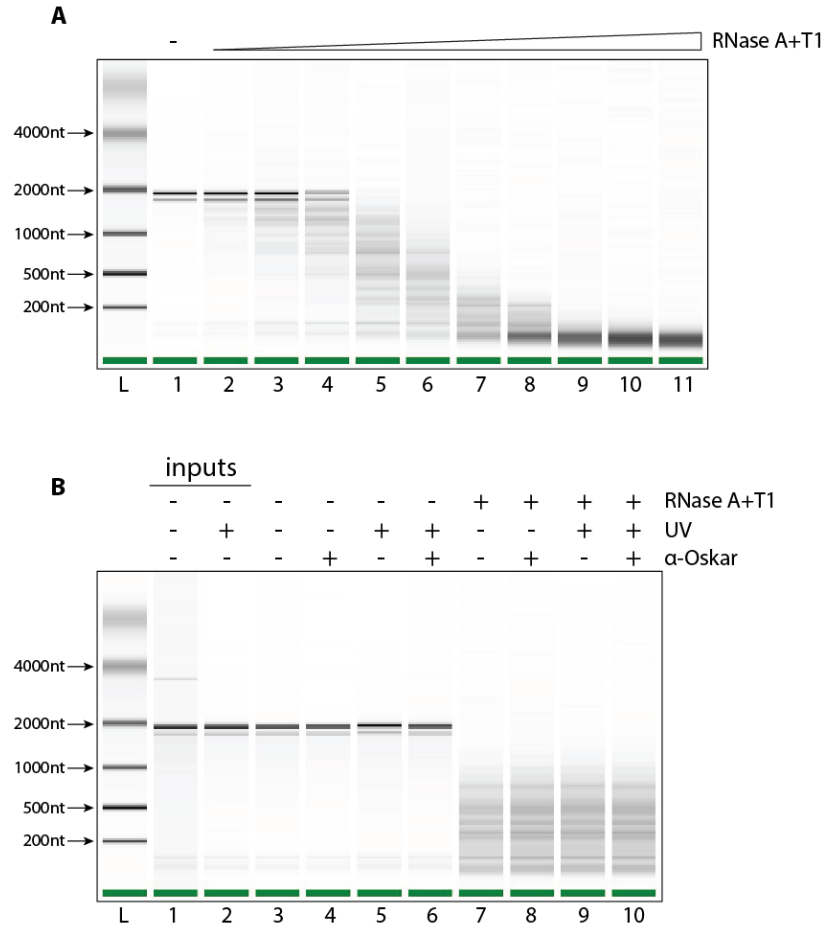

Figure 15: RNase A and T1 partial RNA digestion. (**A**). RNA digestion test using increasing amounts of RNase A and T1. 1ml aliquots of *Drosophila* embryo extracts prepared in denaturing conditions were incubated 3 hours at 8^0C in presence of increasing amounts of RNase A and T1. After incubation, total RNA was extracted and analyzed on an Agilent Bioanalyzer 2100 instrument. RNase A and T1 concentrations used in the assay: sample 2, $4*10^{-6}$U/µl; sample 3, $8*10^{-6}$U/µl; sample 4, $1.6*10^{-5}$U/µl; sample 5, $4*10^{-5}$U/µl; sample 6, $8*10^{-5}$U/µl; sample 7, $4*10^{-4}$U/µl; sample 8, $8*10^{-4}$U/µl; sample 9, $4*10^{-3}$U/µl; sample 10, $8*10^{-3}$U/µl; sample 11, $4*10^{-2}$U/µl. For each sample 50ng of total RNA were loaded into the instrument. L: RNA ladder (**B**). Bioanalyzer traces of RNAs extracted from two Osk CLIP experiments. *Drosophila* embryos, either UV crosslinked (UV +) or non-crosslinked (UV -) were lysed in lysis buffer containing 8M urea. After dilution, 5ml aliquots of the extracts were incubated with protein G sepharose beads conjugated with an anti-Osk antibody (α-Oskar +) or only with beads (α-Oskar -). The samples from the two Osk CLIP experiments were incubated overnight (16 hours) at 8ºC. One set of samples was supplemented with RNase A+T1 mix (final concentration $8*10^{-5}$U/µl) 3 hours before the end of the incubation time (RNase A+T1 +), while the other was not (RNase A+T1 -). After several

washes, proteins were digested and the RNAs extracted. The extracted RNAs were analyzed on an Agilent Bioanalyzer 2100 instrument. For non-crosslinked and UV crosslinked input embryo extracts (lanes 1 and 2), 5ng of total RNA were loaded into the machine. The RNA samples extracted from the two Osk CLIP experiments (from lane 3 to 10) were resuspended in 5µl of H_2O and 1µl was loaded into the machine. L: RNA ladder.

As an additional test, I evaluated the RNase A- and T1-mediated digestion in a standard Osk CLIP experiment, repeating the experiment shown in Figure 14, and performing two parallel sets of Osk IPs in denatured embryo extracts. In one case the samples were not treated with RNase A and T1, whereas in the other, RNase A and T1 ($8*10^{-5}$U/µl) were added during the last 3 hours of IP incubation. After several washes, the proteins were removed by enzymatic digestion, and RNAs were extracted and analyzed on the Agilent Bioanalyzer 2100 instrument (Fig. 15B). As can be seen in Figure 15B, the untreated samples were quite strongly contaminated with rRNAs, probably due to the huge abundance of these RNA species in the input extracts, which tend to stick to the protein G sepharose beads used (Fig. 15B, lanes 3, 4, 5 and 6). In contrast, the four RNase treated samples (Fig. 15B, lanes 7, 8, 9 and 10) contained fragments of a size distribution similar to those obtained under the conditions determined as most optimal in the optimization experiment (Fig. 15A, lane 6). This demonstrated the reproducibility of the partial RNA digestion step. The amount of RNA extracted from the different IP samples in both RNase treated and untreated condition ranged between 1 and 3ng and might be considered at the limit for feasibility of iCLIP library preparation. To additionally analyze the applicability of the RNase A and T1 treatment in the Osk iCLIP, I repeated the experiment described in Figure 15B and after RNA extraction and reverse-transcription I performed the qPCR analysis showed in Figure 14. In such experiments, *nos*, *pgc* and *gcl* mRNAs were significantly enriched after UV crosslinking and Osk IP in both RNase treated and untreated conditions (data not shown), indicating that the conditions that I identified for the partial RNA digestion step can be applied during iCLIP library preparation.

The abundant rRNA content in the samples in Figure 15B was not present using the IP condition established for the first Osk iCLIP experiment (Fig. 11) and was going to represent a

major contamination during iCLIP libraries preparation. Nevertheless, the Osk CLIP protocol in denatured embryo extracts still allowed detection of the UV- and Osk-dependent enrichment of *nos*, *pgc* and *gcl* mRNAs (Fig. 14) and reduced the likelihood of co-IP for proteins that interact with Osk. With this protocol I was more confident that the enriched mRNAs truly reflect crosslinking events that occurred *in vivo* between Osk and its putative mRNA targets. For this reason, I decided that, when preparing iCLIP libraries in the future, I would perform the Osk IP on denatured embryo extracts. The high number of reads generated in such an experiment might allow bioinformatic filtering out the signal from rRNAs, and analysis of the remaining reads for identification of the targets and definition of the binding site(s) of Osk protein.

3.4.2. 3’ end dephosphorylation

After RNase-mediated digestion, the remaining RNA fragments bear a phosphate group at their 3’ end. However, T4 RNA ligase 2 truncated-mediated L3App adapter ligation requires that the RNA substrates harbor an -OH group at the 3’ end, making necessary the removal of the residual phosphate groups. This reaction is catalyzed by the PNK enzyme. I set out to test the efficiency and possibly optimize this essential step of my experiment.

To test the efficiency of PNK-mediated dephosphorylation in my reactions, I took an indirect experimental approach. In this method, two RNA oligos of identical sequence are used as substrates of the dephosphorylation reaction. The oligos differ in that one bears a 3’-OH group (oligo 1), whereas the second bears a 3’ phosphate group (oligo 1-P) (Fig. 16A).

The oligos I chose were 40 nucleotides long and contained a central sequence (28 nucleotide) from the *Renilla reniformis* luciferase gene, flanked by 6 nucleotide-long random stretches. The randomized sequences flanking the core of the oligo would reduce the likelihood of double strand formation between the 5’ and 3’ ends of the same molecule, ensuring that the reaction mix contained single stranded 3’ ends accessible to enzymes.

To test the efficiency of the dephosphorylation reaction, I first incubated increasing amounts of RNA substrate in the reactions, from $8*10^{-4}$ to 2 pmoles. I set up three different reactions for each substrate concentration: a negative control, the experiment itself, and a positive control. In the negative control sample, I incubated the phosphorylated oligo (oligo 1-P) in the iCLIP dephosphorylation reaction buffer but did not add the PNK enzyme; in the experimental sample, I incubated the phosphorylated oligo (oligo 1-P) in the presence of PNK, thus mimicking the experimental conditions of the iCLIP protocol; in the positive control sample I incubated the unphosphorylated oligo (oligo 1) without any PNK (Fig. 16A). All the reactions were performed under the same conditions as the iCLIP protocol [186], in the presence of the same amount of protein G sepharose beads as for Osk IP, at 37ºC for 30'.

The experiment was designed such that, in the experimental sample, PNK activity would be expected to remove the phosphate group from the phosphorylated oligo, producing a 3'-OH substrate required for the following ligation step. In contrast, in the negative and positive control samples, the chosen oligos should not be modified. Consequently, while in the subsequent ligation step, the unphosphorylated oligo of the positive control would be a suitable substrate for adapter ligation, the phosphorylated oligo of negative control sample would not (Fig. 16A). Since in reality not all the phosphorylated oligo molecules contain a phosphate group at their 3' end (as indicated by the manufacturer) and ligation products would most likely be produced also in the negative control, this control would determine the expected background level of ligation for a given amount of oligo used in the assay. Conversely, in the positive control, all the unphosphorylated oligo can be used as substrate for the ligation step, hence this condition can serve to determine the maximum dephosphorylation efficiency in the reactions for a given amount of oligo. Finally, the experimental sample should result in an amount of product somewhere in between the two controls.

After the dephosphorylation step, I performed the L3App ligation using T4 RNA ligase 2 truncated. Next I reverse-transcribed the ligation products using the iCLIP primers, which anneal to the L3App adapter sequence. Finally, I analyzed the cDNAs by PCR and separated

the products on an agarose gel (Fig. 16A). The size of the specific PCR products generated in the experiment were expected to be 89bp long.

As shown in Figure 16B, with the smallest amounts of RNA oligo substrate ($8*10^{-4}$ and $8*10^{-3}$ pmoles), no background signal was detected in the negative control. With higher amounts of substrate (0.08, 0.4 and 2 pmoles) the intensity of the specific PCR product in the negative control conditions increased proportionally with the amount of substrate used. However, in all cases, the intensity of the band in the negative control was lower than that of both the positive control and the experimental condition performed with the same amount of oligo (Fig. 16B).

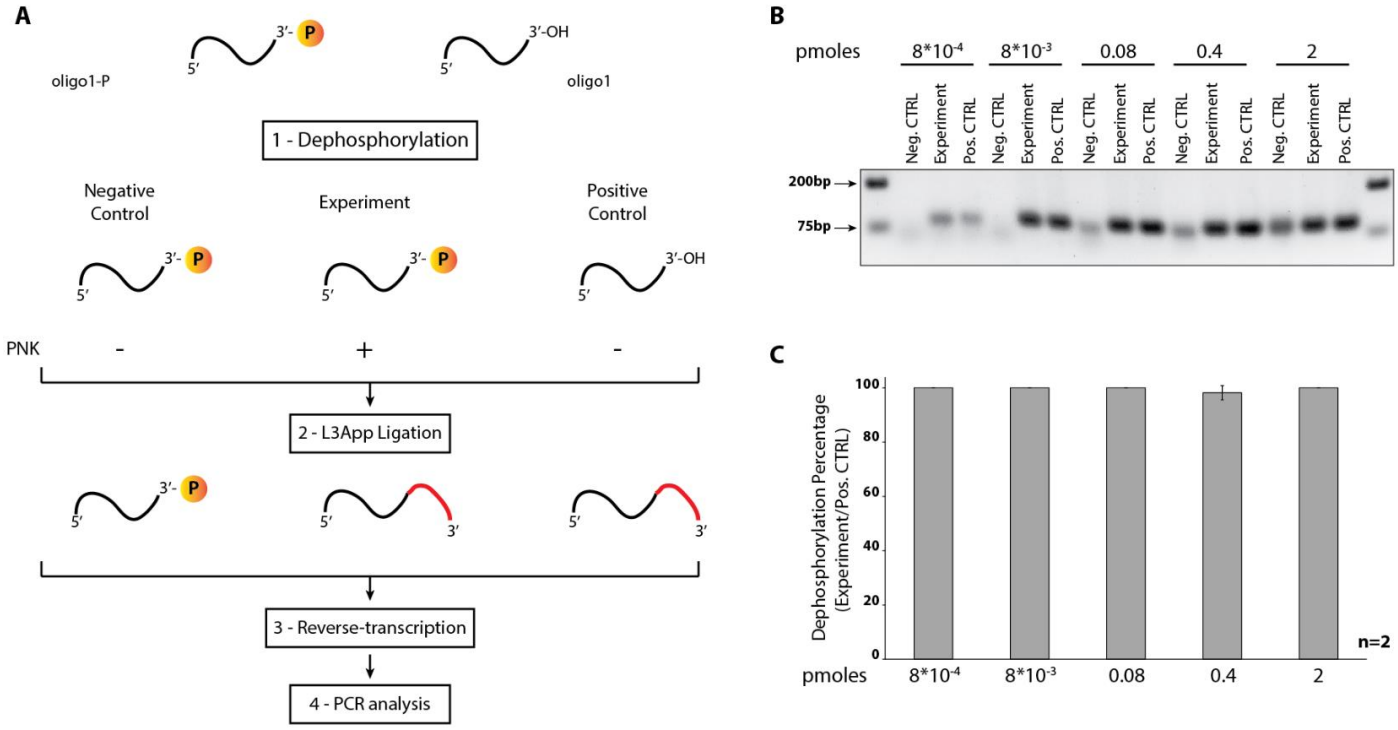

Figure 16: PNK-mediated 3' end dephosphorylation efficiency test. (**A**). Schematic representation of the experimental procedure. Two oligos of identical sequence are used in the experiment: oligo 1-P contains a phosphate group at the 3 'end while oligo 1 does not. For a given amount of oligo three reactions were performed. In the negative and positive control samples, oligo 1-P and oligo 1 respectively, were incubated without PNK. In the experimental sample, oligo 1-P was incubated with PNK. The reactions were performed on protein G sepharose beads in iCLIP dephosphorylation reaction buffer and were incubated at 37ºC for 30'. At the end of the incubation the RNA oligos were extracted and L3App adapter ligation was performed using T4 RNA ligase 2 truncated enzyme. After an additional step of oligo purification, reverse-transcription was performed using primers

that anneal to the adapter sequence. cDNAs were analyzed by PCR using a primer that anneals to the sequence of the primer used in the reverse-transcription step and a primer that anneals to the RNA oligo sequence. (**B**). Agarose gel showing the PCR products from a dephosphorylation efficiency test experiment performed using different amounts of RNA oligo substrate. (**C**). Densitometric analysis of the gel shown in B. For each experimental point, values from the negative control conditions were subtracted from the experimental and positive control conditions. Values obtained from positive controls were used to define 100% dephosphorylation efficiency for each experimental point. Bars represent mean values ± standard deviation.

To quantify the efficiency of the dephosphorylation reactions, I performed a densitometric analysis of the signal from the agarose gel (Fig. 16B). The graph shows the results of this analysis (Fig. 16C). For each experimental condition, I subtracted the values measured in the negative control from those in the experimental and positive control samples. For each oligo amount tested, I considered the values in the positive control samples as representing the maximum level of efficiency (100% dephosphorylation). This analysis showed that, for all the different RNA substrate amounts tested in the assay, the efficiency of dephosphorylation was nearly 100% (Fig. 16C).

This experiment indicated that PNK-mediated dephosphorylation of the RNA substrate was not the limiting step in the iCLIP protocol. Therefore, the low quality of the libraries prepared in my Osk iCLIP experiment (Fig. 13) was probably not due to inefficient 3' end dephosphorylation.

3.4.3. L3App adapter ligation onto RNA

3.4.3.1. L3App adapter ligation efficiency in iCLIP conditions

Ligation of the L3App adapter to the 3' end of RNA fragments crosslinked to Osk is the last enzymatic reaction performed on beads in the Osk iCLIP protocol [186]. I set up an experimental procedure to analyze the efficiency of this reaction under the iCLIP conditions. As a first step, I labelled different amounts of the same RNA oligo (oligo 1) used in the dephosphorylation efficiency test at the 5' end with radioactivity, by PNK-mediated addition of

a ^{32}P containing phosphate group. After RNA extraction, I incubated the oligo with T4 RNA ligase 2 truncated enzyme (200U) and a molar excess of L3App adapter (60 pmoles) in the iCLIP ligation reaction buffer. In this assay, I used increasing amounts of RNA substrate, from $8*10^{-4}$ pmoles to 2 pmoles. The reactions were performed in the presence of the same amount of protein G sepharose beads as in Osk IP and were incubated overnight (16 hours) at 16ºC, in accord with the iCLIP protocol [186]. At the end of the incubation period, I extracted the RNA oligos and separated them on a 10% polyacrylamide denaturing gel that I subsequently exposed to a phosphor imaging plate to detect the radioactive signal. By densitometric analysis, I calculated the percentage of ligation from the ratio between the signal of the ligation product *versus* the total signal in the experiment (ligation product plus free RNA substrate).

The results of the L3App ligation efficiency test performed in iCLIP conditions are shown in Figure 17A and B. The lower band represents the unligated RNA substrate, while the band just above represents the ligation product. In all samples, the majority of the RNA substrate was not ligated, revealing that the ligation reaction was inefficient. In the samples containing the highest amount of oligo (Fig. 17A, lanes 6 and 7), an additional band (*) is present above the first ligation product. This band is most likely the product of two ligation events involving the same RNA substrate. As mentioned previously, the L3App adapter contains two modifications: the 5' end is pre-adenylated and the 3' end is blocked by a dideoxycytidine. As in the case of the oligo 1-P (previous paragraph), not all the oligos are modified, and in this case only 20% of the L3App molecules contain both modifications. Therefore, ligation of an RNA substrate molecule to an adapter lacking the 3' end modification can occur. In this case, the ligation product will contain a 3' -OH group that can undergo a second ligation event. Such additional products can be detected on the gel when high amounts of RNA substrate are used, as more ligation products lacking the 3' end modification are present in the reaction.

As can be seen in the graph in Figure 17B, the ligation efficiency under iCLIP conditions was low, ranging from ~2% ($8*10^{-4}$ pmoles RNA substrate) up to ~12% (2 pmoles) (Fig. 17B). The results of these tests revealed the low efficiency of the reaction and that L3App ligation

was the limiting step in my iCLIP experiment. As the total amount of RNA co-precipitated with Osk using CLIP was ~1ng, it is possible that the conditions of the iCLIP were similar to those in the samples shown in lanes 4 and 5 of Figure 17A and B. The percentage of ligation in these conditions was ~5% and ~8% respectively, possibly explaining why the Osk iCLIP experiment failed (Fig. 13).

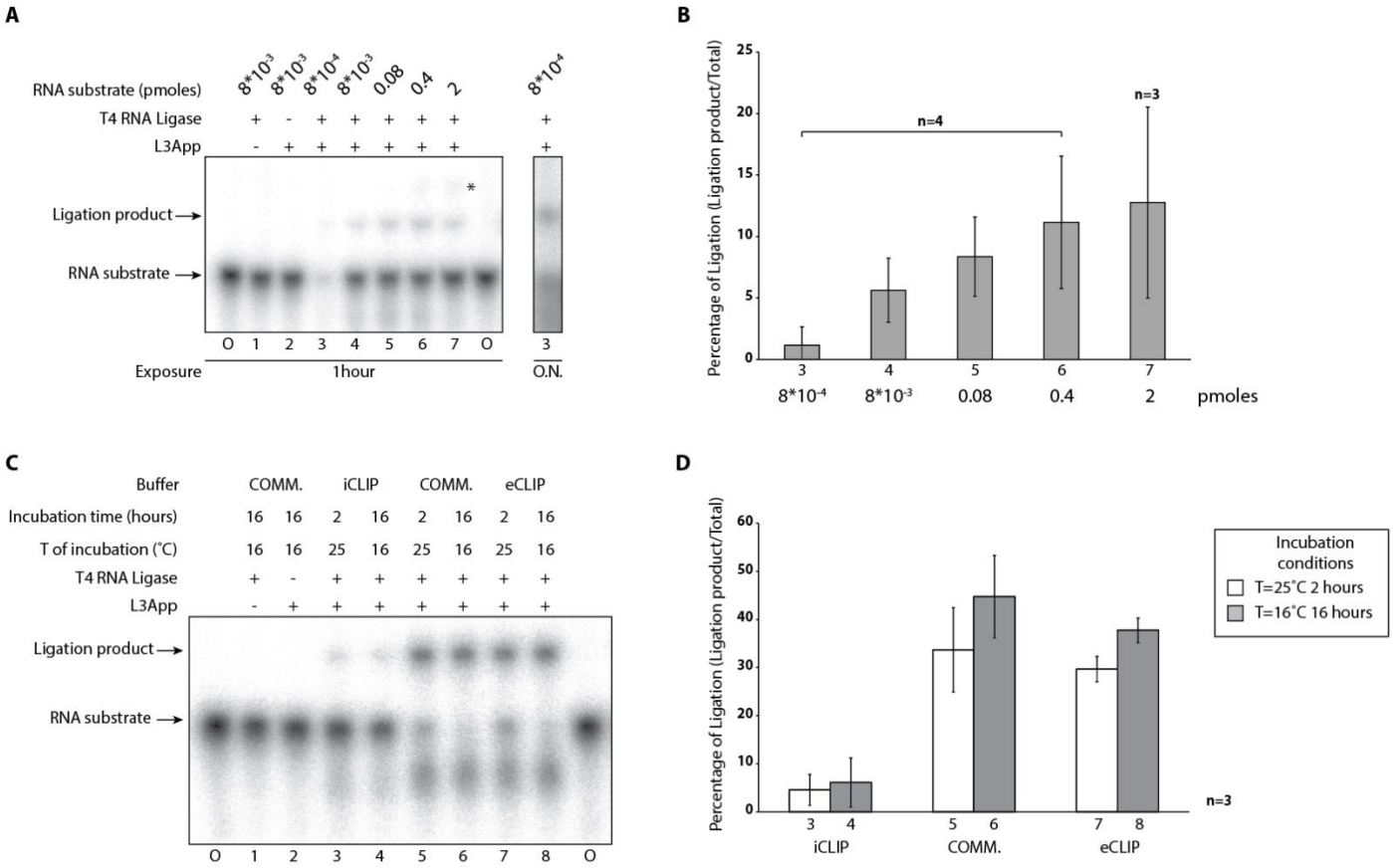

Figure 17: L3App ligation test experiments. (**A**). Ligation test experiment performed using iCLIP conditions. Increasing amounts of radioactively labeled RNA oligo (oligo 1) substrate were incubated with L3App adapter (60 pmoles) in presence of T4 RNA ligase 2 truncated (200U) and protein G sepharose beads. The reactions were performed in iCLIP ligation reaction buffer and were incubated overnight (16 hours) at 16ºC. Following RNA extraction, reaction products were separated on a 10% denaturing polyacrylamide gel. Samples 5, 6 and 7 were diluted in order to load the same amount of RNA oligos used in sample 4. The gel was exposed to a phosphor imaging plate for 1 hour and overnight (O.N.). Reaction products from two ligation events on the same RNA molecules are indicated by the *. O: radioactively labeled free RNA oligo. (**B**). Densitometric analysis of the gel shown in A. Bars represent mean values ± standard deviation. (**C**). Comparison of ligation efficiency in different reaction buffers (samples 3 and 4: iCLIP ligation reaction buffer; samples 5 and 6: commercial ligation buffer; samples 7 and 8: eCLIP ligation buffer). For all samples $8*10^{-3}$ pmoles of RNA substrate were incubated with T4 RNA ligase 2 truncated (200U) in presence of L3App adapter (60 pmoles) and protein G sepharose

beads. For the three different buffer tested two incubation conditions were analyzed: 2 hours at 25ºC or 16 hours at 16ºC. O: radioactively labeled free RNA oligo. (**D**). Densitometric analysis of the gel showed in C. Bars represent mean values ± standard deviation.

3.4.3.2. Effect of buffer composition and incubation temperature

As the L3App ligation was clearly a limiting step in the iCLIP method I was employing, I decided to optimize the reaction and identify conditions that would increase its efficiency. I started by analyzing the effect of different buffer compositions and different times and temperatures of incubation. All reactions were performed in the presence of protein G sepharose beads and a constant amount of T4 RNA ligase 2 truncated (200U). Additionally, I chose to perform the test using $8*10^{-3}$ pmoles of RNA oligo substrate, as in addition to the low ligation efficiency (~5%), this amount of oligo substrate had produced a clear signal on a phosphor imaging plate after 1 hour of exposure (Fig. 17A). Hence these particular conditions seemed a good starting point for optimization of the ligation step.

I first tested three different buffers: the iCLIP ligation reaction buffer, a commercial ligation buffer (T4 RNA ligase reaction buffer), and the enhanced CLIP (eCLIP) ligation buffer. The iCLIP ligation was the one I was aiming to optimize [186]. The commercial buffer is the standard buffer that is commercially provided with the enzyme. The eCLIP ligation buffer was used in a recently published protocol similar to iCLIP [195]. The main difference between the three buffers lies in the presence of PEG8000, rather than of PEG400, and of dimethyl sulfoxide (DMSO) in the commercial and eCLIP buffers. It has been shown that addition of PEG8000 and DMSO increases T4 RNA ligase-dependent ligation [196]. Therefore, I wanted to determine whether the presence of these components in the eCLIP and commercial ligation buffers might improve the efficiency of ligation of my L3App adapter and oligo substrates.

In addition to the different buffers I also tested two incubation conditions: overnight (16 hours) at 16ºC (as stated in the iCLIP protocol) and a shorter (2 hours) incubation at 25ºC (as

recommended by the T4 RNA ligase 2 truncated manufacturer). The overnight ligation introduced a desirable break in the lengthy iCLIP protocol. Although 16ºC is a temperature lower than the optimal one for the enzyme, it should minimize RNA degradation and the long incubation (16 hours) should allow a sufficient level of ligation to be achieved. I also tested if a shorter incubation time at the optimal temperature of 25ºC might improve the outcome of the reaction.

The densitometric analysis of the data confirmed that overnight ligation in the iCLIP ligation reaction buffer leads to low percentage of ligation (~5%) (Fig. 17C and D, lane 4). Increasing the incubation temperature did not produce a higher level of ligation (Fig. 17C and D, lane 3). In both assay conditions, the majority of the RNA substrate was not ligated in the iCLIP ligation reaction buffer (Fig. 17C, lanes 3 and 4). In contrast, use of the commercial and eCLIP ligation buffers both led to a strong increase in the percentage of ligation. In both cases, overnight incubation at 16ºC produced better results compared to the other reaction conditions (Fig. 17C and D, compare lanes 5 *vs* 6, and 7 *vs* 8). Specifically, overnight ligation in the eCLIP ligation buffer resulted in an average of ~38% ligation (Fig. 17D, lane 8) and, in the commercial ligation buffer, of ~45% ligation of the oligo (Fig. 17D, lane 6).

These data show that addition of PEG8000 and DMSO to the ligation buffer increased the ligation reaction efficiency almost 10 fold. The presence of DTT (1mM) in the commercial ligation buffer seems to result in better ligation than what was obtained using the eCLIP ligation buffer, which does not contain any DTT. In view of this, I decided to use the commercial ligation buffer to test additional parameters that might increase L3App ligation efficiency in the Osk iCLIP protocol.

3.4.3.3. Effect of increasing amounts of T4 RNA ligase 2 truncated enzyme

I next tested how increasing the amount of T4 RNA ligase 2 truncated in the reaction might influence L3App ligation efficiency. In this experiment, I incubated a fixed amount of RNA oligo substrate ($8*10^{-3}$ pmoles) with increasing amounts of T4 RNA ligase 2 truncated (200U up to 1200U), in the presence of protein G sepharose beads, in the commercial ligation buffer. As with the previous experiment, I compared two incubation conditions: 2 hours at 25ºC and overnight (16 hours) at 16ºC.

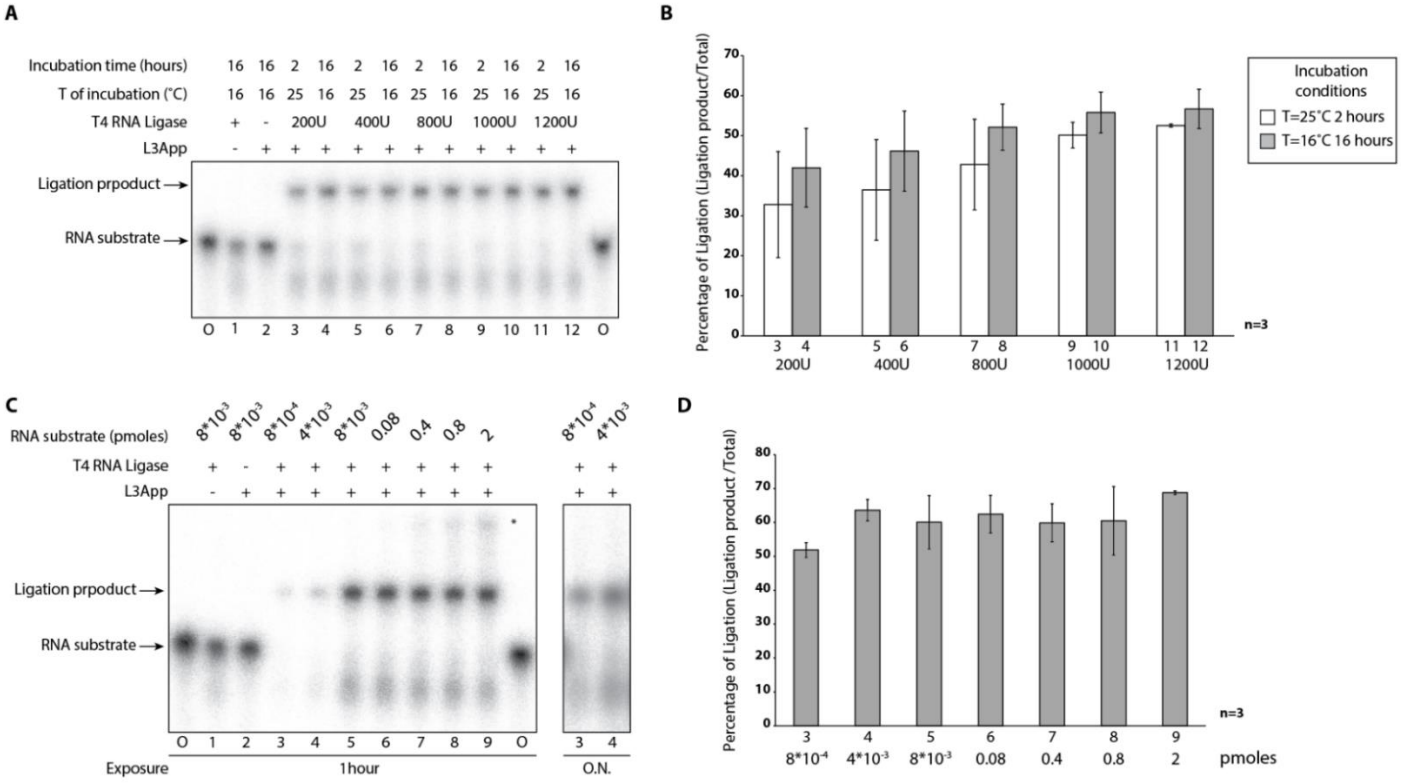

Figure 18: Optimized reaction conditions for L3App ligation. (**A**). Ligation test experiment performed using increasing amounts of T4 RNA ligase 2 truncated. $8*10^{-3}$ pmoles of radioactively labeled RNA oligo 1 were incubated in commercial ligation buffer in presence of increasing amounts of T4 RNA ligase 2 truncated (from 200U to 1200U), L3App adapter (60 pmoles) and protein G sepharose beads. Reactions were incubated 2 hours at 25ºC or 16 hours at 16ºC. Products of ligation were separated on a 10% denaturing polyacrylamide gel subsequently exposed to a phosphor imaging plate for detection. O: radioactively labeled free RNA oligo. (**B**). Densitometric analysis of the gel showed in A. Bars represent mean values ± standard deviation. (**C**). Ligation test experiment performed using increasing amounts of RNA oligo 1 substrate using optimized reaction conditions from the experiment

showed in panels A and B (sample 10). Increasing amounts of radioactively labeled RNA substrate, ranging from $8*10^{-4}$ pmoles to 2 pmoles were incubated in commercial ligation buffer in presence of 1000U of T4 RNA ligase 2 truncated, 60 pmoles of L3App adapter and protein G sepharose beads. After 16 hours of incubation at 16ºC the products of the reactions were separated on a 10% denaturing polyacrylamide gel (samples 6, 7, 8 and 9 were diluted in order to load the same amount of RNA oligos used in sample 5). The gel was exposed to a phosphor imaging plate for 1 hour and overnight (O.N.). Reaction products from two ligation events on the same RNA molecules are indicated by the *. O: radioactively labeled free RNA oligo. (**D**). Densitometric analysis of the gel showed in panel C. Bars represent mean values ± standard deviation.

This experiment revealed that incubating the reaction overnight at 16ºC resulted in a higher ligation efficiency than for 2 hours at 25ºC (Fig. 18A, compare lanes 3 *vs* 4 and 5 *vs* 6). This finding is best illustrated by the graph (Fig. 18B), which shows that, regardless of the amount of enzyme used, 2 hours of incubation at 25ºC always produced a lower percentage of ligation. Increasing amounts of T4 RNA ligase 2 truncated in the reaction mix positively affected the ligation efficiency. When only considering the data from the reactions incubated for 16 hours, the percentage of ligation increased from ~41% when 200U of enzyme were used, to ~56% in the presence of 1000U of enzyme. Amounts of T4 RNA ligase 2 truncated higher than 1000U in the reaction, did not further increase the efficiency of ligation (Fig. 18B).

These data confirm that the best L3App adapter ligation results are achieved when the reaction is incubated for 16 hours at 16ºC, as was shown in the previous experiment (Fig. 17C and D) and show that increasing the amount of enzyme used in the reaction (up to 1000U) further improves reaction efficiency. These new reaction conditions represent an important improvement to the ligation conditions proposed by the standard iCLIP protocol. I decided therefore to use these ligation conditions when performing the final optimization experiments (Fig. 18A and B, lane 10).

3.4.3.4. Effect of different amounts of RNA substrate

The final experiment I performed to optimize L3App ligation was to test the effect of different amounts of RNA substrate on the efficiency of ligation using the optimized

experimental conditions (Fig. 18A and B, lane 10). The results of the experiment are depicted as a graph (Fig. 18D). For the entire range of substrate amounts tested, ligation efficiency was greater than 50% and considerably higher than that achieved using the original iCLIP ligation conditions (Fig. 17B), ranging from ~52% with the smallest amount of oligo substrate ($8*10^{-4}$ pmoles) to close to 70% of ligation when 2 pmoles of RNA substrate were included in the reaction (Fig. 18D).

In conclusion, I identified the ligation of the L3App adapter to the RNA substrate as the limiting step in the iCLIP protocol. By testing different reaction buffers, incubation conditions and amounts of T4 RNA ligase 2 truncated, I managed to optimize the ligation step in the iCLIP protocol, achieving between 50% and 70% of ligation over a wide range of amounts of RNA substrate. The optimization of the ligation step in the iCLIP protocol might help to increase the quality of the libraries produced from the Osk iCLIP experiment possibly allowing me to identify where Osk is crosslinked to its target transcripts *in vivo*.

3.5. The Oskar iCLIP protocol optimized

With the experiments shown in the previous paragraphs, I was able to set up and optimize the three enzymatic steps: partial RNA digestion (Fig. 15), 3' end dephosphorylation (Fig. 16) and L3App adapter ligation (Fig. 17 and 18), all performed before RNA extraction in the iCLIP protocol (Fig. 10). The optimized iCLIP protocol that I set up might help me to increase the overall quality of the libraries produced from an Osk iCLIP experiment. A successful library preparation and analysis might eventually help to identify all the RNAs bound by Osk *in vivo* and the deposition site(s) of Osk on its target transcripts. I therefore decided to proceed and carry out the full experiment.

3.5.1. Preparation of an optimized Oskar iCLIP

For the optimized Osk iCLIP experiment, I began by preparing 5ml of denatured extracts from embryos that had either been, or not been, subjected to UV crosslinking. I incubated the extracts with protein G sepharose beads conjugated with an anti-Osk antibody or, as a control, I performed a parallel IP using beads lacking the antibody. I incubated the samples overnight (16 hours), supplementing the reactions with $8*10^{-5}$U/µl of RNase A and T1 for the last 3 hours of IP incubation. After several washing steps, 1% of the beads used for each sample were subjected to protein extraction and western blot analysis to monitor the IP efficiency. The results of the Osk IP are shown in Figure 19A. Western blotting revealed that Osk protein was specifically precipitated from both UV crosslinked and non-crosslinked samples, and that was strongly depleted from the extracts following the IP. As previously, the IP washing conditions were stringent enough to avoid co-IP of Vas, a well-known Osk interactor [113] [115], and of α-Tubulin, a very abundant cytoplasmic protein (Fig. 19A).

To check integrity of the RNA in the input material and RNA fragmentation pattern after partial digestion with RNases, following the IP incubation I extracted the RNA for analysis on an Agilent Bioanalyzer 2100 instrument. The Bioanalyzer traces from the inputs (UV crosslinked and non-crosslinked) showed that the RNA in the starting material was intact (Fig 19B, lanes 1 and 2). The Bioanalyzer traces of the RNA extracted after RNase A- and T1-mediated digestion showed the level of RNA fragmentation. The four profiles were very similar, indicating an optimal level of digestion of all the samples in the experiment (Fig. 19B, lanes 3, 4, 5 and 6).

Knowing that both the Osk IP and the partial digestion by RNases worked as expected, I proceeded with the iCLIP protocol (Fig. 10), using the optimized ligation conditions that I previously identified. At the end of the protocol, I performed a test PCR, with 28 and 30 cycles for library amplification, and analyzed the products on a 6% native polyacrylamide gel in TBE buffer (Fig. 19C). This analysis showed that with 28 PCR cycles the signals from the four

libraries was not sufficiently strong to appreciate and analyze the differences between the samples (Fig. 19C, lanes 1, 3, 5 and 7).

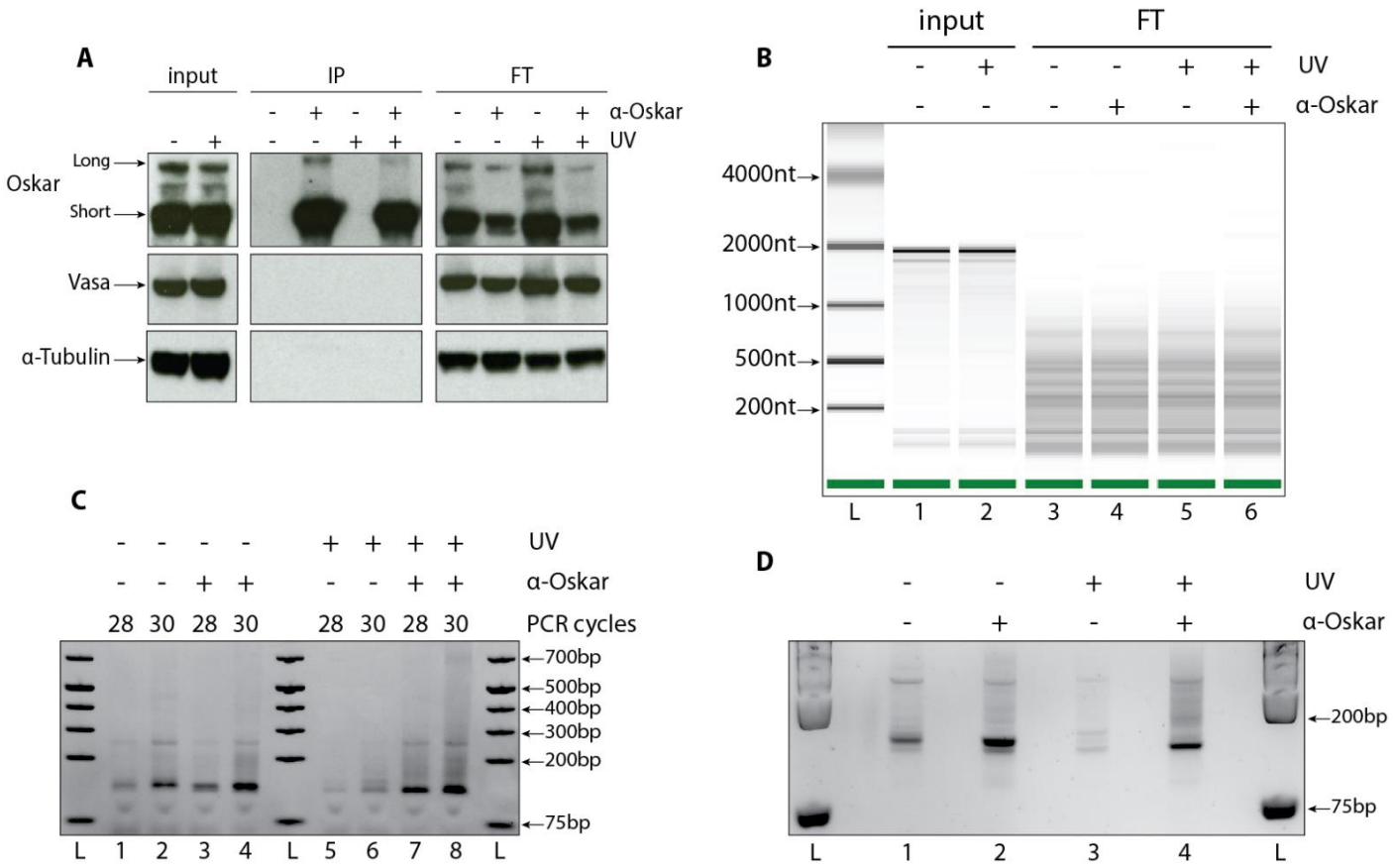

Figure 19: Oskar iCLIP libraries prepared according to the optimized protocol. (**A**). Western blot analysis of the Osk IP step. *Drosophila* embryos collected between 0 and 2 hours after egg laying were irradiated (UV +) or not (UV -) with UV light. Lysis was performed in lysis buffer supplemented with 8M urea. After extract dilution and clarification, 5ml aliquots (0.5g of embryos) were incubated with protein G sepharose beads conjugated with an anti-Osk antibody (α-Oskar +) or, in a parallel control experiment, only with beads (α-Oskar -). After stringent washing steps, 1% of beads from each sample were subjected to protein elution, SDS-PAGE and western blot analysis for: Osk, Vas and α-Tubulin. FT: flow through. (**B**). RNase A and T1 partial RNA digestion. During the last 3 hours of IP incubation, the extracts were supplemented with $8*10^{-5}$U/µl of RNase A and T1. After reaction completion, total RNA was extracted from the extracts used for Osk IP (FT). RNAs have been analyzed on an Agilent Bioanalyzer 2100. L: RNA ladder. (**C**). Test PCR for Osk iCLIP libraries. Final libraries were amplified by PCR using different cycles (28 and 30) and separated on a 6% native polyacrylamide gel in TBE buffer. L: DNA ladder. (**D**). Osk iCLIP library analysis on agarose gel. PCR amplified Osk iCLIP libraries (30 PCR cycles) were separated on a 4% agarose gel in TBE buffer for extraction. L: DNA ladder.

In contrast, the signals produced after 30 PCR cycles allow analysis of the libraries (Fig. 19C, lanes 2, 4, 6 and 8). As with the previous Osk iCLIP experiment (Fig. 13B), all four libraries present a sharp band (127bp) towards the bottom of the gel. This band was produced by residual unligated L3App adapter molecules that were not removed before RNA extraction. All the DNA fragments migrating above this band should in principle represent single ligation events, and therefore should contain useful information. A second sharp band of ~260bp is also detected in all four samples. This band was also present in the previous Osk iCLIP libraries (Fig. 13B), but in this case did not seem to be the major product present in the libraries. Furthermore, after 30 PCR cycles, the two libraries from the non-crosslinked samples looks very similar: in both the beads-only and the Osk IP samples, the signal spreads from the 127bp band to ~260bp band, and no additional product in the upper part of the gel is visible (Fig. 19C, lanes 2 and 4). The library produced from the UV crosslinked beads-only sample shows the least signal between the 127bp and ~260bp bands, and no other PCR products are visible above these two bands (Fig. 19C, lane 6). Interestingly, this particular library looks emptier than the one produced from the non-crosslinked beads-only condition. There are at least two possible explanations for this result: some material from the UV-crosslinked beads-only sample might have been lost during the procedure or, for unknown reasons, during the washing steps performed on this sample, background RNA fragments and unligated L3App adapter molecules might have been removed more efficiently than in the other conditions. Finally, library prepared from the UV crosslinked Osk IP sample looked different from the others. The smearing signal above the 127bp band was stronger than in the other libraries and, interestingly, this library contained an additional smearing signal in the upper part of the gel, between 300bp and 700bp of the DNA ladder (Fig. 19C, lane 8). This feature is completely absent in the libraries from the three control conditions, which might indicate the presence of longer cDNA molecules produced from RNA fragments crosslinked to Osk.

A positive feature of the optimized method of iCLIP library preparation is the number of PCR cycles required to detect their content. In the original iCLIP experiment, 38 PCR cycles were needed to detect a smeared signal in the UV crosslinked Osk IP sample (Fig. 13B, lane

12). The optimized protocol required 8 fewer cycles, which indicates that the cDNA input in the PCR reaction was 1000 times greater than in the previous iCLIP experiment. One possible reason for this difference is the much higher RNA background signal in the samples prepared using the Osk IP protocol from denatured embryo extracts (Fig. 15B). Another possible reason is the improved efficiency of the ligation step (Fig. 18C and D). With the optimized ligation conditions, more than 1 out of 2 RNA molecules is attached to an iCLIP adapter, in a ligation reaction that is ~10 times more efficient than in the standard iCLIP protocol (Fig. 17A and B). Although at this stage of the experiment I could not distinguish which of the two factors was responsible for the fewer PCR cycles needed, the presence of a stronger signal in the UV crosslinked Osk IP library and the fewer PCR amplification cycles required to detect it, indicate that more RNA fragments were used as substrates in the optimized iCLIP experiment. Importantly, this means that the new iCLIP libraries might be less PCR biased than the original ones.

In order to analyze the four new libraries in greater detail, I decided to proceed with the experiment. After PCR amplification, I separated the libraries on a 4% agarose gel (Fig. 19D) and extracted and purified all the DNA fragments running above the 127bp band. PCR fragments from the different samples show a similar distribution to those generated in the test PCR (Fig. 19C). In the non-crosslinked condition, the beads-only sample appears slightly “cleaner” than the Osk IP. In both cases, the signal spreads above the 127bp band representing the unligated L3App adapter, towards the upper part of the gel (Fig. 19D, lanes 1 and 2). The lane containing the UV crosslinked beads-only sample is almost empty compared with the others (Fig. 19D, lane 3). The library prepared from UV crosslinked Osk IP sample shows a stronger signal spreading above the 127bp band (Fig. 19D, lane 4). Curiously, the band running at ~260bp on the test PCR polyacrylamide gel is not detected on the agarose gel (Fig. 19D).

I next analyzed the Osk iCLIP libraries from the four samples on an Agilent Bioanalyzer 2100 instrument to determine with higher accuracy the size of the PCR products (Fig. 20). In

all four, the 127bp band was removed during gel excision, and only ~260bp band is detectable. The two samples from the non-crosslinked condition show very similar profiles: in addition to the ~260bp band, which is the strongest PCR product in the samples, other PCR fragments are mostly distributed between the 150bp and 300bp bands of the DNA ladder. Additional fragments can be detected above 300bp (up to 500bp), but at very low levels (Fig. 20A and B).

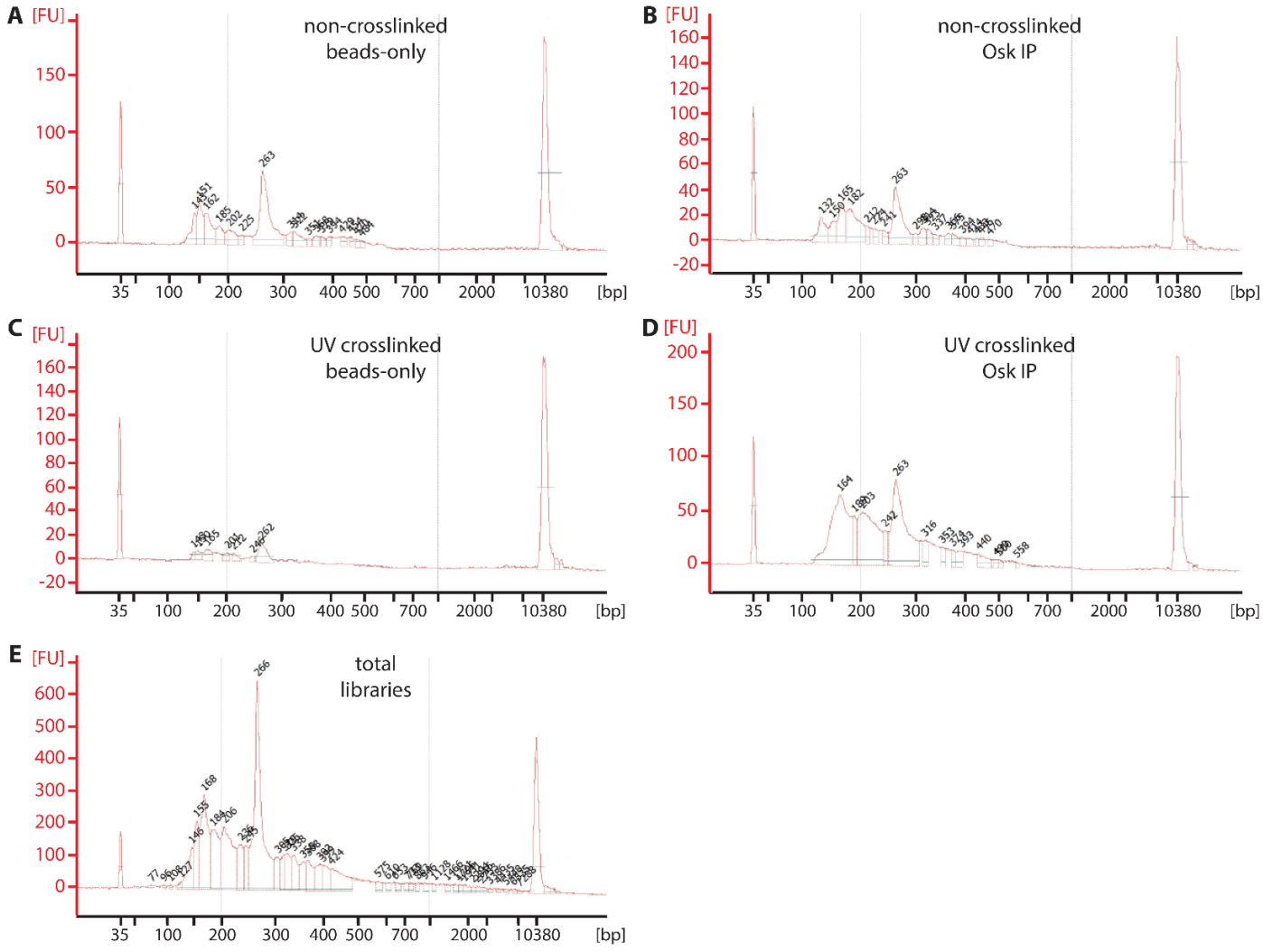

Figure 20: cDNA fragments distribution in Oskar iCLIP libraries. After PCR amplification and gel extraction, DNA fragments in the different Osk iCLIP libraries were analyzed on an Agilent Bioanalyzer 2100 instrument. (**A**, **B**, **C** and **D**). Analysis of single Osk iCLIP libraries from the four different experimental conditions used. (**E**). Analysis of the four libraries produced in the Osk iCLIP experiment pooled and concentrated. On the Y axis, FU represents arbitrary fluorescent units measured by the instrument; on the X axis all samples were aligned respect to a DNA ladder (bp: base pairs).

The library from the UV crosslinked beads-only sample contained very little material. DNA fragments are visible between 150bp and 300bp, but no additional signals are detected in addition to the ~260bp band (Fig. 20C). The UV crosslinked Osk IP specific library had the highest DNA content. Most of the DNA fragments are visible between 150bp and 300bp, and some additional material accumulates up to 600bp (Fig. 20D). It was encouraging to note that this library contained more material between 300bp and 600bp than the libraries from the control conditions. This was a positive indication that might indicate the presence of DNA fragments produced from RNA molecules UV crosslinked to Osk.

It seemed likely that the ~260bp band in the four Osk iCLIP libraries was a contaminant and would not produce any informative data. Indeed, a similar band was present in the previous iCLIP libraries (Fig. 13B) and sequencing analysis of those libraries suggested it was an artifact produced by the L3App adapter. Although in the optimized Osk iCLIP libraries the similarly dominant band was of slightly higher molecular weight, it seemed probable that sequencing of this band would not produce any mappable reads. I considered avoiding this band during gel extraction step, but decided against it, for two reasons. First, during the gel extraction step, the removal of a specific PCR product is not 100% efficient, as demonstrated by the residual presence of the 127bp L3App adapter product at ~130bp, which is visible on the Bioanalyzer trace of the individual iCLIP libraries (Fig. 20). Second, excising the ~260bp band would result in the loss of PCR fragments overlapping with, as well as those migrating just below and above the abundant contaminating band during gel electrophoresis.

After considering the positive and negative points of the experiment, I decided to proceed and have the Osk iCLIP libraries, produced with the optimized protocol, sequenced. The four libraries, after PCR amplification and gel extraction, were pooled together and sent for sequencing. The Bioanalyzer profile of this library preparation is showed in Figure 20E.

3.5.2. Analysis of the Oskar iCLIP libraries

Osk iCLIP libraries produced using the optimized protocol previously described were sequenced on a HighSeq 2000 instrument (Illumina; single-end sequencing). I demultiplexed the raw data [192], aligned the reads against the *Drosophila* genome using the TopHat2 tool [193], and finally I removed PCR duplicates [194].

	total accepted reads	alignment rate	total aligned reads	PCR duplication rate	unique reads mapped
non-crosslinked beads-only	5887649	46.1%	2716758	99.87%	3532
non-crosslinked Osk IP	44558054	59.8%	26667372	99.97%	8000
UV crosslinked beads-only	32267440	63.4%	20449096	99.95%	10225
UV crosslinked Osk IP	31458786	47.3%	14868341	99.92%	11895

Table 1: Statistical analysis of the optimized Oskar iCLIP libraries

The analysis of the data shown in Table 1 revealed a different library size for the different experimental conditions. The library from the non-crosslinked beads-only sample was the smallest (~6 millions of reads), whereas the non-crosslinked Osk IP library was the largest, with ~44 millions of reads. The libraries from the UV crosslinked condition were of a comparable size, with ~32 and ~31 millions of reads in the beads-only and the Osk IP samples, respectively. The percentage of reads successfully aligned in *Drosophila* genome ranged between 46% and 63% depending on the sample. In all four libraries, the reads that could not be aligned mainly contained L3App adapter sequences, probably derived from the ~260bp contaminating band present in all four experimental samples. Unfortunately, all four libraries

showed an extremely high PCR duplication rate (close to 99.9%) that strongly reduced, to just a few thousand, the number of unique reads aligned in each sample (Table 1).

To further analyze the results, I focused on the three mRNA targets identified by candidate approach in the Osk CLIP experiment (Fig. 9 and Fig. 14). I hoped that, even if the experiment was not optimal due to the high rate of PCR duplication in the libraries, it might be still possible to confirm the UV- and Osk-dependent enrichment of *nos*, *pgc* and *gcl* mRNAs and to infer information regarding the position of Osk binding on those RNAs.

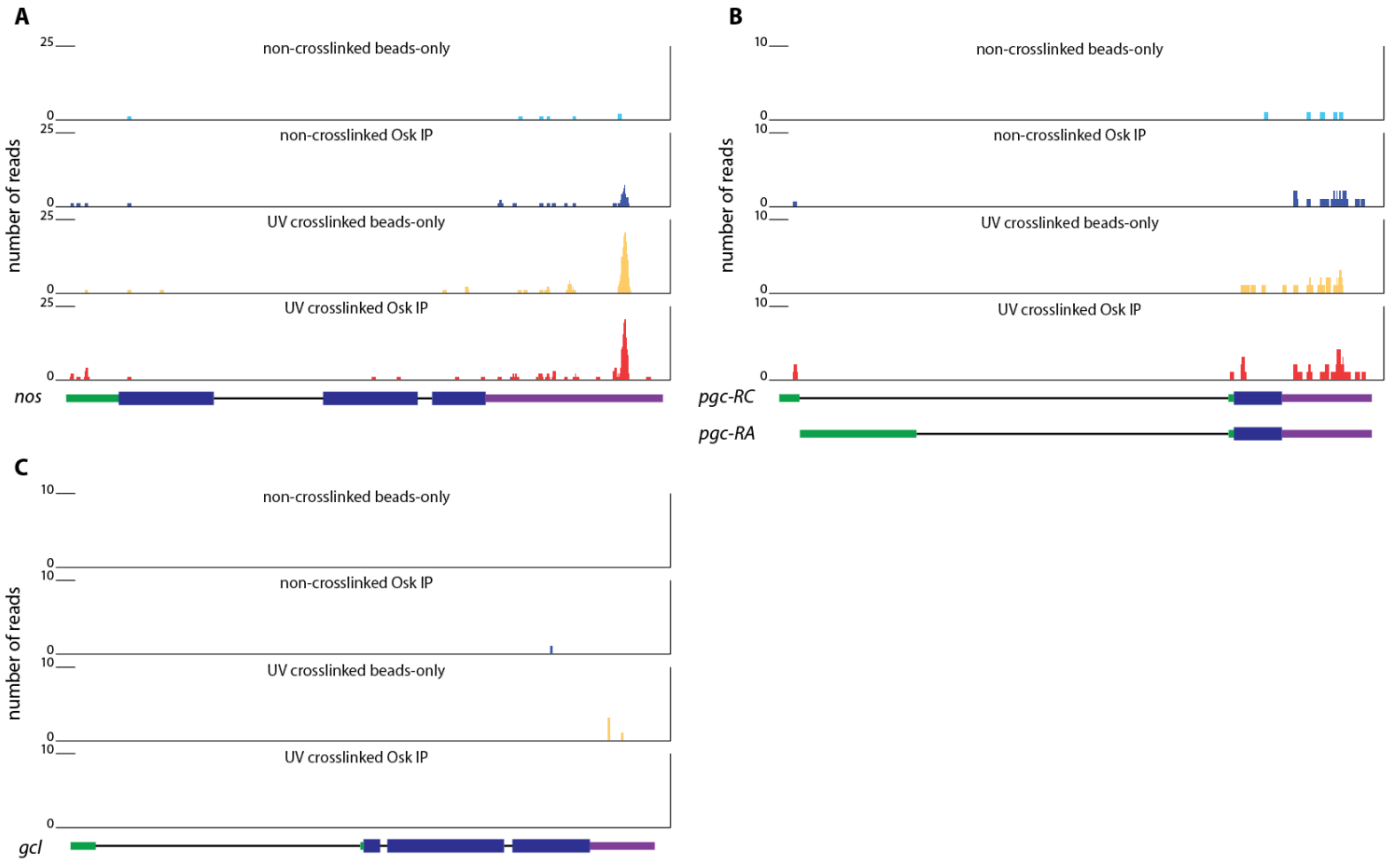

Figure 21: Osk iCLIP libraries alignment. Osk iCLIP libraries were sequenced on Illumina HighSeq 2000 instrument. After data processing the reads from the different libraries were aligned against *Drosophila* genome. Final data were visualized with Integrative Genomics Viewer software [197] [198]. The alignment results for three genes are showed: (**A**) *nos*, (**B**) *pgc* (**C**) *gcl*. In the schematic representation of the genomic sequences, green boxes represent 5'UTRs, blue boxes represent the coding regions, purple boxes represent 3'UTRs and black lines represent introns. In B, both splicing isoforms of *pgc* (*pgc-RA* and *pgc-RC*) are shown.

The results of sequence alignment for *nos* mRNA are shown in Figure 21A. In general, fewer reads could be mapped to *nos* mRNA in the non-crosslinked than in the UV crosslinked condition, and the vast majority of the reads mapped to the *nos* 3'UTR. The major cluster of reads in the 3'UTR is present in both UV crosslinked beads-only and Osk IP libraries with the same intensity. A smaller cluster of reads is present at the same position also in the non-crosslinked Osk IP sample. In the UV crosslinked Osk IP library, in addition to the major cluster of reads from the 3'UTR, several smaller clusters map to the 3'UTR and the 5'UTR of the *nos* sequence. Even if some of these are absent in the control samples (Fig. 21A), the clusters in the UV crosslinked Osk IP library contain too few reads (between 2 and 5) to be considered putative Osk deposition sites, and to conclude that *nos* mRNA was enriched in the UV crosslinked and Osk IP experimental sample.

A similar analysis was carried out for *pgc* mRNA (Fig. 21B). In this case, the number of reads aligned to *pgc* transcripts in the different conditions was generally lower than *nos* mRNA. Similar to *nos*, most of the reads map to the *pgc* 3'UTR and a few more to the 5'UTR. As in the case of *nos*, due to the low number of mappable reads, it was not possible to confirm the UV- and Osk-dependent enrichment of the *pgc* transcript and it was not possible to recognize putative deposition sites of Osk protein on the RNA (Fig. 21B).

The results for *gcl* mRNA were very poor. Only three reads could be mapped to *gcl* in the four iCLIP libraries and none from the UV crosslinked Osk IP sample (Fig. 21C). Therefore, was not possible to confirm *gcl* RNA as a binding partner of Osk protein *in vivo*. The other reads present in the iCLIP libraries mapped randomly throughout the *Drosophila* genome, without a specific enrichment for particular mRNAs other than rRNAs.

3.6. *nos* and *pgc* 3'UTRs are enriched in Oskar CLIP experiments

The results of the Osk iCLIP experiment shown in Figure 21 did not allow me to confirm Osk association with *nos*, *pgc* and *gcl* mRNAs. Given that in the case of *nos* and *gcl* mRNAs,

the majority of the reads mapped to the 3'UTR of both transcripts, I decided to try a different CLIP-based approach to investigate which regions of *nos* and *pgc* mRNAs might be bound by Osk.

To this end, I performed an Osk CLIP experiment on denatured embryo extracts, as described in the Paragraph 3.3. After RNA extraction and reverse-transcription using random hexamer primers, I analyzed the cDNAs by qPCR using primers that amplify short DNA fragments covering different regions of *nos* and *pgc* mRNAs. Most of the primers used in the experiment were designed to anneal in the 3'UTR of *nos* and *pgc* sequences (Fig. 22A and D), as the highest number of unique mappable reads in the Osk iCLIP libraries mapped in these positions (Fig. 21A and B).

The results of the experiment for *nos* mRNA are shown in Figure 22B and C. *A*ll the fragments tested were significantly enriched upon UV crosslinking and Osk IP, but to different degrees. Interestingly, the two fragments located at the 3' end of *nos* 3'UTR (*nos3'UTR7* and *nos3'UTR8*) were the most enriched (Fig. 22B).

One can speculate that, if the efficiency of reverse-transcription along the target mRNA were identical, all fragments produced under the UV crosslinked Osk IP condition should be enriched to the same degree. In reality, a number of factors might influence reverse-transcription such as nucleotide sequence, secondary structure or selective degradation of some mRNAs [199]. As a consequence, some fragments might be over or underrepresented in the pool of cDNA produced in the reaction. In order to control for the efficiency of reverse-transcription along *nos* mRNA, I measured the relative abundance of the different *nos* fragments, normalized with respect to the housekeeping protein encoding *Act42A* mRNA, in the input extracts. The graph in Figure 22C shows the results of the analysis. The first six fragments of *nos* mRNA (*nos5'UTR*, *nosEx2*, *nos3'UTR2*, *nos3'UTR3*, *nos3'UTR5* and *nos3'UTR6*) were all expressed at similar levels, and were between 2 and 4 fold more abundant than *Act42A* mRNA. Conversely, two fragments that mapped to the 3' end of *nos* 3'UTR (*nos3'UTR7* and *nos3'UTR8*) appeared underrepresented compared to the others. This

result was probably the consequence of the random hexamer primers used for reverse-transcription. Indeed, random hexamer primers anneal inefficiently to poly(A) stretches, resulting in underrepresentation of cDNA fragments covering mRNA regions close to the poly(A) tail.

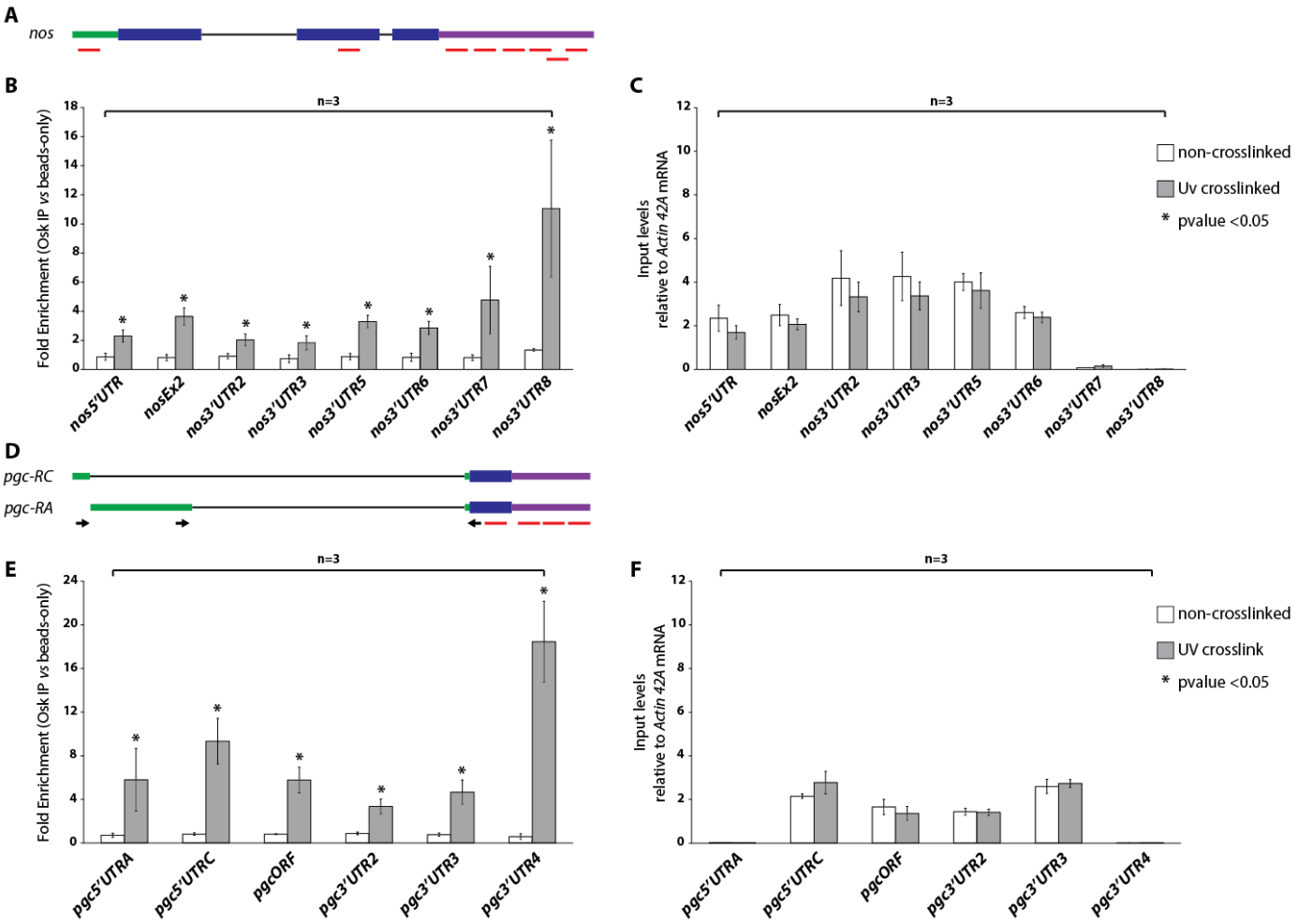

Figure 22: *nos* and *pgc* 3'UTRs are enriched in Oskar CLIP. Osk CLIP experiments were performed as described in Paragraph 3.3. The cDNAs were analyzed by qPCR for the enrichment of different fragments of both *nos* and *pgc* mRNAs. (**A** and **D**). Schematic representation of *nos* and *pgc* genomic regions. The two splicing isoforms of *pgc* (*pgc-RA* and *pgc-RC*) are shown. Green boxes represents 5'UTRs, blue boxes represent coding regions, purple boxes represent 3'UTRs and black lines represent introns. The red lines below the gene structure represent the DNA fragments amplified by qPCR. Black arrows represent the position of PCR primers used for amplification of the two 5'UTR variants of *pgc* (two isoform specific forward primers and a single reverse primer). (**B** and **E**). qPCR analysis of Osk CLIP experiment for *nos* and *pgc* mRNAs. For both UV crosslinked and non-crosslink conditions, the level of enrichment for the different fragments of *nos* and *pgc* measured in the Osk IP was normalized to the level of the beads-only control. Bars represent mean values ± standard deviation. p values were calculated using Student's t test. (**C** and **F**). Relative abundance of *nos* and *pgc* fragments

tested in B and E, measured in the input extracts. The levels were normalized compared to the housekeeping *Act42A* mRNA. Bars represent mean values ± standard deviation.

Taken together, the low relative abundance of *nos3'UTR7* and *nos3'UTR8* fragments in the input samples and the high degree of enrichment of those same fragments in the Osk CLIP experiment might indicate that *nos3'UTR7* and *nos3'UTR8* fragments were specifically enriched, in a UV- and Osk-dependent manner. One can therefore speculate that Osk might bind *in vivo* to the *nos* mRNA, at a position close to the poly(A) tail of the transcript. Indeed, if this were the case, this region of *nos* would preferentially co-precipitate with Osk upon UV crosslinking, whereas fragments located far from the crosslinking site(s) might be "preferentially" lost during the experiment procedure due to RNA degradation, which cannot be completely abolished.

pcg mRNA was analyzed in a similar manner and the results are shown in Figure 22E and F. All *pgc* fragments analyzed in the Osk CLIP experiment were significantly enriched in the UV crosslinked Osk IP condition and, similarly to *nos*, the fragment closest to the poly(A) tail (*pgc3'UTR4*) was the most enriched (Fig. 22E). Analysis of the relative abundance of the *pgc* fragments in the input material revealed that *pgc3'UTR4* was underrepresented compared to the others. As in the case of *nos* mRNA, one can speculate that Osk binds to *pgc* mRNA at a position near the poly(A) tail.

An additional *pgc* fragment is underrepresented in the input material: *pgc5UTRA* (Fig. 22F). There are two splicing isoforms of *pgc* mRNA, *pgc-RA* and *pgc-RC*, that possess different 5'UTRs (Fig. 22D). The qPCR analysis shown in Figure 22F revealed that the two isoforms are expressed at different levels in 0 to 2 hour old *Drosophila* embryos: *pgc-RC* is roughly 100 times more abundant than *pgc-RA*. Both isoforms are enriched in the Osk CLIP experiment in a UV- and Osk-dependent manner and, reflecting their relative abundance in the input, *pgc-RC* is enriched to a higher level than is *pgc-RA* (Fig. 22E). However, considering the different expression levels of the two *pgc* isoforms in the input material, and taking into account the data from the optimized Osk iCLIP experiment, in which no reads could be mapped

to the *pgc-RA* 5'UTR (Fig. 21B), I would not conclude that Osk binds to the 5'UTR of the *pgc-RA* isoform.

In conclusion, the Osk CLIP experiments shown in Figure 22 indicate that fragments close to the poly(A) tail of *nos* and *pgc* mRNAs are specifically enriched in a UV- and Osk-dependent manner. Unfortunately these results were not confirmed in the optimized Osk iCLIP experiment. Lacking additional experimental evidences in support of the results of the Osk CLIP experiments showed in Figure 22, I can only speculate that Osk might associates close to the poly(A) tail of *nos* and *pgc* mRNAs *in vivo*. These data, even if not fully conclusive, might represent a good starting point to further investigate Osk deposition site(s) on *nos* and *pgc* transcripts.

4. Discussion and Conclusions

The accumulation of Osk protein at the posterior pole of the *Drosophila* oocyte triggers assembly of the germ plasm of the animal. The pole plasm contains all the molecular factors, both proteins and RNAs, required for both germ cell formation and posterior patterning (Fig. 6) [47] [88]. The pole plasm inducing activity of Osk relies on its ability to recruit in a step wise process other pole plasm components. Among the pole plasm components, it has been shown that Osk directly interacts with Vas, Vls and Lasp proteins [113] [115] [116] [117] and that Osk activity is necessary for the posterior localization of several mRNAs such as *nos*, *cycB*, *pgc*, *gcl*, *osk* and *Hsp83* [49] [89] [137] [141] [182] [183].

Despite the large amount of information produced in more than 20 years of research, the molecular mechanism underling the pole plasm inducing activity of Osk is still not fully understood. Recently, two independent studies demonstrated that Osk associates with polyadenylated mRNAs in *Drosophila* embryos and indicated that Osk's RNA-binding activity might be exerted, at least in part, by the C-terminal OSK domain [109] [149]. The ability to interact with RNA might represent a crucial feature during pole plasm assembly. Osk might directly recruit some mRNAs at the posterior pole or it might exert regulatory functions in post-transcriptional processes such as mRNA stabilization and translation.

In order to investigate the RNA-binding activity of Osk, I applied the CLIP methodology [161], using UV light as a crosslinking agent to stabilize Osk-RNA interactions *in vivo*. I set up stringent Osk IP conditions in 0 to 2 hour old embryos (Fig. 8) and performed Osk CLIP experiments that I analyzed in a candidate approach, in which I tested by qPCR the UV- and Osk-dependent enrichment of a set of posteriorly localized mRNAs in *Drosophila* early embryo. Among the targets analyzed, I detected a significant enrichment of three mRNAs: *nos*, *pgc* and *gcl* (Fig. 9) [109], all three of which are involved in abdomen and/or germ cell formation. *nos* is not only a morphogen that acts in abdominal patterning, but is also required during PGC formation to inhibit somatic gene expression and mitosis [48] [133] [134] [135] [136]. *pgc* and

gcl are both required to maintain transcriptional quiescence of PGCs during early embryogenesis [138] [139] [145].

The data I obtained showing that Osk associates with *nos*, *pgc* and *gcl* mRNAs *in vivo* suggest that, by direct interaction, Osk might be involved in regulating its target mRNAs. It is noteworthy that Osk activity is required to activate *nos* mRNA translation at the posterior [59] [60]. It is tempting to speculate that, in the case of *nos* mRNA, Osk might interact directly with *nos* in such a manner as to promote its translation at the posterior pole. Only a small number of mRNAs were tested in the candidate approach, hence it is possible that Osk additionally associates with other posteriorly localized transcripts that were not analyzed.

To further characterize Osk's RNA-binding activity I decided to apply the iCLIP protocol, an unbiased transcriptome-wide method that should allow one to identify all the binding targets of a specific RBP, as well as the precise sites of protein binding on the mRNAs [174] (Fig. 13). Unfortunately, the Osk iCLIP experiment was not successful and I was not able to confirm any of the mRNAs identified in the Osk CLIP experiment.

Considering that the iCLIP approach has been used to study several RBPs *in vivo* [174] [175] [176] [177] [178] [179] [180], the negative results of the Osk iCLIP experiment that I performed might indicate that Osk is not a genuine RBP. The enrichment of *nos*, *pgc* and *gcl* mRNAs in the Osk CLIP experiment (Fig. 9) might be due to the residual interaction of Osk with other RBPs in the embryo extracts, rather than to crosslinking events occurring between Osk and these transcripts. Although the stringency of the IP conditions used in the experiment were such that I did not detect the co-IP of Vas and Smg (Fig. 9C and Fig. 13A), two proteins that have been shown to interact with Osk directly [60] [113] [115], I cannot exclude that residual amounts of these or other RBPs crosslinked to their target RNAs, below the detection limit of the western blot assay, co-IPed with Osk nevertheless.

To rule out that the UV- and Osk-dependent enrichment of *nos*, *pgc* and *gcl* mRNAs might be caused by residual interactions between Osk and other proteins associated with these mRNAs, I increased the stringency of the IP conditions by performing the Osk IP on denatured

embryo extracts (Fig. 14A). The denaturation applied during the *Drosophila* embryo extract preparation should consequently prevent interaction between Osk and other proteins, and allow me to exclusively IP Osk. The results of the Osk CLIP experiment from denatured embryo extracts show that *nos*, *pgc* and *gcl* mRNAs are enriched in a UV- and Osk-dependent manner (Fig. 14B). This demonstrates that Osk is UV crosslinked to *nos*, *pgc* and *gcl* mRNAs in 0 to 2 hour old *Drosophila* embryos. In the case of *pgc* mRNA, I measured a similar level of enrichment upon UV crosslinking and Osk IP from non-denatured and denatured experimental conditions (compare Fig. 9B and 14B). This suggests that *pgc* might be a direct binding target of Osk protein *in vivo*. In contrast, the low level of enrichment measured for *gcl* mRNA, even if statistically significant, might indicate an unspecific interaction with Osk (Fig. 14B). The enrichment of *nos* mRNA measured in denaturing conditions (Fig. 14B) was significantly lower than from the Osk CLIP experiment shown in Figure 9B. This might indicate that, in the non-denaturing Osk CLIP experiment, *nos* mRNA was not directly crosslinked to Osk but was crosslinked to some other protein(s) co-precipitating with Osk.

Although the Osk CLIP experiment performed on denatured embryo extracts showed that Osk can be UV crosslinked *in vivo* to *pgc* mRNA and to a lesser extent to *nos* and *gcl* mRNAs (Fig. 14), I cannot conclude that Osk is a genuine RBP. The reason for this is that the UV crosslinking step is performed on intact embryos in which Osk, localized within the pole plasm, is in close proximity to other pole plasm components. The interaction between Osk and other RBPs bound to their own target mRNAs might bring these mRNA molecules close enough to Osk for crosslinking to the protein upon UV light irradiation. The identification of putative Osk binding site(s) on the target mRNAs would allow me to test whether the RNA-binding activity of Osk detected *in vivo* is truly significant. As a genuine RBP, Osk might recognize specific sequence motif(s) or secondary structure element(s) within the target mRNA sequences. Once these are identified, I would design specific mutations that should abolish Osk recognition, either confirming or disproving the RNA-binding activity of the protein.

Since the unsuccessful result of the Osk iCLIP experiment shown in Figure 13 was not due to the lack of mRNAs crosslinked to the protein, other steps in iCLIP library preparation (Fig. 10) might be responsible for the negative results of the experiment. In particular, the enzymatic reactions performed before the RNA extraction step, namely partial RNA digestion, 3' end dephosphorylation and L3App adapter ligation, are crucial for generating a library of good quality. Considering the suboptimal amount of starting material used in the experiment (Fig. 11), insufficiently controlled and optimized reactions might negatively impact on iCLIP library preparation. I therefore decided to focus on the optimization of these enzymatic reactions.

I first established new conditions for the partial RNA digestion step. The new reaction conditions allowed me to reproducibly achieve optimal levels of RNA fragmentation and in general they allowed me to better control this step of the iCLIP protocol (Fig. 15).

Next, I analyzed the efficiency of the 3' end dephosphorylation step performed using the iCLIP protocol conditions. The data obtained suggested that this reaction was not a limiting step in the protocol, as PNK-mediated dephosphorylation worked efficiently under the original iCLIP experimental conditions (Fig. 16).

Finally, my analysis of the L3App adapter ligation step revealed that this reaction was the limiting step in the original iCLIP protocol (Fig. 17A and B). By testing several reaction buffer compositions (Fig. 17C and D) and different amounts of enzyme (Fig. 18A and B), I successfully increased the L3App adapter ligation efficiency over a wide range of RNA substrate amounts (Fig. 18C and D).

Once the iCLIP conditions were optimized, I repeated the Osk iCLIP experiment from denatured embryo extracts using my new protocol (Fig. 19). The iCLIP libraries produced using the optimized conditions contained 1000 time more cDNA material than the previous non-optimized Osk iCLIP experiment (compare Fig. 19C and Fig. 13B). Two factors might account for this: the higher efficiency of the optimized L3App ligation step (Fig.18) and the higher RNA level present in all four experimental conditions in the Osk CLIP assay from denatured embryo

extracts (Fig. 15B). Interestingly, in the optimized Osk iCLIP experiment, the library produced from the UV crosslinked and Osk IP sample contained more DNA material compared to the three control conditions (Fig. 19C, D and Fig. 20).

The analysis of the sequencing data revealed that the optimized iCLIP protocol improved the overall quality of the libraries. The percentage of reads aligned to the *Drosophila* genome ranged between 46% and 63%, depending on the specific experimental condition. Unfortunately the PCR duplication rate of the libraries was high (99.9%), strongly reducing the number of unique mappable reads (Table 1). Indeed, analysis of the alignment results for *nos*, *pgc* and *gcl* mRNAs revealed the negative outcome of the experiment. In the case of *nos* and *pgc*, the few reads aligned on the transcripts mainly clustered in the 3'UTR of the mRNAs. Unfortunately, the low number of mappable reads and the absence of striking differences between the experimental and the control conditions did not allow me to confirm a UV- and Osk-dependent enrichment of *nos* and *pgc* mRNAs, precluding identification of putative Osk binding sites (Fig. 21A and B). The results for *gcl* mRNA were very poor, and from the UV crosslinked Osk IP condition no reads could be mapped to *gcl* (Fig. 21C).

The negative outcome of the optimized Osk iCLIP experiment might be due to two different technical issues. On the one hand, an important negative contribution might be due to those steps of the protocol that I did not optimize, such as reverse-transcription, circularization and linearization (Fig. 10). It is possible that low efficiency of at least one of these steps might have negatively influenced the experiment. However, it is important to keep in mind that the same experimental conditions, including low efficiency ligation conditions were successfully applied by others studying different RBPs *in vivo* [174] [175] [176] [177] [178] [179] [180]. On the other hand, the suboptimal amount of RNA that co-IPed with Osk (Fig. 15B) might have been the major limitation of the experiment. In this regard, one should note that in one study, *in vitro* experiments are presented that would suggest that Osk has a low binding affinity for RNAs (in the low micromolar range) [109].

Considering the number of unique reads that I was able to map in the optimized Osk iCLIP experiment (a few thousand) (Table 1), it might be necessary to scale up the protocol at least 100 times to produce datasets rich enough to be informative. In the Osk iCLIP experiment, 0.5g of *Drosophila* embryos were used as the starting material for preparation of each sample. Therefore the amount of embryos (hundreds of grams), buffers, enzymes and other reagents required for such a scaled up experiment probably exceeds a rational ratio between cost and benefit, and another strategy should be envisioned.

My data suggest that the iCLIP protocol cannot be applied to the study of Osk-RNA interaction *in vivo*. Other approaches, easier from a technical point of view, such as a commercial kits designed for cDNA library preparation from low amounts of RNA, might be used to analyze the RNA interactome bound to Osk *in vivo*. Although they would constitute a first step, none of these experimental strategies provides the high resolution required to identify the binding site(s) of the protein on its target mRNAs. Even if the optimized iCLIP protocol I established did not allow me to carry out a full analysis of Osk-RNA interactions, it should be applicable for the study of other RBPs in *Drosophila* or other model organisms.

The finding that in the optimized Osk iCLIP experiment the majority of the reads mapped to *nos* and *pgc* mRNAs were clustered in the 3'UTR of both transcripts led me to try a different CLIP-based approach to narrow down which part of those mRNAs might be crosslinked to Osk *in vivo*. After performing an Osk CLIP experiment from denatured embryo extracts, as shown in Figure 14, I analyzed the extracted RNAs by qPCR using primers to amplify short stretches of nucleotides spanning the length of the *nos* and *pgc* mRNA sequences. The results of these experiments showed that, for both *nos* and *pgc* mRNAs, the 3'UTR fragments close to the poly(A) tail were more enriched in a UV- and Osk-dependent manner than were the other fragments tested (Fig. 22B and E). Although these data indicate a possible preference of Osk for association with the *nos* and *pgc* 3'UTRs, at this stage there are no additional data to support this hypothesis.

In conclusion, in this study I provided the first experimental evidence regarding the interaction between Osk and three mRNAs (*nos*, *pgc* and *gcl* mRNAs) localized at the posterior pole of live *Drosophila* embryos (Fig. 9) [109]. By performing the same experiment using denatured embryo extracts, I showed that, upon UV light irradiation, Osk is directly crosslinked to *nos*, *pgc* and *gcl* mRNAs, suggesting that Osk-RNA interaction might be mediated by an intrinsic RNA-binding activity of the protein (Fig. 14). To further study Osk-RNA interaction *in vivo*, I applied the iCLIP protocol, an unbiased transcriptome-wide strategy. After a first negative outcome of the experiment, I focused on optimization of three major steps in iCLIP library preparation: partial RNA digestion, 3' end dephosphorylation and L3App adapter ligation. My efficiency test experiment showed that the 3' end dephosphorylation reaction is not the limiting step in the iCLIP protocol (Fig. 16). Conversely the L3App adapter ligation step, due to low reaction efficiency, was suboptimal for the library preparation process (Fig. 17A and B). By changing the reaction conditions, I managed to increase the efficiency of this step (Fig. 17C, D and Fig. 18). In spite of this improvement, performance of the Osk iCLIP experiment using the optimized protocol nevertheless again resulted in a negative outcome (Fig. 21). Finally, by using a different, CLIP-based approach, I showed that *nos* and *pgc* 3'UTR fragments close to the poly(A) tail of both transcripts are preferentially enriched upon UV irradiation and Osk IP (Fig. 22). These data might indicate an interaction of Osk with these particular regions of *nos* and *pgc* mRNAs.

Although I found that Osk associates *in vivo* with *nos*, *pgc* and *gcl* mRNAs, the next steps into the project will require knowing precisely where on the target mRNAs Osk is bound. This information would allow me to design specific mutations to abolish Osk interaction with a specific mRNA and analyze the biological relevance of the Osk RNA-binding activity I detected. It would be possible to address whether and which mRNA post-transcriptional processes Osk might influence, such as localization, translation or stability. In addition, knowledge of the precise Osk binding sequences in the mRNAs would make it possible to study how the protein's RNA-binding activity influences posterior patterning, germ cell formation and germ cells viability.

5. Materials and methods

5.1. UV crosslinking in *Drosophila* embryos *in vivo*

UV crosslinked and non-crosslinked embryo extracts were prepared from *Drosophila melanogaster* Oregon R wild-type animals. Embryos were collected between 0 and 2 hours after egg laying on 1% agar and 30% apple juice plates supplemented with yeast paste, dechorionated in a 50% sodium hypochlorite solution for 4' and extensively washed with water. Dechorionated embryos were resuspended in 15ml of Embryo Washing Buffer (120mM NaCl; 0.04% Triton-X100), spread on 600cm^2 cell culture plates and irradiated with UV light (254nm; 1J*cm^{-2}) on ice in a Stratalinker 2400 (Stratagene) instrument; non-crosslinked control embryos were left on ice.

5.2. Embryo extracts preparation

UV crosslinked and non-crosslinked embryos were collected and resuspended in 1 volume (1g in 1ml) of specific Lysis Buffer (see below).

5.2.1. Non-denatured embryo extracts

Embryos were resuspended in Lysis Buffer 1 (LB1) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.5% LiDS; 1mM EDTA; 1mM DTT) [148] supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific), then lysed using a 15ml Douncer homogenizer (Wheaton). The extracts were centrifuged 10' at 3500g at 4ºC, the supernatants collected were frozen in liquid N_2 and stored at -80ºC for later use. After thawing, extracts were centrifuged 4' at 16000g and the supernatant were diluted 1:10 (1g in 10ml) in LB1 [148] supplemented with complete EDTA-free proteinase inhibitor cocktail

(Roche) and RiboLock RNase inhibitor (Thermo Scientific). 5ml aliquots (0.5g of embryos) were prepared and stored on ice.

5.2.2. Non-denatured embryo extracts supplemented with Na deoxycholate

Embryo lysis was performed as described in 5.2.1. After thawing and centrifugation (4' at 16000g) the supernatants were diluted 1:10 (1g in 10ml) in Lysis Buffer 2 (LB2) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.5% LiDS; 0.5% Na deoxycholate; 1mM EDTA; 1mM DTT) supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific). 5ml aliquots (0.5g of embryos) were prepared and stored on ice.

5.2.3. Denatured embryo extracts

After collection (see 5.1), embryos were resuspended in Lysis Buffer Urea (LBU) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.5% LiDS; 8M urea; 1mM EDTA; 1mM DTT) supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific). Lysis was performed in a 15ml Douncer homogenizer (Wheaton). The extracts were centrifuged 10' at 16000g at 4ºC and the supernatants collected were diluted 1:10 (1g in 10ml) using LB2. 5ml aliquots (0.5g of embryos) of the extracts were prepared, frozen in liquid N_2 and stored at -80ºC.

5.3. CLIP

5.3.1. Immunoprecipitation

Different IP conditions were used in the experiments presented in this work and described in the next paragraphs. For each IP, 20µl of protein G sepharose beads (GE Healthcare) were used.

5.3.1.1. Immunoprecipitation from non-denatured embryo extracts

Beads were washed 3 times for 5' on a rotating wheel in the cold (8ºC) with Coupling Buffer (CB) (20mM Tris-HCl pH 7.5; 150mM NaCl; 0.5mM EDTA; 0.2% Nonidet P-40). 50µl of anti-Osk antibody [89] were added to beads together with 450µl of CB and the samples were incubated overnight on a rotating wheel in the cold (8ºC). Beads-only control samples were incubated in absence of the antibody. Beads were washed 3 times with LB1 [148] supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific), for 5' in the cold (8ºC) and were added to 5ml aliquots of embryo extract prepared according to 5.2.1. Before beads addition, two aliquots of 50µl each were collected from the extracts to analyze proteins and RNAs in the input. IPs were incubated 1 hour at room temperature on a rotating wheel. After incubation, the beads were washed 3 times (1ml) with LB1 [148], 3 times with (1ml) Washing Buffer 1 (WB1) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.1% LiDS; 1mM EDTA; 1mM DTT) [148] and 3 times (1ml) with Washing Buffer 3 (WB3) (20mM Tris-HCl pH 7.5; 150mM LiCl; 0.1% LiDS; 1mM EDTA; 1mM DTT). All buffers were supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific), and all the washes were performed for 10' on a rotating wheel in the cold (8ºC). For the experiment in Figure 8 (lanes 2, 3, 7 and 8), beads

were resuspended in 20µl of 2X NuPage LDS sample buffer (Thermo Scientific) and incubated 10' at 95ºC for protein elution.

5.3.1.2. Immunoprecipitation from non-denatured embryo extracts supplemented with Na deoxycholate

Beads preparation and antibody coupling were performed as described in 5.3.1.1. Beads were washed 3 times with LB2, supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific), for 5' in the cold (8ºC), then were added to 5ml aliquots of embryo extracts prepared as described in 5.2.2. Before beads addition, two aliquots of 50µl each were collected for the analysis of proteins and RNAs in the input. IP incubation was performed as described in 5.3.1.1. Beads were washed 3 times (1ml) with LB2, 3 times (1ml) with Washing Buffer 2 (WB2) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.1% LiDS; 0.1% Na deoxycholate; 1mM EDTA; 1mM DTT) and 3 times (1ml) with WB3. All buffers were supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific) and all the washings were performed for 10' on a rotating wheel in the cold (8ºC). For the experiment in Figure 8 (lanes 4, 5, 9 and 10) beads were resuspended in 20µl of 2X NuPage LDS sample buffer (Thermo Scientific) and incubated 10' at 95ºC for protein elution. For experiments in Figures 9 and 13, 1% of the beads were resuspended in 10µl of 2X NuPage LDS Sample Buffer (Thermo Scientific) for western blot analysis.

5.3.1.3. Immunoprecipitation from denatured embryo extracts

Beads were prepared and coupled to anti-Osk antibody as described in 5.3.1.1. Beads were washed 3 times with LB2 supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific), for 5' in the cold (8ºC).

After washings, the beads were added to embryo extracts prepared as described in 5.2.3. 50µl aliquots were collected from the extracts before bead addition for protein and RNAs analysis. IPs were incubated overnight (16 hours) on a rotating wheel in the cold (8ºC). Beads were washed 3 times (1ml) with Washing Buffer Urea 1 (WBU1) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.5% LiDS; 0.8M urea; 1mM EDTA; 1mM DTT), 3 times (1ml) with Washing Buffer Urea 2 (WBU2) (20mM Tris-HCl pH 7.5; 500mM LiCl; 0.1% LiDS; 0.8M urea; 1mM EDTA; 1mM DTT) and 3 times (1ml) with Washing Buffer 4 (WB4) (20mM Hepes pH 7.8; 125mM NaCl; 0.1% Nonidet P-40). All buffers were supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RiboLock RNase inhibitor (Thermo Scientific) and all the washings were performed for 10' on a rotating wheel in the cold (8ºC). 1% of the beads were resuspended in 10µl of 2X NuPage LDS Sample Buffer (Thermo Scientific) for protein elution and western blot analysis (Fig.14 and Fig. 19).

5.3.2. RNA extraction

At the end of the Osk IP processes described in 5.3.1.2 and 5.3.1.3, beads were resuspended in 100µl of Proteinase K Buffer (PKB) (10mM Tris-HCl pH 7.5, 150mM NaCl; 0.2% SDS; 10mM EDTA; 0.5mM DTT; 5mM $CaCl_2$) [148], supplemented with 80µg of Proteinase K (Promega) and incubated 1 hour at 50ºC in a Thermomixer C (Eppendorf). RNA extraction was performed using TRIzol LS reagent (Invitrogen) according to the manufacturer's protocol and extracted RNAs were resuspended in 6µl diethyl pyrocarbonate (DEPC) treated H_2O. 50µl aliquots from input extracts were supplemented with 40µg of Proteinase K (Promega) and incubated 1 hour at 50ºC. RNA was extracted using TRIzol LS (Invitrogen), treated with Turbo DNase (Invitrogen) according to manufacturer's instructions and again extracted using Phenol/Chloroform/isoamyl alcohol (AppliChem). The RNA was ethanol precipitated overnight at -20ºC in presence of 0.3M Na acetate (pH 5.0) and 15µg of GlycoBlue

(Invitrogen). The RNA was resuspended in 50µl DEPC treated H_2O and RNA concentration was measured using a NanoDrop 8000 instrument (Thermo Scientific).

5.3.3. Reverse-transcription

Reverse-transcription was performed using a SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen) using random hexamer primers, following the manufacturer's instructions. 1µg of total RNA extracted from UV crosslinked and non-crosslinked inputs were used for cDNA synthesis, whereas all the RNA extracted from the IP samples was used in the reaction.

5.3.4. Quantitative real-time PCR

qPCR analysis was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems). Each reaction was prepared in 11µl total volume containing: 1.1µl cDNA; 0.33µl primers mix (10µM); 5.5µl 2X SYBR Green PCR Master Mix (Applied Biosystems); 4.07µl DEPC treated H2O. 10µl of the mix were used for the PCRs. The PCRs were performed using the following conditions:

10' at 95ºC - {15'' at 95ºC - 15'' at (see Table 2) - 1' at 60ºC}.

45 total PCR cycles (steps between {---}).

Results were visualized with StepOne software v2.3 (Applied Biosystems) and analyzed using the ΔΔCt method. All primers were purchased from Eurofins Genomics.

Primer name	Sequence 5'-3'	Ann. T	Figure
nos2 RT Fw	AATCTCGGCGTGGGAATGGG	55ºC	9, 14
nos2 RT Rev	GGGCACAGCACTCGGTTAAAG	55ºC	9, 14
orb1 RT Fw	CGAGATTAGGACCACGAAGTTCAC	55ºC	9
orb1 RT Rev	CTCTGCGACTTGGAGTTGCG	55ºC	9
CycB2 RT Fw	CGAGGACGAGCACCATACG	55ºC	9
CycB2 RT Rev	GCGAGTAGAAGGTCAGAGTGG	55ºC	9
pgc2 RT Fw	GAAGACAGCTCCTGCGAGGG	55ºC	9, 14
pgc2 RT Rev	CATCTATCCGCGATGACCGGC	55ºC	9, 14
gcl2 RT Fw	CGGCACATTCACAACCACTG	55ºC	9, 14
gcl2 RT Rev	CCAGGCTCAAGCAACATACG	55ºC	9, 14
osk4 RT Fw	GCGGAGTCGAACTGGTGCTACC	55ºC	9, 14
osk4 RT Rev	GTGGGCGTCATTTCGTGGCG	55ºC	9, 14
Hsp83-1 RT Fw	ATACACAATTTACTTGCGTTTC	55ºC	9
Hsp83-1 RT Rev	CGACAACTGCTCTTGAATG	55ºC	9
bcd1 RT Fw	TGCTATGCCAAAGTGTCTG	55ºC	9, 14
bcd1 RT Rev	TGCTCTTCTTAGTGATGTAAATAG	55ºC	9, 14
Act42A1 RT Fw	GGCTGGGCGTGGTCGTTC	55ºC	9, 14, 22
Act42A1 RT Rev	AAGTGTTGTGCGCCTGCTCCC	55ºC	9, 14, 22
nos5UTR Fw	GAATTGCCGTACGCTTCGCAG	55ºC	22
nos5UTR Rev	GGGTTAAAATCGTGACGCAGAGGC	55ºC	22
nosEx2 Fw	CATCA CAACAGCAGGGGCTG	55ºC	22
nosEx2 Rev	GTAGTTGTTCAGCCAGTGGGTGG	55ºC	22
nos3UTR2 Fw	ACTTGTTCAATCGTCGTGGC	55ºC	22
nos3UTR2 Rev	GACCAGTTGCAGACCAATTCC	55ºC	22
nos3UTR3 Fw	CGTAAAATGTTCGCTTGCGG	55ºC	22
nos3UTR3 Rev	CAGTTCGAAATTCAAAGTGTTCC	55ºC	22
nos3UTR5 Fw	GACATTCCGCACTTTTTG	50ºC	22
nos3UTR5 Rev	GAAATGAATACTTGCGATACATG	50ºC	22
nos3UTR6 Fw	CATGTATCGCAAGTATTCATTTC	50ºC	22
nos3UTR6 Rev	GAAAAATTGATCAATGGTAAAC	50ºC	22
nos3UTR7 Fw	GTTTACCATTGATCAATTTTTC	55ºC	22
nos3UTR7 Rev	CATTTTCAGAATATGTGTACAC	55ºC	22
nos3UTR8 Fw	GTGTACACATATTCTGAAAATG	50ºC	22
nos3UTR8 Rev	CCTTAAATTGTAACCATTTC	50ºC	22
pgc5UTRA2 Fw	GTGGAAGCCAGGTGCATCTG	55ºC	22
pgc5UTRC2 Fw	CAAGAGAACAAGTTGAGCGTGGC	55ºC	22
pgc5UTR Rev	TCGCACATTTTCGGGTCTTC	55ºC	22
pgcORF Fw	GAAGACAGCTCCTGCGAGGG	55ºC	22
pgcORF Rev	GCAACTCCTCGCGCACTTGATG	55ºC	22
pgc3UTR2 Fw	GCCACATGAGTGGAACATCGTG	55ºC	22
pgc3UTR2 Rev	CATACGATCGCCTCTGAAGGCTG	55ºC	22
pgc3UTR3 Fw	GTCCGTGTATTCAAATGTTTGC	55ºC	22
pgc3UR3 Rev	GACAACCATATGCGATTGATTGG	55ºC	22
pgc3UTR4 Fw	GTATTCAAGAACAAGTAGGGAAG	55ºC	22
pgc3UTR4 Rev	GAACGATTGCGAATCGAAAATA	55ºC	22

Table 2: List of primers used for qPCR analysis

5.4. Radioactive labeling of protein-RNA complexes

Osk IP was performed as described in 5.3.1.2. At the end of the washing steps, beads were resuspended in 1ml of WB3. 10µl of a series of dilutions (1:50, 1:500, 1:1000 and 1:2000) of a 100U/µl stock of RNase I (Invitrogen) were added to the samples (see Fig. 12 for details). In parallel, control samples were not treated with RNase I. The samples were incubated 3' at 37ºC at 1100rpm in a Thermomixer C (Eppendorf) instrument and subsequently washed 2 times (1ml) with 1X PNK Reaction Buffer B (50mM Tris-HCl pH 6.4; 18mM $MgCl_2$; 5mM DTT; 0.1mM spermidine; 0.1mM ADP) (Thermo Scientific). After washings, beads were supplemented with 20µl of T4 PNK Reaction Mix: 2µl 10X PNK Reaction Buffer B (Thermo Scientific); 2µl T4 PNK (10U/µl) (Thermo Scientific); 1µl RiboLock RNase inhibitor (Thermo Scientific); 1µl gamma ^{32}P ATP (10mCi/ml) (Hartmann Analytic); 14µl DEPC treated H_2O. Reactions were incubated 30' at 37ºC at 1100rpm in a Thermomixer C (Eppendorf). At the end of the incubation, beads were resuspended in 10µl of 2X NuPage LDS Sample Buffer (Thermo Scientific) and incubated 10' at 75ºC. Proteins were analyzed by SDS-PAGE and western blot as described in 5.10.

5.5. iCLIP

Osk IP was performed as described in 5.3.1.2 except for the last washing step, which was performed using WB4 (1ml, 3 times). All buffers were supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche), RiboLock RNase inhibitor (Thermo Scientific) and 0.02mg/ml of heparin (Sigma Aldrich). At the end of the washing steps, beads were resuspended in 1ml of WB4 (without heparin) and 10µl of a 1:1000 dilution of RNase I (stock 100U/µl) (Invitrogen) were added. Samples were incubated 3' at 37ºC at 1100rpm in a Thermomixer C (Eppendorf) and subsequently washed 2 times with 1ml of WB4 (supplemented with 0.02mg/ml of heparin) and 2 times with 1ml of WB4 (without heparin). The washings were performed for 5' on a rotating wheel in the cold (8ºC). Beads were equilibrated

2 times in 1ml of Dephosphorylation Buffer for Equilibration (DBE) (70mM Tris-HCl pH 6.5; 10mM $MgCl_2$; 5mM DTT) [186] then supplemented with 40µl of Dephosphorylation Reaction mix: 38µl Dephosphorylation Reaction Buffer (DRB) (74.1mM Tris-HCl pH 6.5; 10.6mM $MgCl_2$; 5mM DTT); 1µl T4 PNK (10U/µl) (Thermo Scientific); 1µl RiboLock RNase inhibitor (Thermo Scientific). Samples were incubated 30' at 37ºC at 1100rpm in a Thermomixer C (Eppendorf) and subsequently washed 2 times with 1ml of WB4 (supplemented with 0.02 mg/ml of heparin) and 2 times with 1ml of WB4 (not supplemented with heparin). The washings were performed for 5' on a rotating wheel in the cold (8ºC). Beads were equilibrated 2 times with 1ml of Ligation Buffer for Equilibration (LBE) (50mM Tris-HCl pH 7.5; 10mM $MgCl_2$; 10mM DTT) [186] then supplemented with 40µl of Ligation Reaction mix: 27µl Ligation Reaction Buffer (75.3mM Tris-HCl pH 7.5; 15mM $MgCl_2$; 15mM DTT); 1µl T4 RNA ligase 2 truncated (200U/µl) (New England BioLabs); 3µl L3App adapter (20µM) (Sigma Aldrich); 1µl RiboLock RNase inhibitor (Thermo Scientific); 8µl PEG400 (20% final concentration) (Sigma Aldrich). Reactions were incubated overnight (16 hours) at 16ºC at 1100rpm in a Thermomixer C (Eppendorf) then washed 2 times with 1ml of WB4 (supplemented with 0.02mg/ml of heparin) and 2 times with 1ml of WB4 (without heparin). The washings were performed for 5' on a rotating wheel in the cold (8ºC). Beads were resuspended in 400µl of PKB [148], supplemented with 200µg of Proteinase K (Promega) and incubated 1 hour at 50ºC at 1100rpm in a Thermomixer C (Eppendorf). RNA-L3App were extracted using Phenol/Chloroform/isoamyl alcohol (AppliChem). Products were ethanol precipitated overnight at -20ºC in presence of 0.3M Na acetate (pH 5.0) and 15µg of GlycoBlue (Invitrogen) then resuspended in 6µl DEPC treated H_2O. For reverse-transcription, samples were transferred to 0.2ml PCR tubes and were supplemented with: 1µl Annealing Buffer (from Invitrogen SuperScript III First-Strand Synthesis SuperMix kit); 1µl RT primers mix (see Table 3). Samples were incubated in a C1000 Touch Thermal Cycler (Biorad): 5' at 70ºC - 5' at 25ºC; followed by addition of: 10µl 2X Reaction Mix; 2µl Superscript III (both from Invitrogen SuperScript III First-Strand Synthesis SuperMix kit). Samples were additionally incubated: 20' at 25ºC - 30' at 42ºC - 50' at 50ºC - 3' at 90ºC then cooled on ice. cDNA fragments were purified using QIAquick PCR Purification Kit (Qiagen) following the

manufacturer's instructions and were eluted in 20µl of DEPC treated H_2O. Samples were transferred to 0.2ml PCR tubes and were supplemented with 20µl of 2X Circularization mix: 4µl 10X CircLigase II Reaction Buffer (Epicentre); 0.4µl $MnCl_2$ (50mM) (Epicentre); 0.6µl CircLigase II ssDNA Ligase (100U/µl) (Epicentre); 15µl DEPC treated H_2O. Samples were incubated in a C1000 Touch Thermal Cycler (Biorad): 1 hour at 60ºC and subsequently supplemented with 17µl of Oligo Annealing mix: 6µl NEBuffer 4 (New England BioLabs); 1.5µl Cut Oligo (10µM) (Sigma Aldrich); 9.5µl DEPC treated H_2O. Reactions were additionally incubated 4' at 85ºC, then slowly cooled down (T decreased 1ºC/10'') to 25ºC. After addition of 3µl of BamHI HF (20U/µl) (New England BioLabs), samples were incubated 30' at 37ºC. Reaction products were purified using QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions and were eluted in 20µl of DEPC treated H_2O.

Oligo name	Sequence 5'-3'	Sample
L3App adapter	App/AGATCGGAAGAGCGGTTCAG/ddC	all
RT10Clip	nnGACCnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	non-crosslinked beads-only
RT13Clip	nnTCCGnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
RT4Clip	nnAGGTnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	non-crosslinked Osk IP
RT8Clip	nnCATTnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
RT2Clip	nnACAAnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	Uv crosslinked beads-only
RT7Clip	nnCTAAnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
RT3Clip	nnATTGnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	UV crosslinked Osk IP
RT9Clip	nnGCCAnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
Cut Oligo	GTTCAGGATCCACGACGCTCTTC	all

Table 3: List of oligos used in the Oskar iCLIP experiment

5.6. RNase A- and T1-mediated digestion

1ml aliquots of embryo extracts prepared as described in 5.2.3 were incubated 3 hours on a rotating wheel in the cold (8ºC), in presence of different amounts of a mix of RNase A (from bovine pancreas) (Sigma Aldrich) and RNase T1 (from *Aspergillus oryzae*) (Sigma Aldrich). Stock mix concentration: 4U/µl (see Fig. 15 for the amounts of RNase A and T1 used). After digestion RNA was extracted as described in 5.3.2.

5.7. Dephosphorylation efficiency test

For each sample, 20µl of protein G sepharose beads (GE Healthcare) were used. Beads were washed 2 times for 5' on a rotating wheel in the cold (8ºC), with 1ml of CB then equilibrated 2 times with 1ml of DBE [186]. Beads were supplemented with 40µl of different reaction mixes according to the experimental condition. Experiment: 37µl DRB; 1µl oligo 1-P (diluted according to the final amount required in the experiment, see Fig. 16) (Integrated DNA Technologies); 1µl T4 PNK (10U/µl) (Thermo Scientific); 1µl RiboLock RNase inhibitor (Thermo Scientific). Positive Control: 38µl DRB; 1µl oligo 1 (diluted according to the final amount required in the experiment, see Fig. 16) (Integrated DNA Technologies); 1µl RiboLock RNase inhibitor (Thermo Scientific). Negative Control: 38µl DRB; 1µl oligo 1-P (diluted according to the final amount required in the experiment, see Fig. 16) (Integrated DNA Technologies); 1µl RiboLock RNase inhibitor (Thermo Scientific). For oligo sequences see Table 4. Beads were incubated 30' at 37ºC at 1100rpm in a Thermomixer C (Eppendorf), RNA was extracted using TRIzol LS reagent (Invitrogen) according to the manufacturer's protocol and resuspended in 5µl DEPC treated H_2O. L3App adapter (Sigma Aldrich) ligation was performed using T4 RNA ligase 2 truncated (New England BioLabs) and products were purified using QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions (eluted in 6µl DEPC treated H_2O). Reverse-transcription using RT Clip primers 3 and 9 (see Table 3)

was performed as described in 5.5; cDNAs were purified using QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions and were eluted in 20µl DEPC treated H_2O.

Oligo name	Sequence 5'-3'
oligo 1	NNNNNNGGCCUCGUGAAAUCCCGUUAGUAAAAGGNNNNNN
oligo 1-P	NNNNNNGGCCUCGUGAAAUCCCGUUAGUAAAAGGNNNNNN-P

Table 4: List of oligos used in the dephosphorylation efficiency test experiments

5.8. L3App adapter ligation efficiency test

Different amounts of oligo 1 (Integrated DNA Technologies) (see Table 4) were radioactively labeled at the 5' end. 20µl labeling reaction composition: 2µl 10X PNK Reaction Buffer A (Thermo Scientific); 1µl oligo 1 (Integrated DNA Technologies) (diluted according to the final amount required in the experiment, see Fig. 17, Fig. 18 and Table 4); 1µl T4 PNK (10U/µl) (Thermo Scientific); 1µl RiboLock RNase inhibitor (Thermo Scientific); 1µl gamma ^{32}P ATP (10mCi/ml) (Hartmann Analytic); 14µl DEPC treated H_2O. Reactions were incubated 30' at 37ºC in a 2720 Thermal Cycler (Applied Biosystems), then RNA was extracted using TRIzol LS reagent (Invitrogen) following the manufacturer's instructions and was resuspended in 4µl DEPC treated H_2O. 20µl of protein G sepharose beads (GE Healthcare) were used for each reaction and were initially washed 2 times for 5' on a rotating wheel in the cold (8ºC) with 1ml of CB. Equilibration was performed using different buffers, according to the specific experimental conditions. For iCLIP conditions (Fig. 17): 2 times with 1ml of LBE [186]. For Commercial conditions (Fig. 17 and Fig. 18): 2 times with 1ml of 1X T4 RNA Ligase Reaction Buffer (50mM Tris-HCl pH 7.5; 10mM $MgCl_2$; 1mM DTT) (New England BioLabs). For eCLIP conditions (Fig. 17): 2 times with 1ml of 1X eCLIP Ligation Buffer (50mM Tris-HCl pH 7.5; 10mM $MgCl_2$) [195]. Ligation reactions were performed in 40µl total volume with different composition according to specific experiments. iCLIP condition (Fig. 17): 23µl Ligation

Reaction Buffer; 1µl T4 RNA ligase 2 truncated (200U/µl) (New England BioLabs); 3µl L3App adapter (20µM) (Sigma Aldrich); 1µl RiboLock RNase inhibitor (Thermo Scientific); 8µl PEG400 (20% final concentration) (Sigma Aldrich); 4µl radioactive labeled oligo 1. Commercial condition (Fig. 17 and Fig. 18): 4µl 10X T4 RNA Ligase Reaction Buffer (New England BioLabs); Xµl T4 RNA ligase 2 truncated (200U/µl) (according to experimental conditions, see Fig. 17C, Fig. 18A and C) (New England BioLabs); 3µl L3App adapter (20µM) (Sigma Aldrich); 1µl RiboLock RNase inhibitor (Thermo Scientific); 14.4µl 50% PEG8000 (18% final concentration) (New England BioLabs); 1.3µl 100% DMSO (3.2% final concentration) (AppliChem); 4µl radioactive labeled oligo 1; Xµl DEPC treated H_2O to 40µl final volume. eCLIP conditions (Fig. 17): 4µl 10X eCLIP Ligation Buffer [195]; 1µl T4 RNA ligase 2 truncated (200U/µl) (New England BioLabs); 3µl L3App adapter (20µM) (Sigma Aldrich); 1µl RiboLock RNase inhibitor (Thermo Scientific); 14.4µl 50% PEG8000 (18% final concentration) (New England BioLabs); 1.3µl 100% DMSO (3.2% final concentration) (AppliChem); 4µl radioactive labeled oligo 1; 13.3µl DEPC treated H_2O. In the control reactions (no L3App adapter and no T4 RNA Ligase 2), the reaction volume of the omitted component was replaced by DEPC treated H_2O. Samples were incubated overnight (16 hours) at 16ºC at 1100rpm or 2 hours at 25ºC at 1100rpm in a Thermomixer C (Eppendorf). RNA oligos were extracted using TRIzol LS reagent (Invitrogen) following the manufacturer's instructions, resuspended in 5µl DEPC treated H_2O and supplemented with 5µl 2X RNA Loading Dye (New England BioLabs).

5.9. Optimized iCLIP

Osk IP was performed as described in 5.3.1.3. During the last 3 hours of IP incubation, the extracts were supplemented with $8*10^{-5}$U/µl of RNase A and T1 mix (Sigma Aldrich) and incubated on a rotating wheel in the cold (8ºC). All buffers were supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche), RiboLock RNase inhibitor (Thermo Scientific). WBU1 and WBU2 were also supplemented with 0.02mg/ml of heparin (Sigma Aldrich). Beads were equilibrated 2 times in 1ml of DBE [186] then supplemented with 40µl of

Dephosphorylation Reaction mix: 38µl DRB; 1µl T4 PNK (10U/µl) (Thermo Scientific); 1µl RiboLock RNase inhibitor (Thermo Scientific). Samples were incubated 30' at 37ºC at 1100rpm in a Thermomixer C (Eppendorf) then washed 2 times with 1ml of WBU2 and 2 times with 1ml of WB4. The washings were performed 10' on a rotating wheel in the cold (8ºC). Beads were equilibrated 2 times with 1ml of 1X T4 RNA Ligase Reaction Buffer (New England BioLabs) then were supplemented with 40µl of Commercial Ligation mix: 4µl 10X T4 RNA Ligase Reaction Buffer (New England BioLabs); 5µl T4 RNA ligase 2 truncated (200U/µl) (New England BioLabs); 3µl L3App adapter (20µM) (Sigma Aldrich); 1µl RiboLock RNase inhibitor (Thermo Scientific); 14.4µl 50% PEG8000 (18% final concentration) (New England BioLabs); 1.3µl 100% DMSO (3.2% final concentration) (AppliChem); 11.3µl DEPC treated H_2O. Samples were incubated overnight (16 hours) at 16ºC at 1100rpm in a Thermomixer C (Eppendorf) then washed 2 times with 1ml of WBU2 and 2 times with 1ml of WB4 (10' on a rotating wheel in the cold at 8ºC). The other steps of the protocol were performed as described in 5.5. For the RT Clip primers used for reverse-transcription see Table 5.

Oligo name	Sequence 5'-3'	Sample
RT2Clip	nnACAAnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	non-crosslinked beads-only
RT7Clip	nnCTAAnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
RT3Clip	nnATTGnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	non-crosslinked Osk IP
RT9Clip	nnGCCAnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
RT4Clip	nnAGGTnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	UV crosslinked beads-only
RT8Clip	nnCATTnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	
RT10Clip	nnGACCnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	UV crosslinked Osk IP
RT13Clip	nnTCCGnnnAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	

Table 5: List of oligos used in the optimized Oskar iCLIP experiment

5.10. SDS-PAGE and western blot

Protein samples collected as described in 5.3.1.1, 5.3.1.2 and 5.3.1.3 were separated on NuPAGE 4-12% polyacrylamide Bis-Tris gels (Invitrogen). Gels were run in 1X NuPAGE MOPS SDS Running Buffer (Invitrogen) using an XCell SureLock Mini-Cell Electrophoresis System (Thermo Scientific), for 1 hour at 180V constant electric current. Gels were blotted on Protran Pure Nitrocellulose membranes (Perkin Elmer) using a PerfectBlue Semi-Dry-Blotter Sedec (VWR Peqlab) for 1 hour at 14V and 2mA*cm^{-2} in 1X Towbin SDS Buffer (25mM Tris; 192mM Glycine; 0.1% SDS; 20% Methanol; pH 8.3). In the case of the experiment described in 5.4 (Fig. 12), the gels were blotted on Protran Pure Nitrocellulose membranes (Perkin Elmer) using a Mini Trans-Blot Cell (Biorad) wet-blotting system, for 1 hour at 130V constant electrical current in 1X Towbin SDS Buffer. For western blots, membranes were blocked in 5% milk (Frema) solution in TNT Buffer (20mM Tris-HCl pH 7.5; 150mM NaCl; 0.05% Tween-20), incubated overnight with primary antibodies and 1h with secondary antibodies. All antibody incubations were performed in 5% milk (Frema) solution in TNT Buffer in the cold (8ºC). The washing steps after both primary and secondary antibody incubations were performed in TNT Buffer at room temperature. For information regarding the antibodies used see Table 6. Protein detection was performed using Western Lightning Plus-ECL (Perkin Elmer) and Amersham Hyperfilm ECL (GE Healthcare) films developed using an RP X-Omat Processor, model M6B (Kodak). Developed films were scanned using a Perfection V750 Pro (Epson) scanner.

Primary Antibodies		
Name	**Dilution**	**Source**
anti-Oskar (rabbit)	1:2000	[89]
anti-Vasa (rat)	1:2000	[125]
anti-α-Tubulin (mouse)	1:10000	Sigma Aldrich
anti-Smaug (rabbit)	1:500	gift from Dr. E. Whale
Secondary Antibodies		
Name	**Dilution**	**Source**
anti-rabbit light chain specific	1:100000	Merck
anti-rabbit whole antibody	1:2000	GE Healthcare
anti-rat whole antibody	1:10000	GE Healthcare
anti-mouse whole antibody	1:10000	GE Healthcare

Table 6: List of antibodies used for western blot analysis

5.11. PCR

Test PCRs for iCLIP libraries generated as described in 5.5 (Fig. 13) and 5.9 (Fig. 19) were performed in 8µl volumes: 4µl 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific); 0.2µl P3/P5 primers mix (10µM each) (Sigma Aldrich) (see Table 7); 1µl cDNA template; 2.8µl DEPC treated H_2O. The reactions were incubated:

20'' at 98ºC - {3'' at 98ºC - 10'' at 65ºC - 15'' at 72ºC} - 1' at 72ºC.

For PCR cycles number (steps between {---}) see Figure 13 and 19.

Final PCRs for iCLIP library generation were performed in a 20µl volume: 10µl 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific); 1µl P3/P5 primers mix (10µM each) (Sigma Aldrich); 4µl cDNA template; 5µl DEPC treated H_2O. Reaction incubation conditions

were the same as described above, except for the number of PCR cycles: Figure 13, 36 cycles; Figure 19, 28 cycles. For the PCR of the dephosphorylation efficiency test experiment described in 6.7 (Fig. 16), reactions were performed in 10µl volumes: 5µl 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific); 0.5µl iCLIP PCR test (forward primer) (10µM) (Eurofins Genomics) (see Table 7); 0.5µl P3 (reverse primer) (10µM) (Sigma Aldrich); 1µl cDNA template; 3µl DEPC treated H_2O. The reactions were incubated:

1' at 98ºC - {10'' at 98ºC - 10'' at 60ºC - 5'' at 72ºC} - 1' at 72ºC.

30 PCR cycles (steps between {---}) were performed. All PCR reactions were performed in a C1000 Touch Thermal Cycler (Biorad).

Primer name	Sequence 5'-3'
P3	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
iCLIP PCR test 1	GGCCTCGTGAAATCCCGT

Table 7: List of primers used for PCR analysis

5.12. RNA and DNA gels

Test PCRs from iCLIP experiments described in 5.5 (Fig. 13B) and 5.9 (Fig. 19C) were supplemented with 6X Gel Loading Dye Blue (New England BioLabs) to a final 1X concentration. Samples were separated on 6% polyacrylamide Novex TBE gels (Invitrogen). Gels were run in 0.5X TBE buffer (45mM Tris-borate pH ~8.3; 1mM EDTA) using an XCell SureLock Mini-Cell Electrophoresis System (Thermo Scientific), for 40' at 60V constant electric current. The gels were stained in 0.5X TBE containing HD Green Plus DNA Stain (Intas) diluted 1:20000. Final PCRs for iCLIP libraries (Fig. 13C and Fig. 19D) were supplemented with 6X Gel Loading Dye Blue (New England BioLabs) to a final 1X concentration. Samples were

separated on 4% MetaPhor Agarose (Lonza) gel in 0.5X TBE buffer. The gels were run in 0.5X TBE buffer, for 3.5 hours at 50V constant electric current, in the cold (8ºC). Gels were stained with SYBR Gold Nucleic Acid Gel Stain (Invitrogen) diluted 1:10000. PCRs from the dephosphorylation efficiency test experiment described in 5.7 (Fig. 16) were supplemented with 6X Gel Loading Dye Blue (New England BioLabs) to a final 1X concentration. The reactions were separated on a 2% agarose (Sigma Aldrich) in 0.5X TBE buffer. Gels were run in 0.5X TBE buffer for 1 hour at 60V constant electric current. The gels were stained using HD Green Plus DNA Stain (Intas) diluted 1:20000. All DNA gels were visualized on a Quantum ST4 instrument (Vilber) using the Quantum-Capt software (Vilber). Reaction products of L3App adapter ligation efficiency test experiment described in 5.8 (Fig. 17 and Fig. 18) were separated on 10% polyacrylamide Novex TBE-Urea Gels (Invitrogen). Gels were pre-run in 0.5X TBE buffer using an XCell SureLock Mini-Cell Electrophoresis System (Thermo Scientific), for 1 hour at 180V constant electric current. After loading the gels were run 1 hour using the same conditions.

5.13. DNA extraction from gel

iCLIP libraries prepared as described in 5.5 (Fig. 13C) and 5.9 (Fig. 19D) and separated on 4% MetaPhor Agarose (Lonza) gels were extracted using QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions. The purified libraries were eluted in 20µl of Buffer EB (10mM Tris-HCl pH 8.5) (Qiagen).

5.14. RNA and DNA analysis

RNA and DNA samples were analyzed on a Bioanalyzer 2100 instrument (Agilent). RNA samples from experiments described in 5.3.1.2 (Fig. 11) and 5.3.1.3 (Fig. 15B) were analyzed on RNA 6000 Pico Chips (Agilent). RNA samples from experiments described in 5.6 (Fig. 15A)

and 5.9 (Fig. 19B) were analyzed on RNA 6000 Nano Chips (Agilent). DNA libraries from experiments described in 5.5 (Fig. 13D) and 5.9 (Fig. 20) were analyzed on High Sensitivity DNA Chips (Agilent). The results were visualized using 2100 Expert software (Agilent).

5.15. Radioactivity detection

Nitrocellulose membranes from experiments described in 5.4 (Fig. 12) were exposed overnight to a BAS-SR 2040 Imaging plate (Fujifilm). Gels from experiments described in 5.8 (Fig. 17 and Fig. 18) were exposed 1 hour and overnight to a BAS-MS 2040 Imaging plate (Fujifilm). The plates were analyzed on a Typhoon FLA 7000 laser scanner (Fujifilm), using the FLA 7000 software (Fujifilm).

5.16. NGS sequencing

iCLIP libraries prepared as described in 5.5 (Fig. 13) were sequenced on a MiSeq platform (Illumina). iCLIP libraries prepared as described in 5.9 (Fig. 19 and Fig. 20) were sequenced on a HiSeq 2000 platform (Illumina). In both cases, single-end sequencing was performed by the EMBL Genomics Core Facility.

5.16.1. Data analysis pipeline

Raw data from iCLIP library sequencing were analyzed using bioinformatic tools available on the EMBL Galaxy server (Genome Biology Unit Computational Support). Data were demultiplexed using the Je-Demultiplex tool [192]. Data were subsequently processed using FASTQ Groomer tool [200]. Reads were aligned against the *Drosophila* genome (release 3) [201] using the TopHat2 tool [193]. PCR duplicates were removed using rmdup application

from SAMTools toolkit [194]. Data were visualized using Integrative Genomics Viewer software [197] [198].

5.17. Software

All graphs and statistical analyses presented in this work were generated using Excel 2013 software (Microsoft). All images were processed using ImageJ (USA National Institutes of Health) and Photoshop CS6 (Adobe) software. Figure assembly was performed using Illustrator CS6 (Adobe) software.

6. References

1 **Spradling, AC. 1993.** Developmental genetics of oogenesis. In: Bate, M., Martinez-Arias, A. eds. The Development of *Drosophila* melanogaster. *Cold Spring Harbour Laboratory Press*, 1-70.

2 **Gates, J. 2012.** Drosophila egg chamber elongation: insights into how tissues and organs are shaped. *Fly (Austin)*. **6**, 213-27.

3 **Becalska, AN. & Gavis, ER. 2009.** Lighting up mRNA localization in Drosophila. *Development*. **136**, 2493-503.

4 **Riechmann, V. & Ephrussi, A. 2001.** Axis formation during Drosophila oogenesis. *Current Opinion in Genetics and Development*. **11**, 374-83.

5 **de Cuevas, M. & Spradling, AC. 1998.** Morphogenesis of the Drosophila fusome and its implications for oocyte specification. *Development.* **125**, 2781-9.

6 **Grieder, NC., de Cuevas, M. & Spradling, AC. 2000.** The fusome organizes the microtubule network during oocyte differentiation in Drosophila. *Development*. **127**, 4253-64.

7 **Mach, JM. & Lehmann, R. 1997.** An Egalitarian-BicaudalD complex is essential for oocyte specification and axis determination in Drosophila. *Genes and Development*. **11**. 423-35.

8 **Suter, B. Romberg, LM. & Steward, R. 1989.** Bicaudal-D, a Drosophila gene involved in developmental asymmetry: localized transcript accumulation in ovaries and sequence similarity to myosin heavy chain tail domains. *Genes and Development*. **3**. 1957-68.

9 **Laver, JD., Marsolais, AJ., Smibert, CA. & Lipshitz, HD. 2015.** Regulation and function of maternal gene products during the maternal-to-zygotic transition in Drosophila. *Current Topics in Developmental Biology*. **113**, 43-84.

10 **Lasko, P. 2012.** mRNA localization and translational control in Drosophila oogenesis. *Cold Spring Harbor Perspectives in Biology*. **4**(10). pii: a012294. doi: 10.1101/cshperspect.a012294.

11 **Vaccari, T. & Ephrussi, A. 2002.** The fusome and microtubules enrich Par-1 in the oocyte, where it effects polarization in conjunction with Par-3, BicD, Egl and dynein. *Current Biology*. **12**, 1524-8.

12 **Gonzales-Reyes, A., Elliott, H. & St Johnston, D. 1995.** Polarization of both major body axes in Drosophila by gurken-torpedo signaling. *Nature*. **375**, 654-8.

13 **Roth, S., Neuman-Silberberg, FS., Barcelo, G. & Schupbach, T. 1995.** cornichon and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in Drosophila. *Cell*. **81**, 967-78.

14 **Steinhauer, J. & Kalderon, D. 2006.** Microtubule polarity and axis fromation in the Drosophila oocyte. *Developmental Dynamics*. **235**, 1455-68.

15 **Zhao, T., Graham, OS., Raposo, A. & St Johnston, D. 2012.** Growing microtubules push the oocyte nucleus to polarize the Drosophila dorsal-ventral axis. *Science*. **336**, 999-1003.

16 **Neuman-Silberberg, FS. & Schupbach, T. 1993.** The Drosophila dorsoventral patterning gene gurken produces a dorsally localized RNA and encodes a TGF alpha-like protein. *Cell*. **75**, 165-74.

17 **Margaritis, LH., Kafatos, FC. & Petri, WH. 1980.** The eggshell of Drosophila melanogaster. I. Fine structure of the layers and regions of the wild-type eggshell. *Journal of Cell Science*. **43**, 1-35.

18 **Kishimoto, T. 2003.** Cell-cycle control during meiotic maturation. *Current Opinion in Cell biology*. **15**, 654-63.

19 **Von Stetina, JR., Tranguch, S., Dey, SK., Lee, LA., Cha, B. & Drummond-Barbosa, D. 2008.** alpha-Endosulfine is a conserved protein required for oocyte meiotic maturation in Drosophila. *Development*. **135**, 3697-706.

20 **Courtot, C., Frankhauser, C., Simanis, V. & Lehner, CF. 1992.** The Drosophila cdc25 homolog twine is required for meiosis. *Development*. **116**, 405-16.

21 **Mirouse, V., Formstecher, E. & Couderc, JL. 2006.** Interaction between Polo and BicD proteins links oocyte determination and meiosis control in Drosophila. *Development*. **133**, 4005-13.

22 **Xiang, Y., Takeo, S., Florens, L., Hughes, SE., Huo, LJ., Gilliland, WD., Swanson, SK., Teeter, K., Schwartz, JW., Washburn, MP., Jaspersen, SL. & Hawley, RS. 2007.** The inhibition of polo kinase by matrimony maintains G2 arrest in the meiotic cell cycle. *PLoS Biology*. **5**, e323.

23 **Horner, VL. & Wolfner, MF. 2008.** Mechanical stimulation by osmotic and hydrostatic pressure activates Drosophila oocytes in vitro in a calcium-dependent manner. *Developmental Biology*. **316**, 100-9.

24 **Mahowald, AP., Goralski, TJ. & Caulton, JH. 1983.** In vitro activation of Drosophila eggs. *Developmental Biology*. **98**, 437-45.

25 **Ferree, PL., Deneke, VE. & Di Talia, S. 2016.** Measuring time during early embryonic development. *Seminars in Cell & Developmental Biology*. **55**, 80-8.

26 **Tadros, W. & Lipshitz, HD. 2009.** The maternal-to-zygotic transition: a play in two acts. *Development*. **136**, 3033-42.

27 **Berleth, T., Burri, M., thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M. & Nusslein-Volhard, C. 1988.** The role of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *The EMBO Journal*. **7**, 1749-56.

28 **Macdonald, PM. & Struhl, G. 1988.** cis-acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. *Nature*. **336**, 595-8.

29 **Macdonald, PM., Kerr, K., smith, JL. & Leask, A. 1993.** RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development*, **118**, 1233-43.

30 **Macdonald, PM. & Kerr, K. 1997.** Redundant RNA recognition events in bicoid mRNA localization. *RNA*. **3**, 1413-20.

31 **Ferrandon, D., Koch, I., Westhof, E. & Nusslein-Volhard, C. 1997.** RNA-RNA interaction is required for the formation of specific bicoid mRNA 3‘ UTR-STAUFEN ribonucleoprotein particles. *The EMBO Journal*. **16**, 1751-8.

32 **Cha, BJ., Koppetsch, BS. & Theurkauf, WE. 2001.** In vivo analysis of Drosophila bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway. *Cell*. **106**, 35-46.

33 **Mische, S., Li, M., Serr, M. & Hays, TS. 2007.** Direct observation of regulated ribonucleoprotein transport across the nurse cell/oocyte boundary. *Molecular Biology of the Cell*. **18**, 2254-63.

34 **Clark, A., Meignin, C. & Davis, I. 2007.** A Dynein-dependent shortcut rapidly deliver axis determination transcripts into the Drosophila oocyte. *Development*. **134**, 1955-65.

35 **Riechmann, V. & Ephrussi, A. 2004.** Par-1 regulates bicoid mRNA localization by phosphorylating Exuperantia. *Development*. **131**, 5897-907.

36 **Weil, TT., Forrest, KM. & Gavis, ER. 2006.** Localization of bicoid mRNA in late oocytes is mediated by continual active transport. *Developmental Cell*. **11**, 251-62.

37 **Schnorrer, F., Luschnig, S., Koch, I. & Nusslein-Volhard, C. 2002.** Gamma-tubulin37C and gamma-tubulin ring complex protein75 are essential for bicoid RNA localization during Drosophila oogenesis. *Developmental Cell*. **3**, 685-96.

38 **Weil, TT., Parton, R., Davis, I. & Gavis, ER. 2008.** Changes in bicoid mRNA anchoring highlight conserved mechanisms during the oocyte-to-embryo transition. *Current Biology*. **18**, 1055-61.

39 **Weil, TT., Xanthakis, D., Parton, R., Dobbie, I., Rabouille, C., Gavis, ER. & Davis, I. 2010.** Distinguishing direct from indirect roles for bicoid mRNA localization factors. *Development*. **137**, 169-76.

40 **Spirov, A., Fahmy, K., Schneider, M., Frei, E., Noll, M. & Baumgartner, S. 2009.** Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient. *Development*. **136**, 605-14.

41 **Driever, W. & Nusslein-Volhard, C. 1988.** The Bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*. **54**, 95-104.

42 **Struhl, G., Struhl, K. & Macdonald, PM. 1989.** The gradient morphogen bicoid is a concentration dependent transcriptinal activator. *Cell*. **57**, 1259-73.

43 **Schroder, C., Tautz, D., Seifert, E. & Jackle, H. 1988.** Differential regulation of the two transcripts from the Drosophila gap segmentation gene hunchback. *The EMBO Journal*. **7**, 2881-7.

44 **Sonoda, J. & Wharton, RP. 2001.** Drosophila Brain tumor is a translational repressor. *Genes and Development*. **15**, 762-73.

45 **Gao, Q. & Finkelstein, R. 1998.** Targeting gene expression to the head: the Drosophila orthodenticle gene is a direct target of the Bicoid morphogen. *Development*. **125**, 4185-93.

46 **Ephrussi, A. & St Johnston, D. 2004.** Seeing is believing: the bicoid morphogen gradient matures. *Cell*. **116**, 143-52.

47 **Ephrussi, A. & Lehmann, R. 1992.** Induction of germ cell formation by oskar. *Nature*. **358**, 387-92.

48 **Wang, C. & Lehmann, R. 1991.** Nanos is the posterior determinant in Drosophila. *Cell*. **66**, 637-47.

49 **Wang, C., Dickinson, LK. & Lehmann, R. 1994.** Genetics of nanos localization in Drosophila. *Developmental Dynamics*. **199**, 103-15.

50 **Bergsten, SE. & Gavis, ER. 1999.** Role for mRNA localization in translational activation but not spatial restriction of nanos RNA. *Development*. **126**, 659-69.

51 **Forrest, KM. & Gavis, ER. 2003.** Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. *Current Biology*. **13**, 1159-68.

52 **Gavis, ER., Curtis, D. & Lehmann, R. 1996.** Identification of cis-acting sequences that control nanos RNA localization. *Developmental Biology*. **176**, 36-50.

53 **Jain, RA. & Gavis, ER. 2008.** The Drosophila hnRNP M homolog Rumpelstilstkin regulates nanos mRNA localization. *Development*. **135**, 973-82.

54 **Becalska, AN., Kin,YR., Belletier, NG., Lerit, DA., Sinsimer, KS. & Gavis, ER. 2011.** Aubergine is a component of a nanos mRNA localization complex. *Developmental Biology*. **349**, 46-52.

55 **Gavis, ER., Lunsford, L., Bergsten, SE. & Lehmann, R. 1996.** A conserved 90 nucleotide element madiates translational repression of nanos RNA. *Development*. **122**, 2791-800.

56 **Kalifa, Y., Huang, T., Rosen, LN., Chatterjee, S. & Gavis, ER. 2006.** Glorund, a Drosophila hnRNP F/H homolog, is an ovarian repressor of nanos translation. *Developmental Cell*. **10**, 291-301.

57 **Smibert, CA., Wilson, JE., Kerr, K. & Macdonald, PM. 1996.** Smaug protein represses translation of unlocalized nanos mRNA in the Drosophila embryo. *Genes and Development*. **10**, 2600-9.

58 **Nelson, MR., Leidel, AM. & Smibert, CA. 2004.** Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *The EMBO Journal*. **23**, 150-9.

59 **Zaessinger, S., Busseau, I. & Simonelig, M. 2006.** Oskar allows nanos mRNA translation in Drosophila embryos by preventing its deadenylation by Samug/CCR4. *Development*. **133**, 4573-83.

60 **Jeske, M., Moritz, B., Anders, A. & Wahle, E. 2011.** Smaug assembles an ATP-dependent stable complex repressing nanos mRNA translation at multiple levels. *The EMBO Journal*. **30**, 90-103.

61 **Barker, DD., Wang, C., Moore, J., Dickinson, LK. & Lehmann, R. 1992.** Pumilio is essential for function but not for distribution of the Drosophila abdominal determinant Nanos. Genes and *Development*. **6**, 2312-26.

62 **Hulskamp, M. & Tautz, D. 1991.** Gap genes and gradients--the logic behind the gaps. *Bioessays*. **13**, 261-8.

63 **Ephrussi, A., Dickinson, LK. & Lehmann, R. 1991.** Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell*. **66**, 37-50.

64 **Kim-Ha, J., Smith, JL. & Macdonald, PM. 1991.** oskar mRNA is localized to the posterior pole of the Drosophila oocyte. *Cell*. **66**, 23-35.

65 **Lehmann, R. 2016.** Germ Plasm Biogenesis--An Oskar-Centric Perspective. *Current Topics in Developmental Biology*. **116**, 679-707.

66 **Jambor, H., Mueller, S., Bullock, SL. & Ephrussi, A. 2014.** A stem-loop structure directs oskar mRNA to microtubule minus ends. *RNA*. **20**, 429-39.

67 **Dienstbier, M., Boehl, F., Li, X. & Bullock, SL. 2009.** Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes and Development*. **23**, 1546-58.

68 **Sanghavi, P., Liu, G., Veeranan-Karmegam, R., Navarro, C. & Gonsalves, GB. 2016.** Multiple roles for Egalitarian in Polarization of the Drosophila Egg Chamber. *Genetics*. **203**, 415-32.

69 **Navarro, C., Puthalakath, H., Adams, JM. Strasser, A. & Lehmann, R. 2004.** Egalitarian binds dynein light chain to establish oocyte polarity and maintain oocyte fate. *Nature Cell Biology*. **6**, 427-35.

70 **Hoogenraad, CC., Wulf, P., Schiefermeier, N., Stepanova, T., Galjart, N., Small, JV., Grosveld, F., de Zeeuw, CI. & Akhmanova, A. 2003.** Bicaudal D induces selective dynein-mediated microtubule minus end-directed transport. *The EMBO Journal*. **22**, 6004-15.

71 **Brendza, RP., Serbus, LR., Duffy, JB. & Saxton, WM. 2000.** A function for kinesin I in the posterior transport of oskar mRNA and Stufen protein. *Science*. **289**, 2120-2.

72 **Zimyanin, VL., Belaya, K., Pecreaux, J., Gilchrist, MJ., Clark, A., Davis, I. & St Johnston, D. 2008.** In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. *Cell*. **134**, 843-53.

73 **Hachet, O. & Ephrussi, A. 2004.** Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature*. **428**, 959-63

74 **Ghosh, S., Marchand, V., Gaspar, I. & Ephrussi, A. 2012.** Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. *Nature Structural and Molecular Biology*. **19**, 441-9.

75 **Simon, B., Masiewicz, P., Ephrussi, A. & Carlomagno, T. 2015.** The structure of the SOLE element of oskar mRNA. *RNA*. **21**, 1444-53.

76 **Moore, MJ., & Proudfoot, NJ. 2009.** Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*. **136**, 688-700.

77 **Palacios, IM., Gatfield, D., St Johnston, D. & Izaurralde, E. 2004.** An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature*. **427**, 753-7.

78 **Newmark, PA. & Boswell, RE. 1994.** The mago nashi locus encodes an essential product required for germ plasm assembly in Drosophila. *Development*. **120**, 1303-13.

79 **Hachet, O. & Ephrussi, A. 2001.** Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Current Biology*. **11**, 1666-74.

80 **van Eeden, FJ., Palacios, IM., Petronczki, M., Weston, MJ. & St Johnston, D. 2001.** Barentsz is essential for the posterior localization of oskar mRNA and colocalizes with it to the posterior pole. *The Journal of Cell Biology*. **154**, 511-23.

81 **Jambor, H., Brunel, C. & Ephrussi, A. 2011.** Dimerization of oskar 3'UTRs promotes hitchhiking for RNA localization in the Drosophila oocyte. *RNA*. **17**, 2049-57.

82 **Erdelyi, M., Michon, AM., Guichet, A., Glotzer, JB. & Ephrussi, A. 1995.** Requirement for Drosophila cytoplasmic tropomyosin in oskar mRNA localization. *Nature*. **377**, 524-7.

83 **Gaspar, I., Sysoev, V., Komissarov, A. & Ephrussi, A. 2017.** An RNA-binding atypical tropomyosin recruits kinesin-1 dynamically to oskar mRNPs. *The EMBO Journal*. **36**, 319-33.

84 **Huynh, JR., Munro, TP., Smith-Litiere, K., Lepesant, JA. & St Johnston, D. 2004.** The Drosophila hnRNPA/B homolog, Hrp48, is specifically required for a distinct step in osk mRNA localization. *Developmental Cell*. **6**, 625-35.

85 **Yano, T., Lopez de Quinto, S., Matsui, Y, Shevchenko, A., Shevchenko, A. & Ephrussi, A. 2004.** Hrp48, a Drosophila hnRNPA/B homolog, binds and regulates translation of oskar mRNA. *Developmental Cell*. **6**, 637-48.

86 **Micklem, DR., Adams, J., Grunert, S. & St Johnston, D. 2000.** Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. *The EMBO Journal*. **19**, 1366-77.

87 **Sinsimer, KS., Jain, RA., Chetterjee, S. & Gavis, ER. 2011.** A late phase of germ plasm accumulation during Drosophila oogenesis requires lost and rumpelstiltskin. *Development*. **138**, 3431-40.

88 **Markussen, FH., Michon, AM., Breitwieser, W. & Ephrussi, A. 1995.** Translational control of oskar generates short Oskar, the isoform that induces pole plasm assembly. *Development*. **121**, 3723-32.

89 **Vanzo, NF. & Ephrussi, A. 2002.** Oskar anchoring restricts pole plasm formation to the posterior of the Drosophila oocyte. *Development*. **129**, 3705-14.

90 **Tanaka, T. & Nakamura, A. 2008.** The endocytic pathway acts downstream of Oskar in Drosophila germ plasm assembly. *Development*. **135**, 1107-17.

91 **Babu, K., Cai, Y., Bahri, S., Yang, X. & Chia, W. 2004.** Roles of Bifocal, Homer, and F-actin in anchoring Oskar to the posterior cortex of Drosophila oocytes. *Genes and Development*. **18**, 138-43.

92 **Vanzo, N., Oprins, A., Xanthakis, D., Ephrussi, A. & Rabouille, C. 2007.** Stimulation of endocytosis and actin dynamics by Oskar polarizes the Drosophila oocyte. *Developmental Cell*. **12**, 543-55.

93 **Smith, JL., Wilson, JE. & Macdonald, PM. 1992.** Overexpression of oskar directs ectopic activation of nanos and presumptive pole cell formation in Drosophila embryos. *Cell*. **70**, 849-59.

94 **Cook, HA., Koppetsch, BS., Wu, J. & Theurkauf, WE. 2004.** The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell*. **116**, 817-29.

95 **Chekulaeva, M. & Ephrussi, A. 2004.** Drosophila development: RNA interference ab ovo. *Current Biology*. **14**, R428-30.

96 **Kim-Ha, J., Kerr, K. & Macdonald, PM. 1995.** Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. *Cell*. **81**, 403-12.

97 **Snee, M., Benz, D., Jen, J. & Macdonald, PM. 2008.** Two distinct domains of Bruno bind specifically to the oskar mRNA. *RNA Biology*. **5**, 1-9.

98 **Wilhelm, JE., Hilton, M., Amos, Q. & Henzel, WJ. 2003.** Cup is an eIF4E binding protein required for both the translational repression of oskar and the recruitment of Barentsz. *The Journal of Cell Biology*. **163**, 1197-204.

99 **Igreja, C. & Izaurralde, E. 2011.** CUP promotes deadenylation and inhibits decapping of mRNA targets. *Genes and Development*. **25**, 1955-67.

100 **Castagnetti, S. & Ephrussi, A. 2003.** Orb and a long poly(A) tail are required for efficient oskar translation at the posterior pole of the Drosophila oocyte. *Development*. **130**, 835-43.

101 **Chekulaeva, M., Hentze, MW. & Ephrussi, A. 2006.** Bruno acts as a dual repressor of oskar translation, promoting mRNA oligomerization and formation of silencing particles. *Cell*. **124**, 521-33.

102 **Besse, F., Lopez de Quinto, S., Marchand, V., Trucco, A. & Ephrussi, A. 2009.** Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. *Genes and Development*. **23**, 195-207.

103 **Reveal, B., Yan, N., Snee, MJ., Pai, CI., Gim, Y. & Macdonald, PM. 2010.** BREs mediate both repression and activation of oskar mRNA translation and act in trans. *Developmental Cell*.**18**, 496-502.

104 **Munro, TP., Kwon, S., Schnapp, BJ. & St Johnston, D. 2006.** A repeated IMP-binding motif controls oskar mRNA translation and anchoring independently of Drosophila melanogaster IMP. *The Journal of Cell Biology*. **172**, 577-88.

105 **Juge, F., Zaessinger, S., Temme, C., Wahle, E. & Simonelig, M. 2002.** Control of poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in Drosophila. *The EMBO Journal*. **21**, 6603-13.

106 **Benoit, P., Papin, C., Kwak, JE., Wickens, M. & Simonelig, M. 2008.** PAP- and GLD-2-type poly(A) polymerase are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. *Development*. **135**, 1969-79.

107 **Jenny, A., Hachet, O., Zavorszky, P., Cyrklaff, A., Weston, MD., Johnston, DS., Erdelyi, M. & Ephrussi, A. 2006.** A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development*. **133**, 2827-33.

108 **Kanke, M., Jambor, H., Reich, J., Marches, B., Gstir, R., Ryu, YH., Ephrussi, A. & Macdonald, PM. 2015.** oskar RNA plays multiple noncoding roles to support oogenesis and maintain integrity of the germline/soma distinction. *RNA*. **21**, 1096-109.

109 **Jeske, M., Bordi, M., Glatt, S., Muller, S., Rybin, V., Muller, CW. & Ephrussi, A. 2015.**The Crystal Structure of the Drosophila Germline Inducer Oskar Identifies Two Domains with Distinct Vasa Helicase- and RNA-Binding Activities. *Cell Reports*. **12**, 587-98.

110 **Yang, N., Yu, Z., Hu, M., Wang, M., Lehmann, R. & Xu, RM. 2015.** Structure of Drosophila Oskar reveals a novel RNA binding protein. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, 11541-6.

111 **Anantharaman, V., Zhang, D. & Aravind, L. 2010.** OST-HTH: a novel predicted RNA-binding domain. *Biology Direct*. **5**, 13.

112 **Callebaut, I. & Mornon, JP. 2010.** LOTUS, a new domain associated with small RNA pathways in the germline. *Bioinformatics*. **26**, 1140-4.

113 **Jeske, M., Muller, CW. & Ephrussi, A. 2017.** The LOTUS domain is a conserved DEAD-box RNA helicase regulator essential for the recruitment of Vasa to the germ plasm and nuage. *Genes and Development*. **31**, 939-952.

114 **Rongo, C., Gavis, ER. & Lehmann, R. 1995.** Localization of oskar RNA regulates oskar translation and requires Oskar protein. *Development*. **121**, 2737-46.

115 **Breitwieser, W., Markussen, FH., Horstmann, H. & Ephrussi, A. 1996.** Oskar protein interaction with Vasa represents an essential step in polar granule assembly. *Genes and Development*. **10**, 2179-88.

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

117 **Suyama, R., Jenny, A., Curado, S., Pellis-van Berkel, W. & Ephrussi, A. 2009.** The acting-binding protein Lasp promotes Oskar accumulation at the posterior pole of the Drosophila embryo. *Development*. **136**, 95-105.

118 **Hurd, TR., Herrmann, B., Sauerwald, J., Sanny, J., Grosch, M. & Lehmann, R. 2016.** Long Oskar Controls Mitochondrial Inheritance in Drosophila melanogaster. *Developmental Cell*. **39**, 560-71.

119 **Extavour, CG. & Akam, M. 2003.** Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development*. **130**, 5869-84.

120 **Hathaway, DS. & Selman, GG. 1961.** Certain aspects of cell lineage and morphogenesis studied in embryos of Drosophila melanogaster with an ultra-violet micro-beam. *Journal of Embryology and Experimental Morphology*. **9**, 310-25.

121 **Warn, R. 1975.** Restoration of the capacity to form pole cells in u.v.-irradiated Drosophila embryos. *Journal of Embryology and Experimental Morphology*. **33**, 1003-11.

122 **Illmensee, K. & Mahowald, AP. 1974.** Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg. *Proceedings of the National Academy of Sciences of the United States of America*. **71**, 1016-20.

123 **Lehmann, R. & Nusslein-Volhard, C. 1986.** Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. *Cell*. **47**, 141-52.

124 **Schpbach, T. & Wieschaus, E. 1986.** Germline autonomy of maternal-effect mutations altering the embryonic body pattern of Drosophila. *Developmental Biology*. **113**, 443-8.

125 **Tomancak, P., Guichet, A., Zavorszky, P. & Ephrussi, A. 1998.** Oocyte polarity depends on regulation of gurken by Vasa. *Development*. **125**, 1723-32.

126 **Johnstone, O. & Lasko, P. 2004.** Interaction with eIF5B is essential for Vasa function during development. *Development*. **131**, 4167-78.

127 **Boswell, RE. & Mahowald, AP. 1985.** tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell*. **43**, 97-104.

128 **Cote, J. & Richard, S. 2005.** Tudor domains bind symmetrical dimethylated arginines. *The Journal of Biological Chemistry*. **280**, 28476-83.

129 **Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y. & Bedford, MT. 2006.** Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Reports*. **7**, 397-403.

130 **Arkov, AL., Wang, JY., Ramos, A. & Lehmann, R. 2006.** The role of Tudor domains in germline development and polar granule architecture. *Development*. **133**, 4053-62.

131 **Anne, J. & Mechler, BM. 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**, 2167-77.

132 **Anne, J., Ollo, R., Ephrussi, A. & Mechler, BM. 2007.** Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in Drosophila. *Development*. **134**, 137-46.

133 **Kobayashi, S., Yamada, M., Asaoka, M. & Kitamura, T. 1996.** Essential role of the posterior morphogen nanos for germline development in Drosophila. *Nature*. **380**, 708-11.

134 **Hayashi, Y., Hayashi, M. & Kobayashi, S. 2004.** Nanos suppresses somatic cell fate in Drosophila germ line. *Proceedings of the National Academy of Sciences of the United States of America*. **10**, 10338-42.

135 **Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K. & Kobayashi, S. 1999.** Maternal Pumilio acts together with Nanos in germline development in Drosophila embryos. *Nature Cell Biology*. **1**, 431-7.

136 **Kadyrova, LY., Habara, Y., Lee, TH. & Wharton, RP. 2007.** Translational control of maternal Cyclin B mRNA by Nanos in the Drosophila germline. *Development*. **134**, 1519-27.

137 **Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S. & Lasko, PF. 1996.** Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. *Science*. **274**, 2075-9.

138 **Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P. & Nakamura, A. 2008.** Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature*. **451**, 730-3.

139 **Timinszky, G., Bortfeld, M. & Ladurner, AG. 2008.** Repression of RNA polymerase II transcription by a Drosophila oligopeptide. *PLos One*. **3**, e2506.

140 **Nakamura, A. & Seydoux, G. 2008.** Less is more specific: specification of the germline by transcriptional repression. *Development*. **135**, 3817-27.

141 **Jongens, TA., Hay, B., Jan, LY. & Jan, YN. 1992.** The germ cell-less gene product: a posteriorly localized component necessary for germ cell development in Drosophila. *Cell*. **70**, 569-84.

142 **Robertson, SE., Dockendorff, TC., Leatherman, JL., Faulkner, DL. & Jongens, TA. 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. *Developmental Biology*. **215**, 288-97.

143 **Moore, J., Han, H. & Lasko, P. 2009.** Bruno negatively regulates germ cell-less expression in a BRE-independent manner. *Mechanisms of Development*. **126**, 503-16.

144 **Lerit, DA., Shebelut, CW., Lawlor, KJ., Rusan, NM., Gavis, ER., Schedl, P. & Deshpande, G. 2017.** Germ Cell-less Promotes Centrosome Segregation to Induce Germ Cell Formation. *Cell Reports*. **18**, 831-39.

145 **Leatherman, JL., Levin, L., Boero, J. & Jongens, TA. 2002.** germ cell-less acts to repress transcription during the establishment of the Drosophila germ cell lineage. *Current Biology*. **12**, 1681-5.

146 **Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, TR., Tomancak, P. & Krause, HM. 2007.** Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*. **131**, 174-87.

147 **Jambor, H., Surendranath, V., Kalinka, AT., Mejstrik, P., Saalfeld, S. & Tomancak, P. 2015.** Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. *eLife*. **4**, doi: 10.7554/eLife.05003.

148 **Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, LM., Krijgsveld, J. & Hentze, MW. 2013.** System-wide identification of RNA-binding proteins by interactome capture. *Nature Protocols*. **8**, 491-500.

149 **Wessels, HH., Imami, K., Baltz, AG., Kolinski, M., Beldovskaya, A., Selbach, M., Small, S., Ohler, U. & Landthaler, M. 2016.** The mRNA-bound proteome of the early fly embryo. *Genome research*. **26**, 1000-9.

150 **Scherrer, T., Mittal, N., Janga, SC. & Gerber, AP. 2010.** A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One*. **5**,e15499.

151 **Tsvetanova, NG., Klass, DM., Salzman, J. & Brown, PO. 2010.** Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. *PLoS One*. **5**,e12671.

152 **Baltz, AG., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C. & Landthaler, M. 2012.** The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular Cell*. **46**, 674-90.

153 **Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, BM., Strein, C., Davey, NE., Humphreys, DT., Preiss, T., Steinmetz, LM., Krijgsveld, J. & Hentze, MW. 2012.** Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. **149**, 1393-406.

154 **Sysoev, VO., Fischer, B., Frese, CK., Gupta, I., Krijgsveld, J., Hentze, MW., Castello, A. & Ephrussi, A. 2016.** Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nature Communications*. **7**:12128. doi: 10.1038/ncomms12128.

155 **Metzker, ML. 2010.** Sequencing technologies - the next generation. *Nature Reviews. Genetics*. **11**, 31-46.

156 **Milek, M., Wyler, E. & Landthaler, M. 2012.** Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. *Seminars in Cell and Developmental Biology*. **23**, 206-12.

157 **Keene, JD., Komisarov, JM. & Friedersdorf, MB. 2006.** RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature Protocols*. **1**, 302-7.

158 **Mili, S. & Steitz, JA. 2004.** Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA*. **10**, 1692-4.

159 **Wagenmakers, AJ., Reinders, RJ. & van Venrooij, WJ. 1980.** Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. *European Journal of Biochemistry*. **112**, 323-30.

160 **Zhang, L., Zhang, K., Pradl, R. & Schoffl, F. 2004.** Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. *Biochemical and Biophysical Research Communications*. **322**, 705-11.

161 **Ule, J., Jensen, KB., Ruggiu, M., Mele, A., Ule, A. & Darnell, RB. 2003.** CLIP identifies Nova-regulated RNA networks in the brain. *Science*. **302**, 1212-5.

162 **Licatalosi, DD., Mele, A., Fak, JJ., Ule, J., Kayikci, M., Chi, SW., Clark, TA., Schweitzer, AC., Blume, JE., Wang, X., Darnell, JC. & Darnell, RB. 2008.** HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. **456**, 464-9.

163 **Sanford, JR., Wang, X., Mort, M., Vanduyn, N., Cooper, DN., Mooney, SD., Edenberg, HJ. & Liu, Y. 2009.** Splicing factor SRSF1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Research*. **19**, 381-94.

164 **Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, YS., Zhang, C., Yeo, G., Black, DL., Sun, H., Fu, XD. & Zhang, Y. 2009.** Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Molecular Cell*. **36**, 996-1006.

165 **Yeo, GW., Coufal, NG., Liang, TY., Peng, GE., Fu, XD. & Gage, FH. 2009.** An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature Structural and Molecular Biology*. **16**, 130-7.

166 **Saito, Y., Miranda-Rottmann, S., Ruggiu, M., Park, CY., Fak, JJ., Zhong, R., Duncan, JS., Fabella, BA., Junge, HJ., Chen, Z., Araya, R., Fritzsch, B., Hudspeth, AJ. & Darnell, RB. 2016.** NOVA2-mediated RNA regulation is required for axonal pathfinding during development. *eLife*. **5**, pii: e14371. doi: 10.7554/eLife.14371.

167 **Chi, SW., Zang, JB., Mele, A. & Darnell, RB. 2009.** Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. **460**, 479-86.

168 **Zhang, C. & Darnell, RB. 2011.** Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnology*. **29**, 607-14.

169 **Darnell, RB. 2010.** HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdisciplinary Reviews. RNA*. **1**, 266-86.

170 **Hafner, M., Landthaler, M., Burger L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr., Jungkamp, AC., Munschauer, M., Ulrich, A., Wardle, GS., Dewell, S., Zavolan, M. & Tuschl, T. 2010.** Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. **141**, 129-41.

171 **Hamilton, MP., Rajapakshe, KI., Bader, DA., Cerne, JZ., Smith, EA., Coarfa, C., Hartig, SM. & McGuire, SE. 2016.** The Landscape of microRNA Targeting in Prostate Cancer Defined by AGO-PAR-CLIP. *Neoplasia*. **18**, 356-70.

172 **Chang, X., Li, B. & Rao, A. 2015.** RNA-binding protein hnRNPLL regulates mRNA splicing and stability during B-cell to plasma-cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, E1888-97.

173 **Yoon, JH., De, S., Srikantan, S., Abdelmohsen, K., Grammatikakis, I., Kim, J., Kim, KM., Noh, JH., White, EJ., Martindale, JL., Yang, X., Kang, MJ., Wood, WH. 3rd., Noren Hooten, N., Evans, MK., Becker, KG., Tripathi, V., Prasanth, KV., Wilson, GM., Tuschl, T., Ingolia, NT., Hafner, M. & Gorospe, M. 2014.** PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nature Communications*. **5**:5248. doi: 10.1038/ncomms6248.

174 **Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, DJ., Luscombe, NM. & Ule, J. 2010.** iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural and Molecular Biology*. **17**, 909-15.

175 **Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, NM., Rot, G., Zupan, B., Curk, T. & Ule, J. 2010.** iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biology*. **8**, e1000530.

176 **Tollervey, JR., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., Konig, J., Hortobagyi, T., Nishimura, AL., Zupunski, V., Patani, R., Chandran, S., Rot, G., Zupan, B., Shaw, CE. & Ule, J. 2011.** Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nature Neuroscience*. **14**, 452-8.

177 **Rossbach, O., Hung, LH., Khrameeva, E., Schreiner, S., Konig, J., Curk, T., Zupan, B., Ule, J., Gelfand, MS. & Bindereif, A. 2014.** Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biology*. **11**, 146-55.

178 **Ma, X., Zhu, X., Han, Y., Story, B., Do, T., Song, X., Wang, S., Zhang, Y., Blanchette, M., Gogol, M., Hall, K., Peak, A., Anoja, P. & Xie, T. 2017.** Aubergine Controls Germile Stem Cell Self-Renewal and Progeny Differentiation via Distinct Mechanisms. *Developmental Cell*.**41**, 157-69.

179 **Bruun, GH., Doktor, TK., Borch-Jensen, J., Masuda, A., Krainer, AR., Ohno, K. & Andresen, BS. 2016.** Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biology*. **14**:54. doi: 10.1186/s12915-016-0279-9.

180 **River, C., Idris, J., Scott, H., Rogers, M., Lee, YB., Gaunt, J., Phylactou, L., Curk, T., Campbell, C., Ule, J. Norman, M. & Uney, JB. 2015.** iCLIP identifies novel roles for SAFB1 in regulating RNA processing and neuronal function. *BMC Biology*. **13**:111. doi: 10.1186/s12915-015-0220-7.

181 **Neugebauer, JM. 1990.** Detergents: an overview. *Methods in Enzymology*. **182**, 239-53.

182 **Raff, JW., Whitfield, WG. & Glover, DM. 1990.** Two distinct mechanisms localise cyclin B transcripts in syncytial Drosophila embryos. *Development*. **110**, 1249-61.

183 **Ding, D., Parkhurst, SM., Halsell, SR. & Lipshitz, HD. 1993.** Dynamic Hsp83 RNA localization during Drosophila oogenesis and embryogenesis. *Molecular and Cellular Biology*. **13**, 3773-81.

184 **Lantz, V., Ambrosio, L. & Schedl, P. 1992.** The Drosophila orb gene is predicted to encode sex-specific germline RNA-binding proteins and has localized transcripts in ovaries and early embryos. *Development*. **115**, 75-88.

185 **Chang, JS., Tan, L. & Schedl, P. 1999.** The Drosophila CPEB homolog, orb, is required for oskar protein expression in oocytes. *Developmental Biology*. **215**, 91-106.

186 **Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, DJ., Luscombe, NM. & Ule, J. 2011.** iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *Journal of Visualized Experiments*. pii: 2638. doi: 10.3791/2638.

187 **Ho, CK., Wang, LK., Lima, Cd. & Shuman, S. 2004.** Structure and mechanism of RNA ligase. *Structure*. **12**, 327-39.

188 **Urlaub, H., Hartmuth, K. & Luhrmann, R. 2002.** A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*. **26**, 170-81.

189 **Greenberg, JR. 1969.** Synthesis and properties of ribosomal RNA in Drosophila. *Journal of Molecular Biology*. **46**, 85-98.

190 **Winnebeck, EC., Millar, CD. & Warman, GR. 2010.** Why does insect RNA look degraded? *Journal of Insect Science*. **10**, doi: 10.1673/031.010.14119.

191 **Huppertz, I., Attig, J., D'Ambrosio, A., Easton, LE., Sibley, CR., Sugimoto, Y., Tajnik, M., Konig, J. & Ule, J. 2014.** iCLIP: protein-RNA interactions at nucleotide resolution. *Methods*. **65**, 274-87.

192 **Girardot, C., Scholtalbers, J., Sauer, S., Su, SY. & Furlong, EE. 2016.** Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics*. **17**, 419.

193 **Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, SL. 2013.** TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. **14**, R36. doi: 10.1186/gb-2013-14-4-r36.

194 **Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R.; 1000 Genome Project Data Processing Subgroup. 2009.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078-9.

195 **Van Nostrand, EL., Pratt, GA., Shishkin, AA., Gelboin-Burkhart, C., Fang, MY., Sundararaman, B., Blue, SM., Nguyen, TB., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M. & Yeo, GW. 2016.** Robust transcriptoiome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*. **13**, 508-14.

196 **Shishkin, AA., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, RP., Rudy, RF., Patel, MM., Novod, N., Hung, DT., Gnirke, A., Garber, M., Guttman, M. & Livny, J. 2015.** Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature Methods*. **12**, 323-5.

197 **Robinson, JT., Thorvaldsdottir, H., Winckler, W., Guttman, M., Landers, ES., Getz, G. & Mesirov, JP. 2011.** Integrative genomics viewer. *Nature Biotechnology*. **29**, 24-6.

198 **Thorvaldsdottir, H., Robinson, JT. & Mesirov, JP. 2013.** Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*. **14**, 178-92.

199 **Bustin, S. & Nolan, T. 2017.** Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *European Journal of Clinical Investigation*. doi: 10.1111/eci.12801.

200 **Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J. & Nekrutenko, A.; Galaxy Team. 2010.** Manipulation of FASTQ data with Galaxy. *Bioinformatics*, **26**, 1783-5.

201 **Celniker, SE., Wheeler, DA., Kronmiller, B., Carlson, JW., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, SP., Frise, E., Hodgson, A., George, RA., Hoskins, RA., Laverty, T., Muzny, DM., Nelson, CR., Pacleb, JM., Park, S., Pfeiffer, BD., Richards, S., Sodergren, EJ., Svirskas, R., Tabor, PE., Wan, K., Stapleton, M., Sutton, CG., Venter, C., Weinstock, G., Scherer, SE., Wyers, EW., Gibbs, RA. & Rubin, GM. 2002.** Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. *Genome Biology*. **3**, RESEARCH0079.