universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

„The Biology of piRNA clusters in

Drosophila melanogaster“

Verfasser

Franz Sebastian Gruber

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt:	A 490
Studienrichtung lt. Studienblatt:	Diplomstudium Molekulare Biologie
Betreuerin / Betreuer:	Dr. Julius Brennecke

Table of Contents

Acknowledgments

to my beloved, generous and ever supporting parents, grandparents and sister

to my encouraging, unique and best friend and mentor, Rolf

to my inspiring, great and noble boss and mentor, Julius

to my ever helpful, wise and courteous lab mates, Derya, Katharina, Greg, Kirsten, Fabio, Dominik, Daniel O., Daniel J. and Christopher

to my companion and friend, Florian

to my love, Veronika

Abstract English

Transposable elements (TEs) are selfish genetic elements in the DNA of all uni- and multicellular organisms. Their major goal is to maintain themselves by invading the genome of gonads and by mobilizing within the genome of their host. This enables them to propagate vertically from generation to generation.

In *Drosophila melanogaster* a phenomenon called hybrid dysgenesis causes sterile F1 progeny and the reason for this are TEs, which are active due to a lack of control. As these selfish elements can interfere with a host's fitness (either positive or negative) a tight and accurate control mechanism has to be active.

Drosophila melanogaster harbors three distinct classes of small RNAs: siRNAs, miRNAs and piRNAs. The latter class evolved sperately and in a Dicer-independent fashion. Recently, they have been linked to TE silencing, which is now known as the piRNA pathway.

A 23-29 nt long piRNA is loaded into an Argonaut protein (Piwi, Aub, AGO3) and navigates this protein-RNA complex to complementary TE sequences, which results in TE silencing. In *Drosophila* the major source for piRNAs are piRNA clusters. These are long stretches of DNA (ranging from 40 to 260 kb) containing degenerated and immobile transposon sequences. piRNA clusters usually reside at euchromatin-heterochromatin boundaries, which might display a special form of chromatin modification. It is believed that the cluster is transcribed as a long single stranded RNA. All subsequent steps, which are involved in generating smaller piRNAs or loading into an effector complex, are hardly understood. To reveal novel information about the biology of piRNA clusters, especially when and where during development clusters are expressed and where within the cell the transcript localizes, are the main aims of my thesis.

In this study, I was able to create a germline specific sensor, which allows the analysis of all stages during gonadal development. My results show, how powerful a sensor based approach is in answering certain questions (e. g. the spatiotemporal expression of a piRNA cluster). Besides giving insights in the expression pattern of clusters, the sensors display an optimal tool for further investigations (e. g. as a

readout in a reverse genetic screen). To this end, I was also able to map putative promoter sequences, using a reporter based system and therefore supporting a single-promoter model. By applying different sophisticated approaches I could show, that the site of cluster transcription localizes close to a piRNA processing body. The transcript itself seems to localize nearby or within this body, according to my results. This confirms our expectations and these assays are useful for further investigations on piRNA biogenesis.

Abstract German

Transposons (“springende Gene”) sind eigennützige Abschnitte im Erbgut (DNA) aller Ein- und Mehrzeller. Um Teil der DNA zu bleiben, versuchen Transposons in die Erbinformation von Keimzellen zu gelangen, um von dort an weitere Generationen vererbt zu werden.

Die Gemeine Fruchtfliege *Drosophila melanogaster* weist ein interessantes Verhalten auf, wenn man männliche Tiere aus freier Wildbahn mit weiblichen Tieren aus einem Fliegenlabor miteinander kreuzt. Dieses Phämonen nennt man Hybriddysgenese und sie führt zu Infertilität der F1 Tochtergeneration. Die Ursache wurde auf aktive Transposons (die sogenannten *P*- oder *I*-Elemente) und auf einen nicht vorhandenen Kontrollmechanismus zurückgeführt. Da Transposons auf Grund ihres Verbreitungsmechanismus ausschlaggebenden Einfluss auf die evolutionsbiologische Fitness haben, müssen sie a priori unter ständiger Kontrolle stehen.

Drosophila melanogaster besitzt neben kurzen interferierenden RNAs („siRNA“) und micro RNAs („miRNA“) noch eine dritte Klasse von kleinen RNAs: Piwi interagierende RNAs („piRNAs“). Diese Klasse von kurzen RNAs entstand separat und wurde in Verbindung mit Transposonkontrolle gebracht, was man nun als „piRNA Pathway“ versteht.

piRNAs sind 23-29 Nukleotide lang, binden an so genannte Argonaut Proteine (Piwi, AGO3, Aub) und dienen in diesem Protein-RNA Komplex als Navigationssystem, um komplementäre Sequenzen zu erkennen und weiters die Aktivität von Transposons zu unterbinden. Eine Quelle für piRNAs im Genom der Fruchtfliege sind so genannte „piRNA clusters“, bei welchen es sich um eine Aneinanderschachtelung von degenerierten und inaktiven Transposonfragmenten handelt. piRNA Cluster können bis zu 260 Kilobasenpaare lang sein und sind meist an der Grenze zwischen Heterochromatin und Euchromatin zu finden. Das Wissen über diese Regionen, die ausschlaggebend für Transposonregulation sind ist sehr rar. Man vermutet, dass piRNA Cluster als Ganzes transkribiert werden und nach dem Export aus dem Zellkern zu kurzen piRNAs verarbeitet werden.

Wo und wann ein solcher Cluster aktiv ist und wo in der Zelle das lange Vorläufertranskript lokalisiert, sind die Hauptfragen dieser Arbeit.
Während meiner Studien ist es mir gelungen einen Gewebsspezifischen biologischen Sensor zu etablieren, mit welchem man die Expression eines Clusters messen kann. Desweiteren gelang es mir Sequenzen proximal eines Clusters liegend mit einer möglichen Promotoraktivität in Verbindung zu bringen. Weiters, konnte ich Protokolle von hochsensitiven Methoden, welche zur Lokalisierung von RNA Molekülen dienen, adaptieren. Mit Hilfe dieser Methoden ist es nun ermöglicht, tiefer in die Materie des piRNA pathways einzutauchen und neue Erkenntnisse über z. B. die Biogenese der kurzen piRNAs zu erhalten.

The importance of this study

Selfish genetic elements like transposable elements (TEs) are present in almost every genome. Due to the fact, that TEs are able to mobilize, they can pose an enormous threat to the genome. Therefore this has to be tightly controlled. A recently discovered small RNA pathway, called the piRNA pathway has been linked to TE repression and therefore acts as a guardian of the genome. Interestingly this pathway is conserved from sponges up to mammals but it is best understood in *Drosophila melanogaster*. One hallmark of the piRNA pathway are piRNA clusters, which give rise to piRNAs, which are essential for TE silencing. In simple words, piRNA clusters serve as an important memory for TE repression. In order to understand more about this genetic memory, I followed up strategies to elucidate novel insights about piRNA clusters. My results will help to get a better understanding of piRNA cluster biology and thus the pathway itself.

Introduction

The battle of the genomes

Sequencing genomes has revealed that DNA does not only contain genes. It was surprising to see, that genes exist of exons and introns and that there are large spacers between genes, sometimes referred as ‘junk DNA’. Indeed even more surprising was to see, that this ‘junk DNA’ is not just unused sequence, but is able to perform a lot of different actions. Parts of this ‘junk DNA’ have been assigned as transposable elements (TEs), which have great impact on genomes. Every genome is thought to have a genomic conflict. This conflict displays how different parts of a genome try to outnumber other parts, by taking the risk of harming the host. As TEs are able to mobilize within a genome, they are part of this conflict. TEs have evolved in a very sophisticated manner and besides moving vertically from species to species, they are also capable of moving within the host to multiply. This mobilization presents a source of danger, as TEs can easily jump into the coding region of a gene and thus disrupt it. That is why the hosts of TEs evolved an elaborate mechanism to silence TEs. In animals, a specialized small RNA silencing pathway, called the piRNA pathway is the major system that keeps TEs silent (Senti and Brennecke, 2010; Hurst and Werren, 2001; Kazazian, 2011).

Drosophila melanogaster has been the workhorse for geneticists for over a century. Not surprisingly it was and still is of great help in revealing novel insights into the piRNA pathway. During the development of the oocyte in *Drosophila melanogaster* TEs are transcribed and try to invade the genome of the oocyte. To suppress all the deleterious effects TEs can exert on a genome upon mobilization, the piRNA pathway is active in the ovary. Hence, *Drosophila melanogaster*, with all its established tools, is a powerful model organism for studying this recently discovered pathway. A brief summary about oogenesis will be given below to further understand, why the piRNA pathway is active in the ovary of *Drosophila melanogaster*.

Oogenesis in *Drosophila melanogaster*

One fundamental property of life is reproduction, i.e. maintenance of the own species (Reece et al., 2010). In the common fruitfly *Drosophila melanogaster,* one part of reproduction takes place in the ovary of the female fly. Each ovary is made up of several aligned ovarioles (about 16 per ovary), all held together by a peritoneal sheat, consisting of muscle tissue (King, 1970). An ovariole is the functional unit of an ovary and has three particular structures: the germarium, in which the germline stem cells (GSCs) reside within a niche, the vitellarium, which is comprised of egg chambers that encapsulated from the germarium and the egg itself, which gets fertilized by sperm and laid in the end (see Figure 1). During early ovariole development a GSC divides asymmetrically, giving rise to a new GSC and to a cystoblast. This cystoblast further divides mitotically four times, building up a 16-cell cyst. The cells within this cyst are interconnected by structures composed of an actin meshwork, so called ring canals. The consequence is a common cytoplasm within the 16-cell cyst (King, 1970). After leaving the germarium the 16-cell cyst subsequently develops into an egg chamber: one out of the 16 cells has entered meiosis and has become the future oocyte, whereas the other 15 cells are supporting nurse cells. Nevertheless the 16 cells are still connected by ring canals (King, 1970). In addition, these 16 cells comprising the germline tissue are surrounded by epithelial cells (i. e. follicle cells), which are of somatic origin (King, 1970; Bate and Arias, 1993; Wolpert and Tickle, 2010).

During later stages of development a process called nurse cell breakdown leads to a massive increase in cell mass of the oocyte. The supporting nurse cells start to deliver nutrients and a certain amount of biological material (e.g. RNA or proteins) into the oocyte (King, 1970). It is important to mention, that the oocyte is transcriptionally inert and therefore only transcripts from nurse cells are transported into the oocyte. A dangerous situation now arose for the oocyte: the promoters of certain TEs are activated in the nurse cells and in addition the flow of nutrients displays a nice route to enter the oocyte and invade its genome. (Levin and Moran, 2011; Bate and Arias, 1993). In addition, there is a second threat: retroelements such as the *gypsy* element are capable of forming virus like particles, which are able to

take a vesicular route (i. e. the vitelline traffic) to enter the underlying oocyte from the follicle cells (Pélisson et al., 1994).

This twofold threat, has to be controlled, as selfish genetic elements like TEs are able to change their place and also to multiply within genomes (Levin and Moran, 2011). How TEs can do this and which impact this can have to the host's fitness, will be discussed below. This will further help to understand the important role of the piRNA pathway to silence TEs.

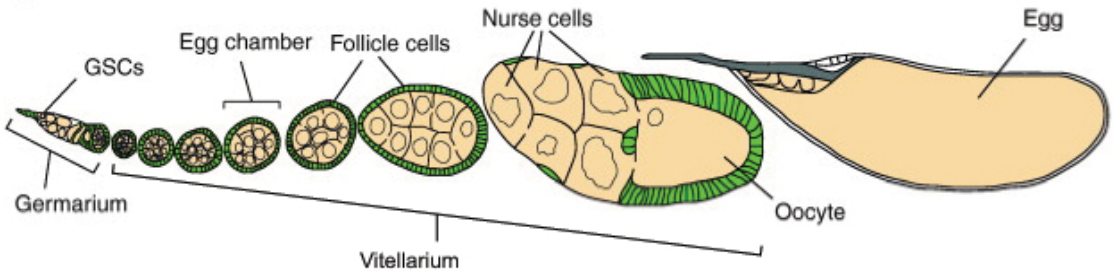


Figure 1 Oogenesis of *Drosophila melanogaster*. One ovariole is shown in the cartoon, starting with the germarium, which contains the germline stem cells (GSCs), followed by egg chambers, which consist of germline cells in beige and somatic cells in green, respectively. The prospective oocyte is transcriptionally silent, is nourished by nurse cells and arrested within meiosis I. After fertilization by the sperm the oocyte further transforms into the egg (adapted from (Senti and Brennecke, 2010)).

Transposable elements

Hallmarks of the genomes of all organisms are TEs. TEs are one subclass of selfish genetic elements. These parasitic elements try to invade and mobilize within the genome of a host cell and to further gain advantage over transmission compared to the rest of the genome. Once it invaded the genome of a gonad, a TE has now two ways of propagations: vertical transmission from generation to generation and multiplying within the host genome, by using one of the mobilizing mechanisms introduced later (Kazazian, 2011; Hurst and Werren, 2001).

TEs have been first described by Barbara McClintock decades ago (McCLINTOCK, 1950) and have been of great interest since then. Especially new high-throughput sequencing technologies (e. g. 454 or Solexa platforms) have helped to elucidate more about the biology of TEs in diverse organisms recently. In the following, I will give a brief overview of the classification and biology of TEs. For more detailed

information, please refer to (Hurst and Werren, 2001; Goodier and Kazazian, 2008; Kazazian, 2011; Levin and Moran, 2011).

Class I and Class II transposons

TEs achieve their mobilization by two general mechanisms. The first strategy is called ‘cut-and-paste’ mechanism and utilizes an enzyme called transposase (encoded by the element, see Figure 2). This strategy is found in class II transposons (also known as DNA transposons). By recognizing inverted terminal repeats (ITRs, see Figure 3), which flank the element, the transposase excises the whole element and inserts it again in proximity to the original site. This can lead to a phenomenon called ‘local hopping’. Using this mobilization strategy the element can cause harm the genome due to e. g. imprecise excision or insertional mutagenesis (Kazazian, 2004; Levin and Moran, 2011; Castro and Carareto, 2004).
A prominent candidate of class II transposons in *Drosophila melanogaster* is the *P* element. This 2.9 kb element has been linked to a phenomenon called hybrid dysgenesis (discussed later) and moreover has been used as toolbox for generating a plethora of useful transgenic flies (Castro and Carareto, 2004; Zhang and Spradling, 1993; Bellen et al., 2011)

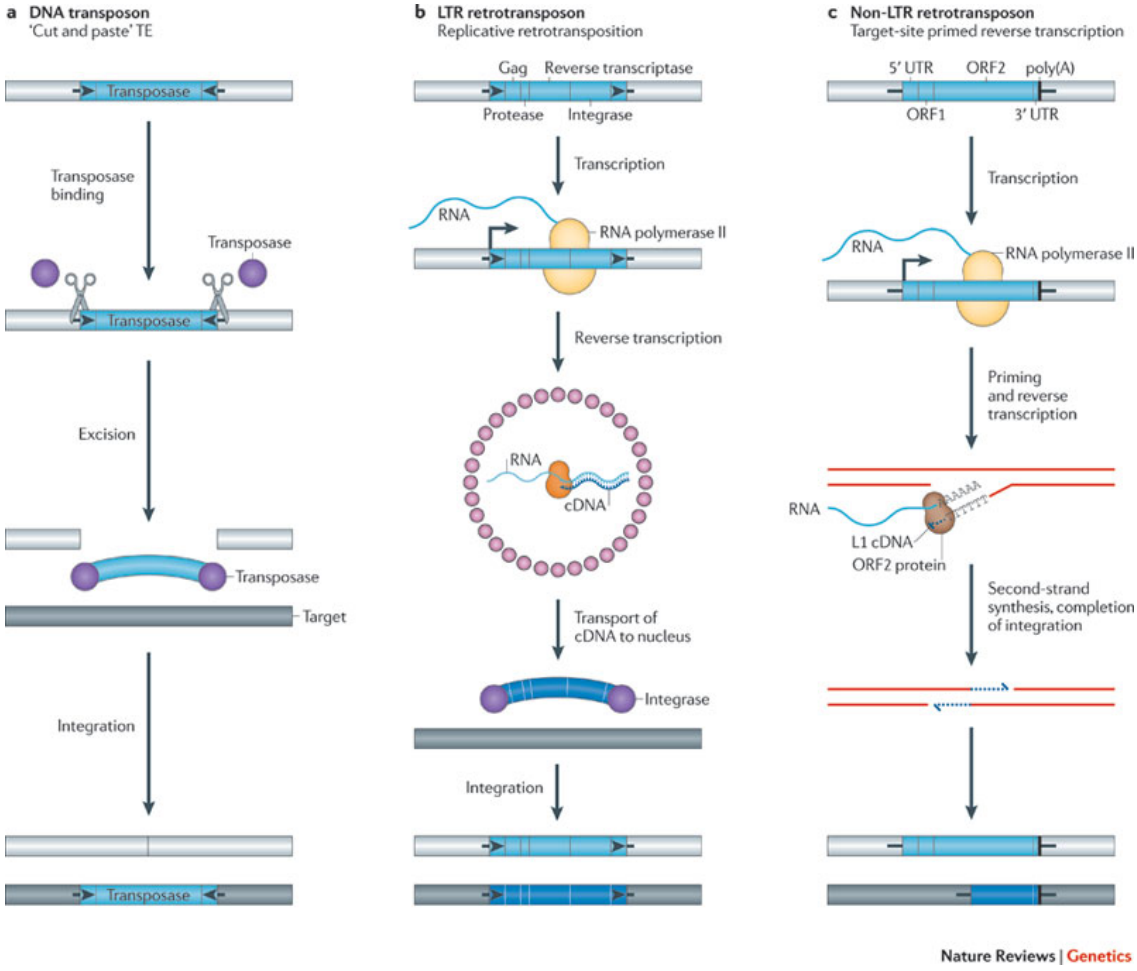


Figure 2 Comparison of class I and class II transposon mobilization mechanisms. (a) DNA transposons are cut out and integrated to a new site by the transposase enzyme. The cleavage at the target site generates staggered strands and subsequently to target-site duplication (TSD) of 4-8 nt (black flanks). Retroelement mobilization starts with the transcription of the element (b) and (c). LTR retrotransposons (b) are reverse-transcribed within a viral particle (encoded by the element) and the copied DNA is integrated into the genome after transport to the nucleus. Non-LTR retroelements (c) follow a slightly different way after transcription: a process called target-site-primed reverse transcription (TPRT) leads to the generation of a DNA intermediate, which is integrated into the genome. TPRT utilizes a nick, which is induced by an endonuclease (element encoded), for priming RT reaction. Like DNA transposons (a), LTR- (b) and Non-LTR retroelements (c) create TSD at the site of integration (picture from Levin & Moran, 2011).

The second strategy is used by class I transposons (e. g. LTR retrotransposon and Non-LTR retrotransposons, respectively). They mobilize by a 'copy-and-paste' mechanism presented in Figure 2 (Goodier and Kazazian, 2008). By using this strategy, the element is first transcribed into an RNA intermediate, which is further

reverse-transcribed into cDNA and subsequently integrated into the genome. All the required enzymes as well as the important recognition sites are encoded by the elements as depicted in Figure 3 (Kazazian, 2004). Depending on their ends, these elements either contain long terminal repeats (LTRs) or instead have a 5' and 3' UTR, respectively and a polyadenylate sequence at their 3' ends (Non-LTRs) (Kazazian, 2004). This strategy offers the TE an efficient way to generate copies within the genome. These class I elements pose danger to the host's genome because of uncontrolled mobilization, which can lead e. g. to DNA double strand breaks (Kazazian, 2011).

Particular TEs and their peculiarities in *Drosophila melanogaster*

Approximately 15 % of the *Drosophila melanogaster* genome is annotated as TEs (Hurst and Werren, 2001). The fact that TEs can either be harmful (insertion may lead to gene disruption; chromosomal rearrangements due to sequence similarity; etc.) or useful (of evolutionary benefit; protecting chromosome ends) to their hosts, has opened a fascinating area of research (Hurst and Werren, 2001; Shpiz et al., 2007). The following section highlights the dualism of TEs in terms of their harm or benefit for an organism like *Drosophila melanogaster*.

The endogenous retroelements *gypsy* and *ZAM* are the best examples for tissue specific activity and for their evolved and sophisticated spreading within the *Drosophila* ovary (Leblanc et al., 2000; Prudhomme et al., 2005; Pélisson et al., 1994).

Both elements mobilize only in certain genetic backgrounds (Prudhomme et al., 2005). The soma specific *gypsy* element e. g. has been shown to be regulated by the *flamenco* locus, which is located on the X-chromosome. In the absence of *flamenco* the element is activated and has now the ability to invade the germline genome (Mével-Ninio et al., 2007). The same holds true for the *ZAM* element (Prudhomme et al., 2005). As both elements are restricted to the somatic follicular cells surrounding the oocyte, the elements somehow have to find a way to enter the oocyte. This might occur by using the vitellogenic vesicular traffic (Prudhomme et al., 2005; Pélisson et al., 1994).

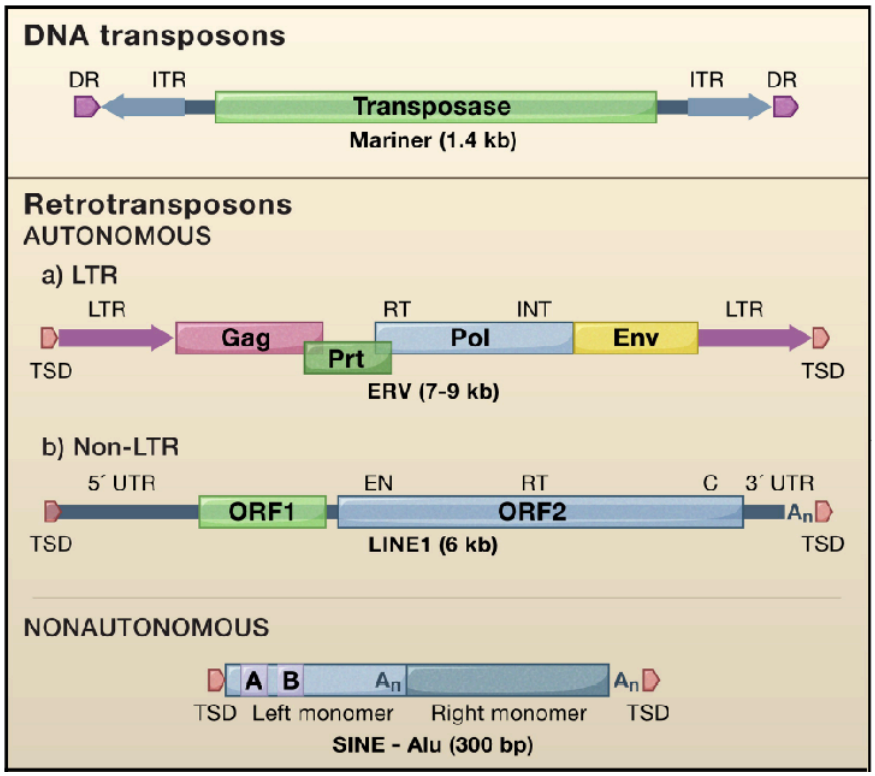


Figure 2 Overview of the structure of TEs. DNA transposons (class II) encode the enzyme needed for transposition (i. e. transposase), which recognizes the flanking inverted terminal repeats (ITRs). Direct repeats (DRs) arise upon the integration to new site. Retrotransposons (class I) can be further classified into autonomous and nonautonomous elements. LTR elements posses Gag (group specific antigen, capsid protein), Pol (polymerase containing reverse transcriptase and integrase domain, respectively), Env (envelope protein) and Prt (protease) in their reading frame, which is flanked by long-terminal repeats (LTRs). Non-LTR elements encode an EN (endonuclease), RT (reverse transcriptase) and C (Zinc knuckle domain). These elements do have a 5' UTR as well as 3' UTR followed by a poly(A) signal. Like the LTRs, Non-LTRs also exhibit TSD, which flank the element after integration to the new site. Nonautonomous elements are dependent on the transposition machinery from an autonomous element. Besides the TSDs there is no homology to other elements (Goodier and Kazazian, 2008).

Besides the destructive effects of uncontrolled transposition, Non-LTR retroelements like *HeT-A*, *TAHRE* and *TART* have also a vital role in establishing and maintaining telomeres, as *Drosophila melanogaster* lacks the telomerase enzyme ().

Having in mind the harm and benefits of TEs in *Drosophila melanogaster*, I will now briefly discuss an observation called 'hybrid dysgenesis', which has been connected to TEs and recently to the piRNA pathway as well.

Hybrid dysgenesis

An interesting phenomenon was observed decades ago, when *Drosophila* strains from the wild were crossed with laboratory-strains. If the cross was set up with a wild-caught male and a laboratory-strain female, the F1 female progeny was either sterile, or laid no eggs at all (Bregliano et al., 1980). If the cross was set up reciprocally, the female F1 generation did not show this effect. This was very surprising as the F1 generations of both crosses are genetically identical. This observation was coined 'hybrid dysgenesis' (Kidwell et al., 1973; 1977; Picard, 1976). Two types of hybrid dysgenesis have been described. The P-M system (P cytotype present in wild-caught flies and M cytotype present in laboratory-strain flies) as well as the I-R system (Inducer strain from the wild, Reactive laboratory strain) involve the *P*-element or the *I*-element, respectively. Neither element is present in the laboratory-strain used for over hundreds of years of *Drosophila* research (Bregliano et al., 1980; Griffiths et al., 2010; Kazazian, 2011; Rubin and Spradling, 1982; Castro and Carareto, 2004). These studies clearly demonstrate, that a single TE is able to pose an enormous threat to a whole species. To this end it also became clear, that TE repression involves a maternal component (Bregliano et al., 1980; Kidwell et al., 1977; Brennecke et al., 2008).
As we have seen so far, TEs can exert a tremendous danger, because a lack of regulation can lead to sterility of the host. In the following pages, I will describe how TEs are kept silent by the piRNA pathway.

The piRNA pathway

To briefly sum up the previous paragraphs: during oogenesis in *Drosophila melanogaster*, TEs try to invade the genome of the oocyte. In addition to moving within the oocyte's genome, this ensures also propagation to the next generations. This mobilization of TEs can pose a threat to the genome, as it can have deleterious effects. 'Hybrid dysgenesis' is the best example of harmful effects of TEs: sterility in F1 generations. Therefore a sophisticated strategy is needed to keep TEs silent. Research over the last decade has shown that this silencing system is a specialized

small RNA pathway, called the piRNA pathway. How this pathway is acting in *Drosophila* is the topic of the following sections.

piRNAs not siRNAs not miRNAs

Remarkable progress has led to a broad understanding of small regulatory RNAs since their discovery in 1993 (Lee et al., 1993). Three types of small RNAs can be distinguished in most animals: siRNAs, miRNAs and piRNAs (Ghildiyal and Zamore, 2009; Senti and Brennecke, 2010; Siomi et al., 2011).

Short interfering RNAs (siRNAs) are 21 nucleotides (nt) long and were first discovered in plants (Hamilton and Baulcombe, 1999). siRNAs can be subdivided into endo-siRNAs and exo-siRNAs. They play a pivotal role in post-transcriptional gene regulation, antiviral defense, transposon repression and transcriptional gene silencing (Ghildiyal and Zamore, 2009).

Another class of small RNAs are microRNAs (miRNAs). Their size can range from 20-25 nt and they control stability of their mRNA targets as well as their translation (Ghildiyal and Zamore, 2009). Both, siRNAs and miRNAs are processed from a double stranded precursor by Dicer (Bernstein et al., 2001). The mature siRNA or miRNA is the loaded into an Argonaute protein to form an RNA induced silencing complex (RISC). The small RNA (i. e. siRNA or miRNA) acts as a guide to direct RISC to complementary target RNAs within the cell (Ghildiyal and Zamore, 2009; Siomi and Siomi, 2009).

The third class of small regulatory RNAs are piRNAs. They are distinct from the other two classes in how and where they act. First of all piRNAs are Dicer independent (Vagin et al., 2006). Their size profile ranges from 23-29 nt and as indicated by the name piRNAs, they are loaded into an Argonaute protein of the PIWI-clade. In *Drosophila melanogaster* the three PIWI proteins are Argonaute3 (AGO3), Aubergine (Aub) and Piwi, which are all three associated with piRNAs and are highly expressed in the gonads of flies (Senti and Brennecke, 2010; Siomi et al., 2011; Khurana and Theurkauf, 2010; Ghildiyal and Zamore, 2009). The first hint at the biological function of piRNAs came from studies in the *Drosophila* male germline. A certain locus called *stellate* has been shown to be controlled by piRNAs (at this time called ‘repeat

associated RNAs – rasiRNAs') encoded by the *suppressor of stellate [su(ste)]* locus (Aravin et al., 2001; 2003). Follow up studies confirmed that piRNAs bound to PIWI proteins map to transposon and other repetitive sequences. Hence the main function of the piRNA pathway is believed to be TE silencing (Brennecke et al., 2007; Gunawardane et al., 2007). As this study is conducted in *Drosophila melanogaster* and most of our current understanding of the piRNA pathway stems from *Drosophila*, I will now introduce some concepts of this hardly understood pathway.

The somatic piRNA pathway

The *Drosophila* ovary consists of two major tissues, a somatic tissue, which mainly consists of follicle cells and the germline tissue, which includes nurse cells and the oocyte (King, 1970) (see Figure 1). The piRNA pathway is essential in both tissues but it has been shown that the complexity of the pathway differs in both tissues with two independent pathways being active in the soma and germline (Malone et al., 2009). In the soma the so called primary piRNA pathway keeps TEs silent, whereas in the germline those primary piRNAs are amplified in a target dependent manner (Brennecke et al., 2007; Malone et al., 2009; Senti and Brennecke, 2010). I will focus this paragraph on the somatic primary piRNA pathway.

The simplest way one could think about silencing of TEs mediated by the piRNA pathway is a linear model. A piRNA source encodes for a long single stranded transcript. This transcript is recognized by specific factors and localized to a site of processing, where it gets parsed into smaller pieces (i. e. 23-29 nt piRNAs). Finally, mature piRNAs need to be loaded into an Argonaute protein, to form the active effector complex (Senti and Brennecke, 2010; Khurana and Theurkauf, 2010; Siomi et al., 2011). In the primary piRNA pathway of the somatic follicle cells (Brennecke et al., 2007), only one Argonaute protein (Piwi) of the PIWI-clade is present. It has been shown, that Piwi, once bound to a piRNA shuttles to the nucleus. Recent studies have identified several factors that are essential for Piwi-piRNA biogenesis (Olivieri et al., 2010; Saito et al., 2010). The RNA helicase Armitage has been identified using tissue specific RNAi and a sensor system in ovaries, which detects impairments of the somatic piRNA pathway (Olivieri et al., 2010) as well as in an accompanying

study using RNAi in an ovarian somatic cell line system (OSCs) (Saito et al., 2010). Armitage resides within the cytoplasm and co-localizes with the tudor-domain containing RNA helicase Yb (fs(1)Yb) in so called Yb-bodies, which are discrete cytoplasmic foci typically close to the nuclear membrane (Olivieri et al., 2010). A third essential factor is the Phospholipase D protein Zucchini. This putative nuclease localizes to mitochondria (Huang et al., 2011; Watanabe et al., 2011) and is necessary for Piwi-piRNA maturation (Brennecke et al., 2007; Pane et al., 2007; Olivieri et al., 2010; Saito et al., 2010). Recent additions to the pathway are Vreteno (Handler et al., 2011; Zamparini et al., 2011) and the two Tdrd12 family proteins CG11133 and CG 31755 (Handler et al., 2011). All three proteins are essential factors in the piRNA pathway and contain a tudor-domain, which seems to be an important protein domain in the piRNA pathway. Tudor-domains are able to bind symmetrically methylated arginine residues, which have been found on PIWI proteins (Vagin et al., 2009; Handler et al., 2011; Nishida et al., 2009; Kirino et al., 2009). Despite this progress, our understanding of piRNA biogenesis at the mechanistic level is still very poor.

Playing ping-pong with TEs

The linear model of the primary piRNA pathway is currently widely accepted. Nevertheless the piRNA pathway contains a second module, which acts only in the germline: the ping-pong amplification loop (Brennecke et al., 2007; Gunawardane et al., 2007).

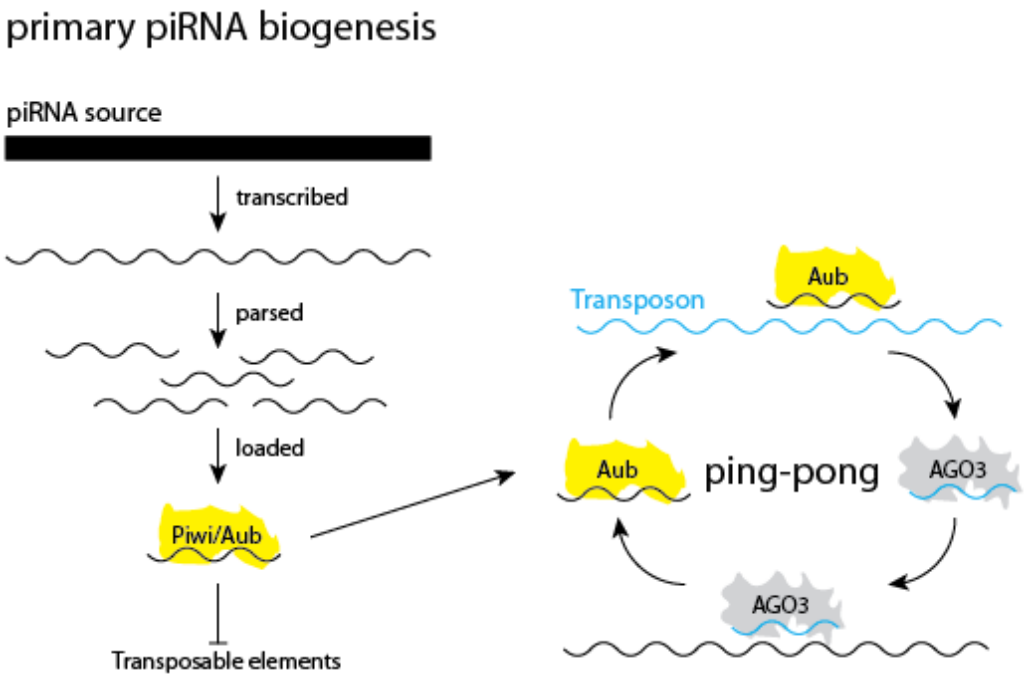


Figure 4. The primary piRNA biogenesis and the ping-pong amplificiation loop. The cartoon depicts both pathway modules in very simplified version. The primary piRNA biogenesis takes place in the somatic tissue exclusively, whereas in the germline tissue the primary piRNAs bound to Piwi or Aub, are integrated into a target amplification loop.

According to the ping-pong model, primary PIWI-piRNA complexes as well as maternally deposited piRNAs feed into a sophisticated amplification loop, whose essential core consists of AGO3 and Aub (Malone et al., 2009; Brennecke et al., 2008). Both proteins localize to a perinuclear electron dense cloud called nuage (French word for cloud) (Eddy, 1974). Information from deep-sequencing experiments suggest that Aub is loaded with antisense piRNAs, which guide Aub to the sense transcript of an active transposon (Brennecke et al., 2007; Senti and Brennecke, 2010). Slicing of this transcript then generates a sense piRNA, which is loaded into AGO3. Subsequently, AGO3 cleaves the transcript of the piRNA source triggering synthesis of the initiating piRNA. This amplification, which resembles a ping-pong game, is a hallmark of the piRNA pathway in the germline of *Drosophila melanogaster* as well as other organisms (Senti and Brennecke, 2010) (e. g. sponges, planarians or mammals) and is thought to be a mechanism for fine tuning the signal towards active elements (Brennecke et al., 2007; Gunawardane et al., 2007; Senti and Brennecke, 2010).

An overview of the primary pathway and the ping-pong-cycle is given in Figure 4.

piRNA clusters

The starting point of the piRNA pathway are piRNA clusters, which are one of the main sources for piRNAs besides transcripts of active TEs and certain mRNAs (Senti and Brennecke, 2010; Khurana and Theurkauf, 2010; Saito et al., 2009; Robine et al., 2009). These clusters are long stretches of DNA sequence, ranging up to 260 kb (Brennecke et al., 2007). They harbor a diverse accumulation of degenerate, immobile and therefore inactive transposon fragments (Brennecke et al., 2007). piRNA clusters have been elucidated by mapping those piRNAs to the genome, which map only a single time (Brennecke et al., 2007). One could describe a piRNA cluster as a graveyard for TEs, which serves as a systemic memory for genome immunity (Senti and Brennecke, 2010). *Drosophila melanogaster* has about 15 major piRNA clusters (Brennecke et al., 2007). Currently, there is only very little information about these specific genomic regions. The best understood clusters are *flamenco* (also known as the *COM* locus (Desset et al., 2003)), *cluster 20A* and *cluster 42AB* (the numbers arise from their cytological location) (Brennecke et al., 2007). All three are located at the euchromatin-heterochromatin boundary, which could implicate a special epigenetic mark and hence regulation (Senti and Brennecke, 2010; Khurana and Theurkauf, 2010). Indeed it has been shown, that mutations in the *rhino* locus are coherent with transposon upregulation (Klattenhoff et al., 2009). Rhino encodes an HP1 family protein and is found at the borders of *cluster 42AB* (Khurana and Theurkauf, 2010).

Consistent with the two modules (i. e. primary biogenesis and ping-pong) that are active in different tissues, piRNA clusters also show tissue specificity. While *flamenco* is thought to be active only in the somatic follicle cells and *cluster 42AB* is thought to be germline exclusive, *cluster 20A* seems to be active in both tissues. Evidence for this came from comparing piRNA populations from ovaries to those from early embryos (representing germline piRNAs) (Malone et al., 2009; Brennecke et al., 2007; Senti and Brennecke, 2010). It is believed that piRNAs from *flamenco* and *cluster 20A* arise from a long single stranded cluster transcript. Evidence for this came from a *P*-element insertion in a region upstream of the first piRNAs mapping uniquely to *flamenco*. This leads to a massive reduction of *flamenco* derived piRNAs

along the entire length of the 180 kb *flamenco* cluster (Sarot et al., 2004; Brennecke et al., 2007; Prud'homme et al., 1995; Robert et al., 2001; Desset et al., 2003; Mével-Ninio et al., 2007).

In contrast to the unidirectional clusters (*flamenco* and *cluster 20A*) most of the piRNA clusters (e. g. *cluster 42AB*) must be transcribed bidirectionally as uniquely mapping piRNAs originate from both genomic strands (Brennecke et al., 2007; Malone et al., 2009; Senti and Brennecke, 2010; Khurana and Theurkauf, 2010). Figure 5 gives an overview of the piRNA profiles of *flamenco* and *cluster 42AB*.

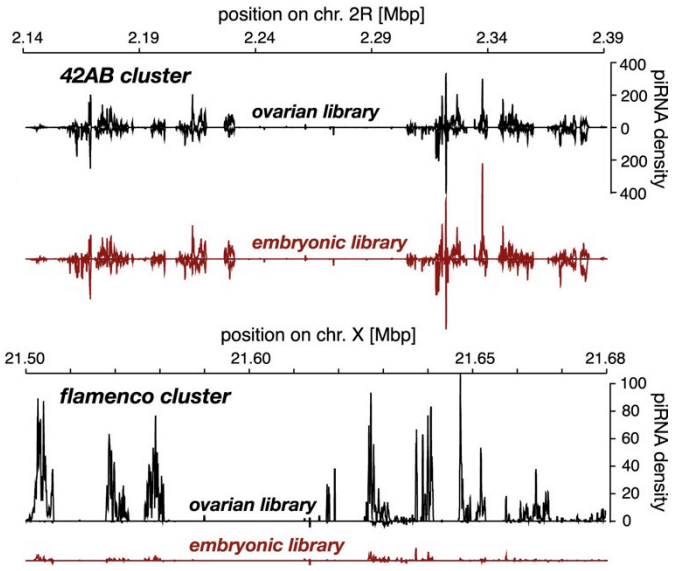


Figure 5 piRNA profiles of *flamenco* and *cluster 42AB*. Uniquely mapping piRNAs to piRNA clusters are shown according to their genomic region. The comparsion of ovarian and embryonic libraries gives insights into tissue specific expression. (picture taken from (Malone et al., 2009))

Interestingly, germline piRNA clusters have also been mapped to telomeric regions (Brennecke et al., 2007). As already indicated in a previous section *Drosophila* telomeres consists of repeating arrays of *HeT-A*, *TAHRE* and *TART* elements (Pardue and Debaryshe, 2000). This brings telomere biology in touch with the piRNA pathway and demonstrates once more the diverse biology surrounding this pathway as well as its importance (Senti and Brennecke, 2010; Khurana and Theurkauf, 2010).

Aims

One of the most fascinating aspects of the piRNA pathway are piRNA clusters. Up to now we have only very sparse information on their biology. It has been shown that they typically reside at the euchromatin-heterochromatin boundary and they encode for a genetic memory, which is used for TE silencing (Brennecke et al., 2007). piRNA clusters are either uni- or bidirectionally transcribed and some evidence is given, that the somatic and germline tissue discriminates between distinct clusters. To understand the logic of the piRNA pathway, it will be essential to reveal novel insights about the biology and regulation of piRNA clusters.

The main problems I would like to address: When and where during development are specific clusters active/expressed? Does a cluster have a single promoter and can we map it? What is the subcellular localization of the transcript from certain clusters? Does the transcript localize to the proposed biogenesis factories like nuage or Yb-bodies?

To address these questions I followed several complementary approaches described in the result section. Based on my results, the various established tools will be powerful in addressing new questions in the biology of piRNA clusters.

Results

Generation of suitable tissue specific sensors

In *Drosophila melanogaster* two major tissues build up an ovariole: germline and somatic tissue (depicted in Figure 6a). This holds also true for the *Drosophila* testis, although the morphology is completely different (Figure 6b) (King, 1970; Bate and Arias, 1993).

One ultimate goal of this thesis is to generate a new set of biological sensors, which are able to report, when and where during development of an ovariole a distinct piRNA cluster is active. Previous work demonstrated how powerful sensors are in their application to address piRNA pathway related questions (Handler et al., 2011; Olivieri et al., 2010; Sarot et al., 2004). Nevertheless these sensors have some limitations: early developmental timepoints (e. g. germarium) are not detected well or at all and to this end single-cell resolution is not provided.

In this study I want to create new sensors with the following features: first, a tissue specific promoter for early and late gonadal development, second eGFP for high resolution single cell analysis, third LacZ for visual comparisons and fourth a nuclear localizing signal (NLS) to localize the signal to the nucleus. In simple words, a tissue specific promoter driving the expression of a NLS-eGFP-LacZ reporter. In addition the sensor should contain two restriction sites (rare 8 base cutters), which allow the insertion of a piRNA target sequence (see Figure 6c). An 18 aa stretch of the *serendipity δ* gene was already convincing in a previous study as an NLS (Noselli and Vincent, 1991), therefore I decided to use it in this approach. To this end a GFP-LacZ fusion has also been shown to work in *Drosophila* and in addition this fusion seems to enhance the GFP signal remarkably (Shiga et al., 1996). Therefore I planned to use it for my sensors as well. In addition two different reporters make the sensor more flexible in terms of output.

The function of the sensor is outlined in Figure 6d. In the normal situation a sensor containing a piRNA target, is downregulated due to the piRNA pathway and therefore stated as ‘OFF’. Upon impairment of the piRNA pathway, the sensor is translated

and, either β-galactosidase or eGFP can be detected within the nuclei due to the NLS (see Figure 6d).

In order to see, in which tissue a certain cluster is active at distinct developmental timepoints, I had to find a suitable promoter, that is active in a desired tissue. The aim was to generate a germline specific sensor along with a sensor, which is expressed in both tissues, germline and soma.

Previous studies showed, that a construct containing a fragement of the putative *vasa* promoter would be sufficient for driving eGFP in the germline tissue only (Sano et al., 2002). Cloning a 1 kb promoter fragment upstream of the NLS-eGFP-LacZ reporter resembled a convincing strategy for me.

The second sensor was based on the *piwi* promoter, as Piwi is expressed in both, the germline and somatic tissue, mainly localizing within the nucleus (Cox et al., 2000). Combining a 1.6 kb putative *piwi* promoter with the NLS-eGFP-LacZ reporter was the approach of choice for the germline/soma sensor.

Both promoter constructs were injected into embryos. Using an attB containing integration vector and embryos expressing site specific integrase ϕC31, the construct was integrated into two different landing sites: *attP2* on chromosome 3 and *attP40* on chromosome 2 (for more details on landing sites please refer to (Markstein et al., 2008)). After balancing the transgene homozygous using either *TM3* or *CyO* as balancer chromomes, ovaries were analysed for eGFP and β-galactosidase expression.

The *vasa* sensor showed faint, but detectable signal in the somatic cells (data not shown), although expected to be germline specific. The *piwi* sensor was active in both tissues, but showed somatic patches, in which the reporter was not detectable (data not shown). Both promoters were therefore useless for my study. All the further constructs were injected into flies containing the *attP2* landing site. This site seems to work more accurate in terms of sensor expression compared to the *attP40* site (data not shown).

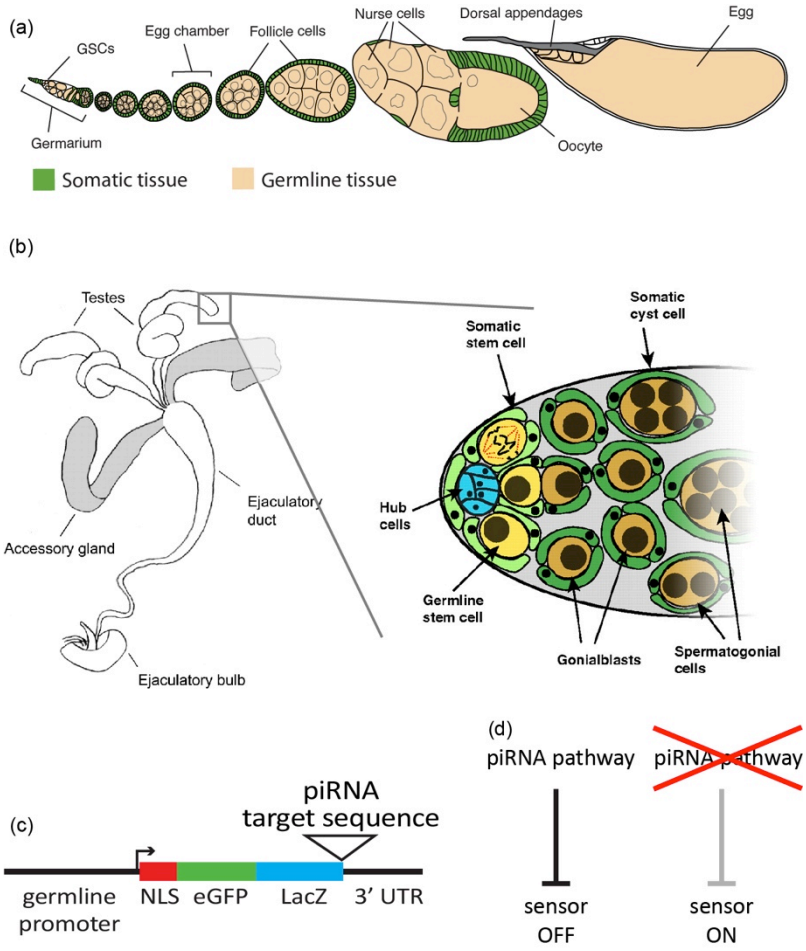


Figure 6 Overview of the *Drosophila* male and female reproductive organs and a cartoon showing the sensor system. The functional unit of an ovary is an ovariole as depicted in (a). It consists mainly of two cell types: germline cells (nurse cells and oocyte, in beige) and somatic cells (follicle cells, in green). Development starts in the germarium, which contains the germline stem cells (GSCs). (b) Cartoon showing testis: the GSCs are in contact with their nice. The GSC divedes further into a gonialblast, which further becomes the spermatogonial cell. All cells coming from the GSC lineage are surrounded by somatic cyst cells, whereas the GSC and the hub are surrounded by somatic stem cells. (c) Scheme of the sensor consisting of an NLS fused to eGFP and LacZ. A target insertion site is given in front of the 3' UTR. (d) Workflow of the sensor under different under normal situation with functional piRNA pathway and under impaired situation ((a) and (b) adapted from (Arbouzova and Zeidler, 2006; Ravi Ram and Wolfner, 2007)).

The optimal tissue specific promoter

As the first generation of sensors (using *vasa*/*piwi* promoter) was not successful I had to find an alternative. This new strategy included a 0.8 kb putative *nanos* promoter fragment, which has been demonstrated to drive expression of a reporter from early stages up to the oocyte in the germline only (Chen and McKearin, 2003; Ali et al., 2010; Van Doren et al., 1998). Following up this approach (Figure 7a), I was able to obtain a sensor, which is active mainly in germline tissue (see Figure 7b/c). The sensor (without a piRNA target) shows bright eGFP signal in early stages of development. All the nuclei in the germarium arising from the germline lineage are eGFP positive. To this end also later stages (egg chambers) as depicted in Figure 7b report remarkable eGFP signal in the polyploid nurse cells. In some ovarioles, egg chambers also show faint eGFP signal in few somatic cells at the posterior end. In all experiments the eGFP signal was overlapping with the DAPI signal. The X-gal staining (Figure 7c) revealed, that the sensor is expressed in testis. Staining gonads of female and male L3 larvae revealed the germline specific expression of the sensor (Figure 7c).

These results suggest, that the *nanos* sensor is an optimal backbone for the further generation and analysis of germline specific sensors containing distinct piRNA targets. As the sensor is not gender limiting it further provides a tool for analyzing the piRNA pathway in male testis. Most strikingly the sensor even shows expression in larval gonads, and therefore the whole development of reproductive organs in *Drosophila melanogaster* can be analyzed.

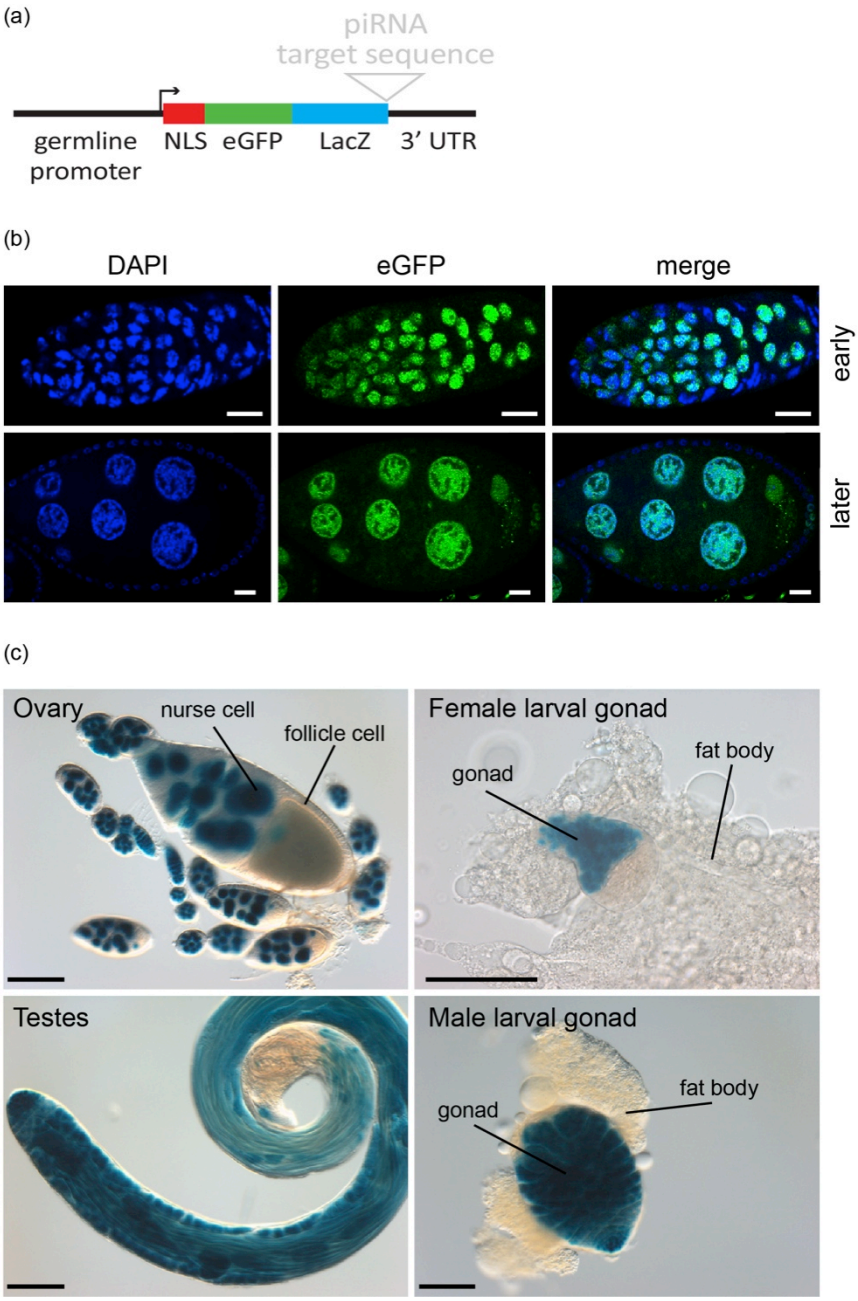


Figure 7 Overview of the *nanos* sensor and its reporter activity in *Drosophila* gonads. (a) Cartoon showing the germline specific *nanos* sensor without a piRNA target. (b) Images showing eGFP signal in early (germarium) and later stages (egg chamber) of ovariole development (white bar = 10 μm). (c) X-gal staining of ovaries, testis, and male/female larval gonads (black bar = 100 μm).

Cluster specific target regions

piRNA clusters are large genomic regions containing fragmented and nested pieces of TEs. In *Drosophila melanogaster* they have been discovered by looking at genome unique mapping piRNAs, which were bound by a protein of the PIWI-clade (Brennecke et al., 2007). For the generation of sensors, which detect the activity of a specific cluster, I created piRNA density profiles along the best charactarized clusters: *flamenco*, *cluster 20A* and *cluster 42AB* (Figure 8a, 9a and 10a). My aim was to find genome unique sequences giving rise to piRNAs. I was able to spot two regions within the *flamenco* cluster, two regions within *cluster 42AB* and one region in *cluster 20A* showing a convincing uniquess profile (Figure 8b, 9b and 10b). Comparing all mappers (i. e. all piRNAs mapping to this region; size > 22 nt) to unique mappers (i. e. only piRNAs arising from this genomic region; size > 22 nt and mapping-number = 1) I could select for regions which seemed to be more unique (*flamenco* region 1, *cluster 42AB* region 1) compared to less unique regions (*flamenco* region 2, *cluster 42AB* region 2). *Cluster 20A* was more difficult due to the lack of true 'unique' stretches. Nevertheless, I came up with the region depicted in Figure x. In addition to the piRNA density profiles I also performed bioinformatic analysis on the siRNA profiles (size = 21 nt) of these regions, to exclude for cross-talk from the siRNA pathway (data not shown). Finally the regions were further scanned to omit putative polyadenylation sites (i. e. AATAAA), which could lead to unfavored processing of the sensor transcript.

Previous work on a different set of germline/soma sensors performed in the lab, used target inserts about 1.5 kb. But it is important to mention, that such big inserts are prone to have siRNA target regions and are therefore not suitable. Hence, the target regions were selected between 300-450 bp in size, as I aimed to titrate out the smallest possible region needed for silencing the sensor.

Taking into account all the bioinformatics analysis and the target size issue, I followed up to generate sensors with cluster specific targets, based on the germline specific *nanos* sensor.

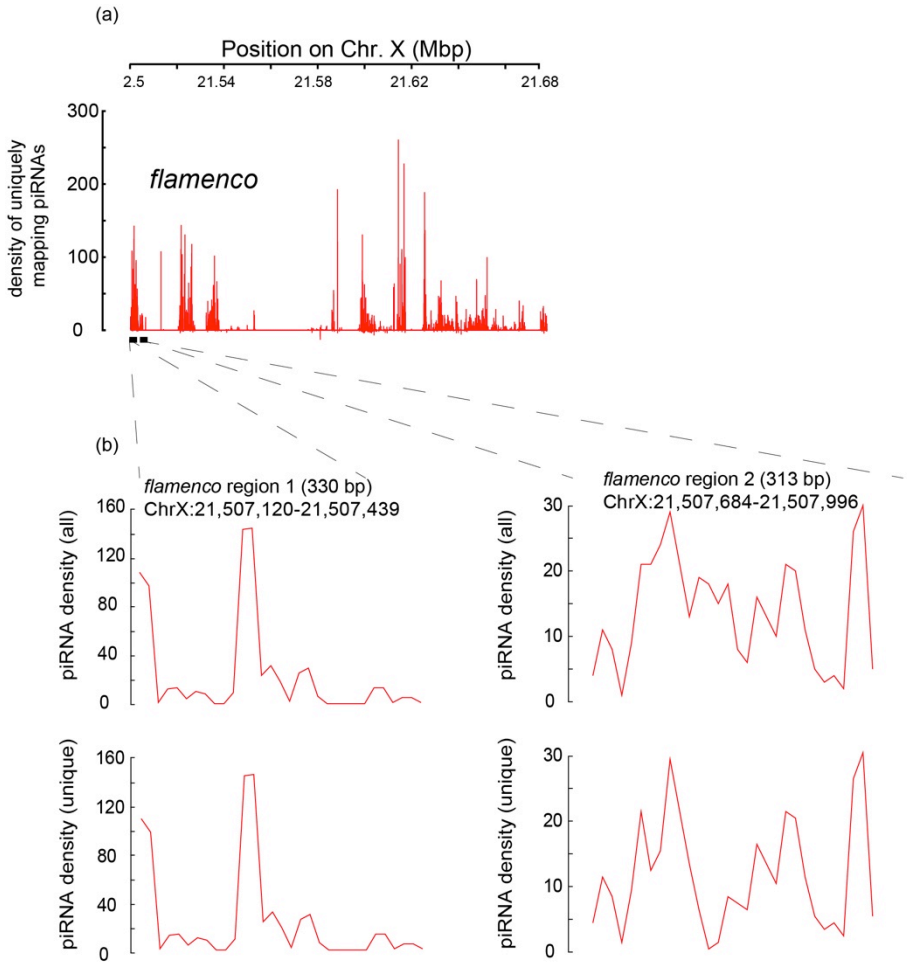


Figure 8 piRNA density profiles of the *flamenco* cluster. (a) Uniqueness piRNA profile of *flamenco*. (b) Comparison of all mappers and uniquely mapping piRNAs of selected target regions 1 and 2.

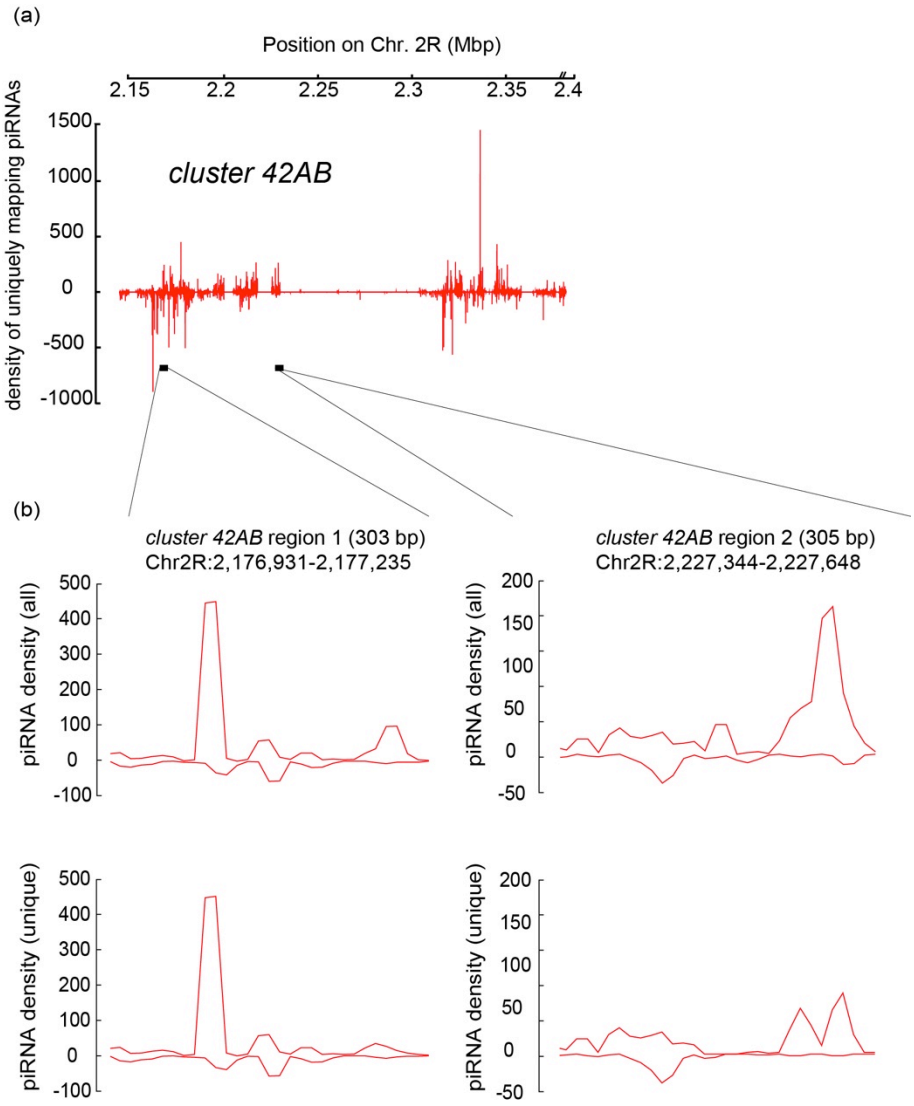


Figure 9 piRNA density profiles for cluster 42AB. (a) Uniqueness profile of *cluster 42AB*. (b) Comparison of all mappers and unique mappers of selected target region 1 and 2.

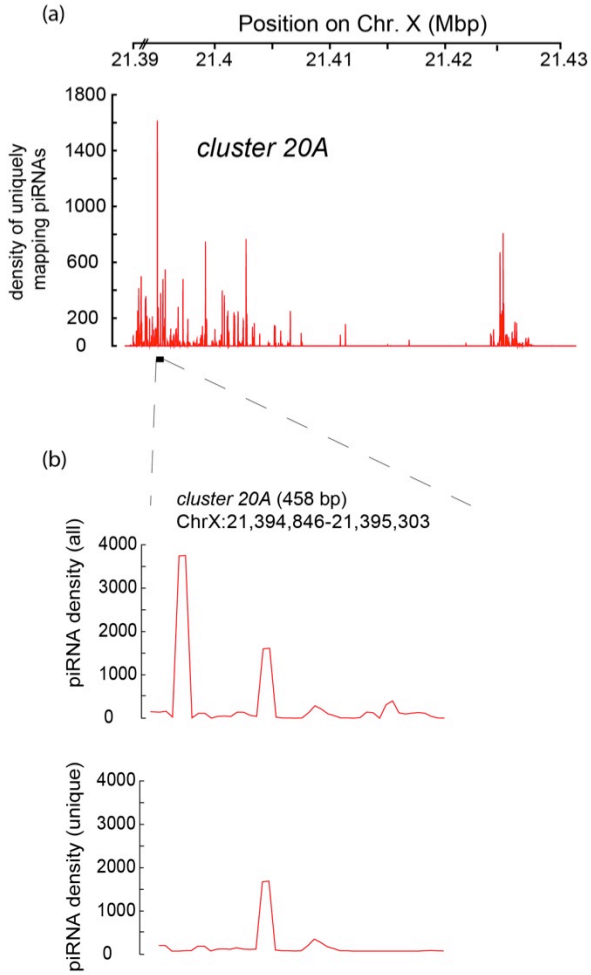


Figure 10 piRNA density profile of *cluster 20A*. (a) Uniqueness profile of *cluster 20A*. (b) Selected target region with either all mappers or uniquely mapping piRNAs.

Probing for the optimal insert

The empty *nanos* sensor (i. e. without target) reports tissue specific expression of either eGFP or LacZ in the germline (see Figure 7b/c, 11a). Upon addition of a piRNA target sequence (see Figure 11b) it is expected to lose the sensor signal.

The orientation of the target sequence in the sensor is crucial. Two of the analyzed clusters (*flamenco* and *cluster 20A*) are both thought to be unidirectionally transcribed (Brennecke et al., 2007). Both clusters have transposon fragments inserted antisense to the transcription direction. Therefore the extracted targets

(Figure 8b and 10b) have to be in antisense direction relative to the cluster sequence, in order to be recognized and silenced by the pathway. Nevertheless, the target regions of *flamenco* and *cluster 20A* were cloned in sense orientation as a control. The third piRNA cluster, which is important for this study, the germline specific *cluster 42AB* has been shown to be transcribed in both directions (Brennecke et al., 2007). Hence, the two selected target regions for *cluster 42AB* (Figure 9b) were cloned in sense, as well as in antisense direction relative to the cluster sequence.

For a more detailed analysis each target sensor was brought into a genetic background, which drives the expression of a short-hairpin (SH) and therefore induces a gene knock-down (for detailed description of the SH-system, please refer to (Ni et al., 2011)). Besides a control hairpin (ControlSH), which should not impair the piRNA pathway, two other hairpins were used, which upon activation deplete the levels of essential piRNA pathway components: Armitage (ArmiSH) and AGO3 (AGO3SH). This depletion should lead to derepression of the sensor.

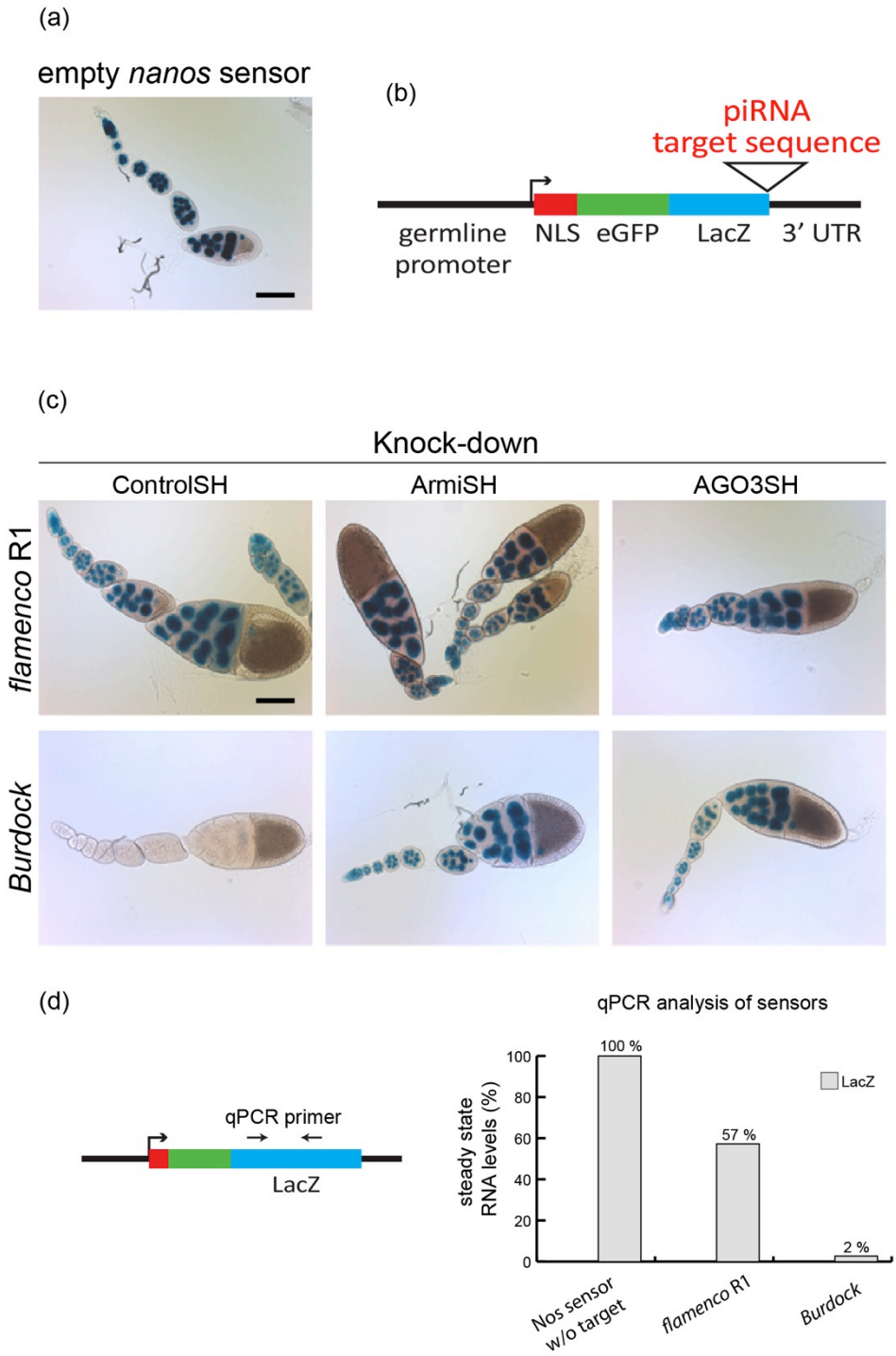


Figure 11 Applications of the *nanos* sensor upon insertion of a piRNA target. (a) X-gal staining of the empty *nanos* sensor. (b) Cartoon showing the insertion of a piRNA target into the sensor. (c) Probing for the piRNA pathway to act on the sensor in certain knock-down backgrounds. (d) qPCR as a read out for sensor expression levels. (black bar = 100 µm)

Upon insertion of a 300 bp fragment antisense to the *flamenco* cluster (*flamenco* R1) I expect no downregulation of the sensor, as *flamenco* has been suggested to be soma specific (Brennecke et al., 2007).

Figure 11c shows the *flamenco* R1 sensor in control and knock-down background along with a *nanos* sensor containing an 800 bp fragment of the TE *Burdock*. The *flamenco* R1 sensor gets only slightly repressed in the germline compared to the empty *nanos* sensor (Figure 11a). Using quantitative PCR (qPCR) as a readout for sensor expression, I observed the *flamenco* R1 sensor to be expressed to 57% of the levels of the empty *nanos* sensor (Figure 11d). The qPCR results have to be treated with caution, as due to time constraints no biological replicate was performed.

The *Burdock* sensor in contrast displays the behavior of an optimal sensor: fully silenced in control background, desilenced upon knock-down of Armitage or AGO3. To this end, also the expression levels detected by qPCR are correlating: the *Burdock* sensor is only expressed to about 2% of the levels of the empty *nanos* sensor (Figure 11d).

From these results it becomes clear, that the *nanos* sensor performs perfectly upon addition of an appropriate piRNA target: no signal in normal situation, signal upon impairment of the piRNA pathway. In addition the *nanos* sensor is detectable during all stages of development.

Cluster 20A sensors

The unidirectionally transcribed *cluster 20A*, which is thought to be expressed in both tissues in the ovary (Brennecke et al., 2007) does not contain suitable stretches of sequence fulfilling the criteria for a target sequence (unique profile, no AATAAA, no siRNA target). The 450 bp region selected for cloning into the *nanos* sensor is therefore not optimal to show cluster specific expression, but still is the best region I found for my analysis. The region was cloned in both orientations to get yet another control for my experiments. To show, that the regulation of the sensor is due to the the piRNA pathway, *the cluster 20A* sensors were crossed into knock-down backgrounds (ControlSH, ArmiSH, AGO3SH).

The results I obtained with this sensor were very surprising. The control sensor (target region in sense orientation) showed no repression in the normal background (ControlSH) and was not altered in the knock-down background (ArmiSH, AGO3SH) (Figure 12). In contrast, the *cluster 20A* sensor with the target cloned in antisense direction was completely silenced in the normal background (like the *Burdock* sensor). After depletion of AGO3 the *cluster 20A* sensor was fully desilenced as indicated by the nuclear X-gal stainings. However, in an Armitage depleted background the *cluster 20A* sensor was only faintly (but detectably) desilenced in early stages. Quantification of the *cluster 20A* sensor expression by qPCR was not possible due to time constraints.

Looking at these results, it is crucial to acknowledge that the sensor approach can reveal novel information about the behavior of a certain cluster upon removal of an essential piRNA pathway factor. But it is also crucial to carefully analyze such serendipity to overcome any misleading interpretations.

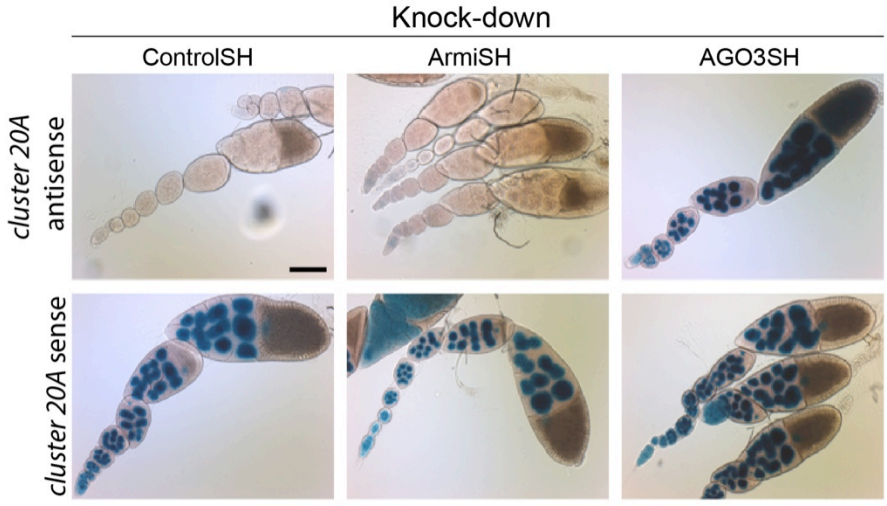


Figure 12 *Cluster 20A* sensors in knock-down backgrounds. Upper panel shows expression levels of the sensor with the target in antisense direction. Lower panel represents target in sense direction (black bar = 100 µm)

Cluster 42AB sensors

The 260 kb long *cluster 42AB* is bidirectionally transcribed and active in the germline tissue only (Brennecke et al., 2007). Hence, this cluster is best suited for analysis with the germline specific *nanos* sensor. To this end I have choosen two target regions that differ in their uniqueness profile to get a glimpse in how important the

uniqueness is for the sensor. Figure 13 represents the analysis of the two target regions in both orientations each. All sensors were tested in normal (ControlSH) and knock-down background (ArmiSH, AGO3SH). Unfortunately, the extent of repression in the various sensors was rather low. Apparently the target size (300 bp) was selected to small, as only one sensor (*cluster 42AB* R2 antisense) reports detectable repression and derepression. In all the other cases the expression of the sensor was not significantly changed (Figure 13). Quantification of the sensors using qPCR was not possible, because of time constraints.

From these sensors it gets clear that a target size of 300 bp seems to be too small, to show desired silencing/desilencing of a *cluster 42AB* sensor.

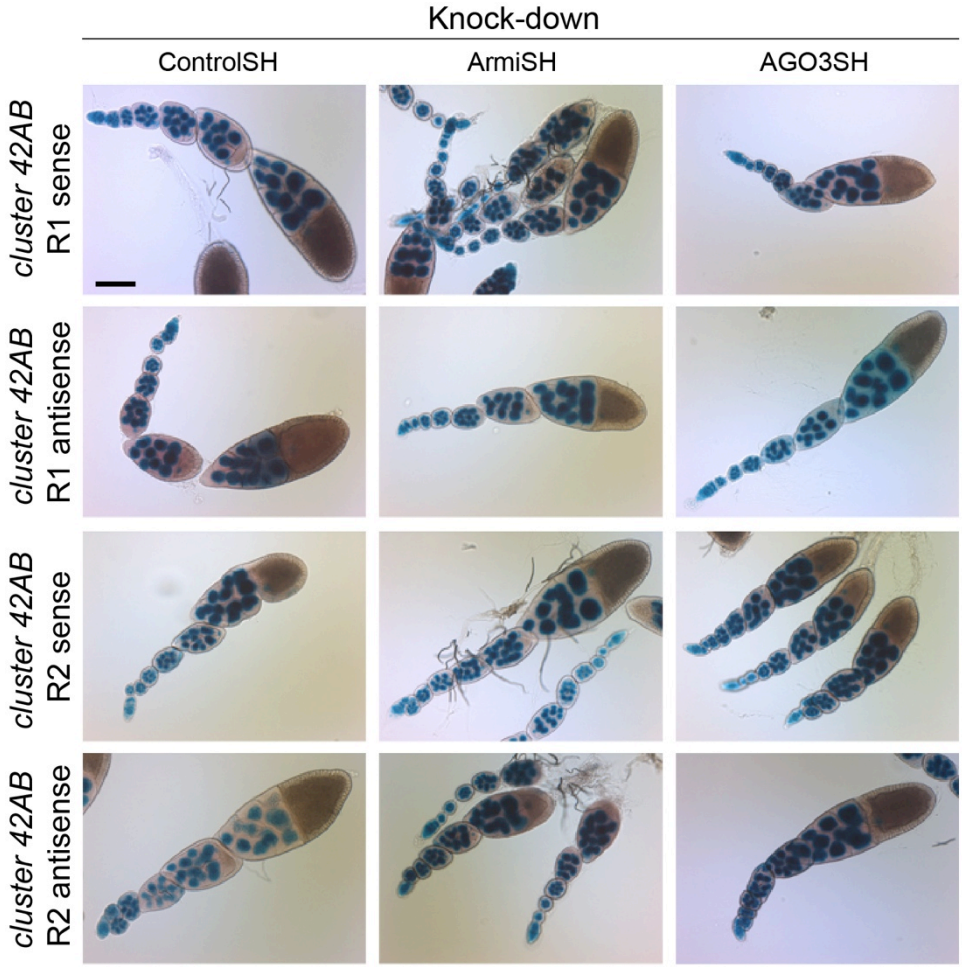


Figure 13 Analysis of cluster 42 sensors. Both target regions were analyzed in sense as well as in antisense orientation. Only *cluster 42AB* R2 antisense showed a detectable downregulation in the control background (last panel). (black bar = 100 µm)

The putative *flamenco* promoter

An alternative strategy to look at the spatiotemporal expression of piRNA clusters would be by mapping their promoter regions. Up to now there is no evidence, how and when a piRNA cluster is transcribed. It is not known, whether there is a special transcription machinery and where around the cluster it is assembled. It is also not clear, if a cluster has one single or multiple promoters to generate a precursor transcript, which is further parsed into smaller piRNAs.

Out of this I wanted to see, whether it is possible to drive a reporter upon cloning several upstream sequences (differing in size) of piRNA clusters (Figure 14a). By looking at a global nuclear run on sequencing data (GRO-seq) generated previously in our lab, I cloned different pieces with putative promoter activity (Figure 14a).

Starting from the nearest upstream gene (i. e. *DIP1*) of *flamenco*, four constructs (ranging from about 2-3.7 kb) were cloned into a vector containing the NLS-eGFP-LacZ reporter and a 400 bp ribosomal 3'UTR (rpl32).

Figure 14b shows the outcome, based on these reporters. Construct 1 drives the reporter in ovaries in germline tissue only as indicated by blue nurse cell nuclei. The signal was stronger in later stages of development. This was unexpected, as *flamenco* is expected to be active in the somatic tissue only (Brennecke et al., 2007). The testis also showed staining in some cells. The staining pattern looks different compared to the empty *nanos* sensor in testis (Figure 7c and 14b). Construct 2 showed no reporter signal in ovaries or testis. Construct 3 displayed the expected outcome in ovaries: somatic staining only. It is important to note, that construct 3 leads to a patchy expression of the reporter (i. e. stretches of somatic follicle cells lacked the reporter expression). In contrast to construct 1, the X-gal staining in testis of construct 3 resembles more the pattern of the empty *nanos* sensor in testis (Figure 7c and 14b). The last *flamenco* promoter construct analyzed (construct 4), had no visible X-gal staining in ovaries but some nuclei in testis showed blue X-gal signal.

These experiments hinted at certain elements upstream of a piRNA cluster to be sufficient for driving the expression of a reporter. Still these results have to be taken with care can only serve as a starting point to further dissect the *flamenco* cluster promoter.

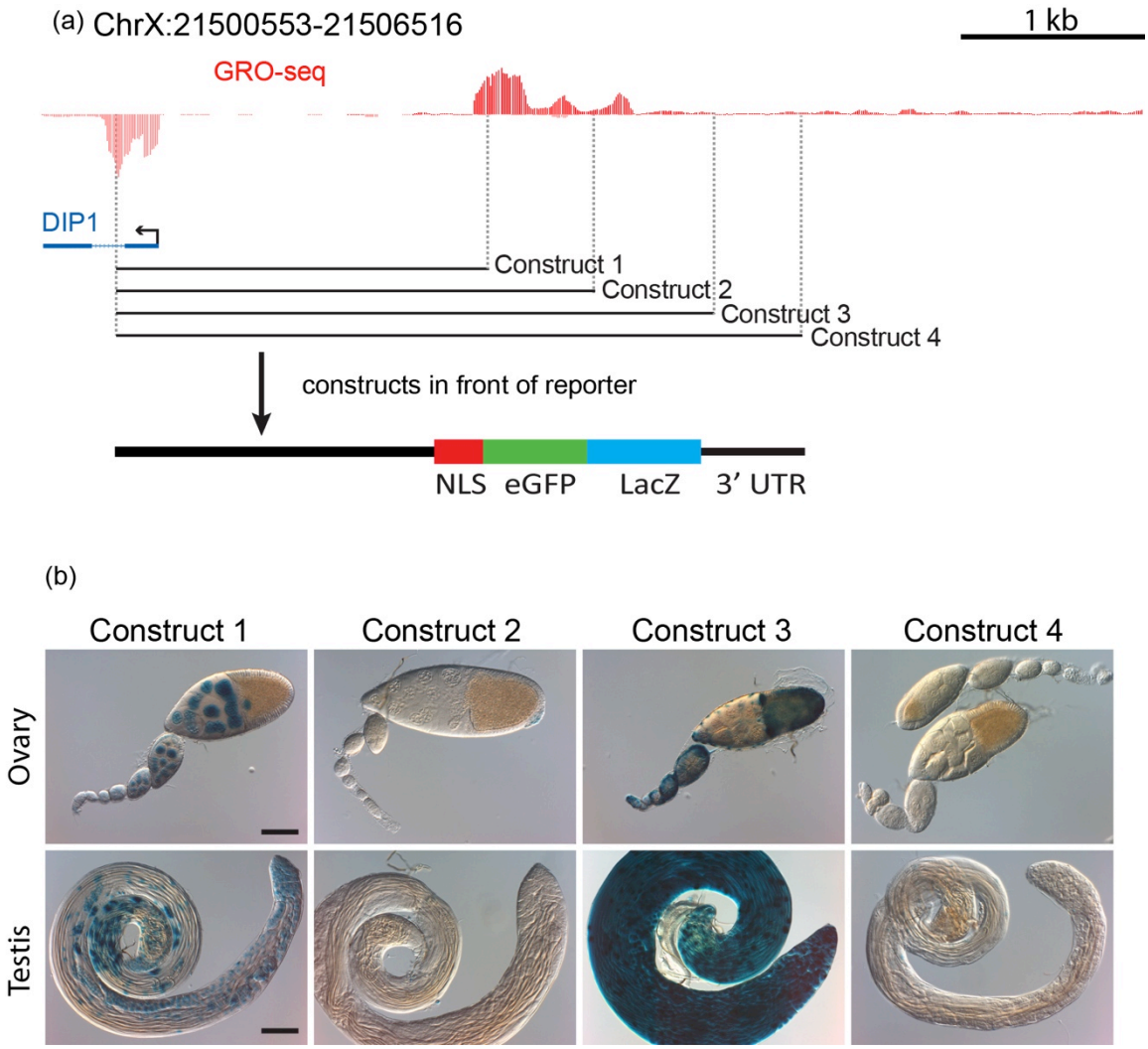


Figure 14 Design and results of promoter mapping studies of *flamenco*. (a) GRO-seq profiles plotted at the region around a putative *flamenco* promoter. A cartoon depicts the cloning into a vector containing the reporter. (b) X-gal stainings of ovaries and testis for the indicated constructs (black bar = 100 µm)

The putative *cluster 20A* promoter

Following up the same strategy as for the *flamenco* promoter studies, I cloned four different constructs for *cluster 20A* (Figure 15a), the second major uni-directional piRNA cluster (ranging from 0.5-3.5 kb) to see their potential of driving a NLS-eGFP-LacZ reporter. Here, only construct 2 showed expression (Figure 15b) of the reporter in ovaries as well as in testis. In ovaries construct 2 is expressed in both tissues, germline and soma. The soma signal is noticeable weaker compared to e. g. the

signal in the *flamenco* promoter construct 3. In testis *cluster 20A* promoter construct 2 drives an expression pattern similar to the empty *nanos* sensor. All the other constructs for this promoter analysis were not able to drive expression of the reporter.

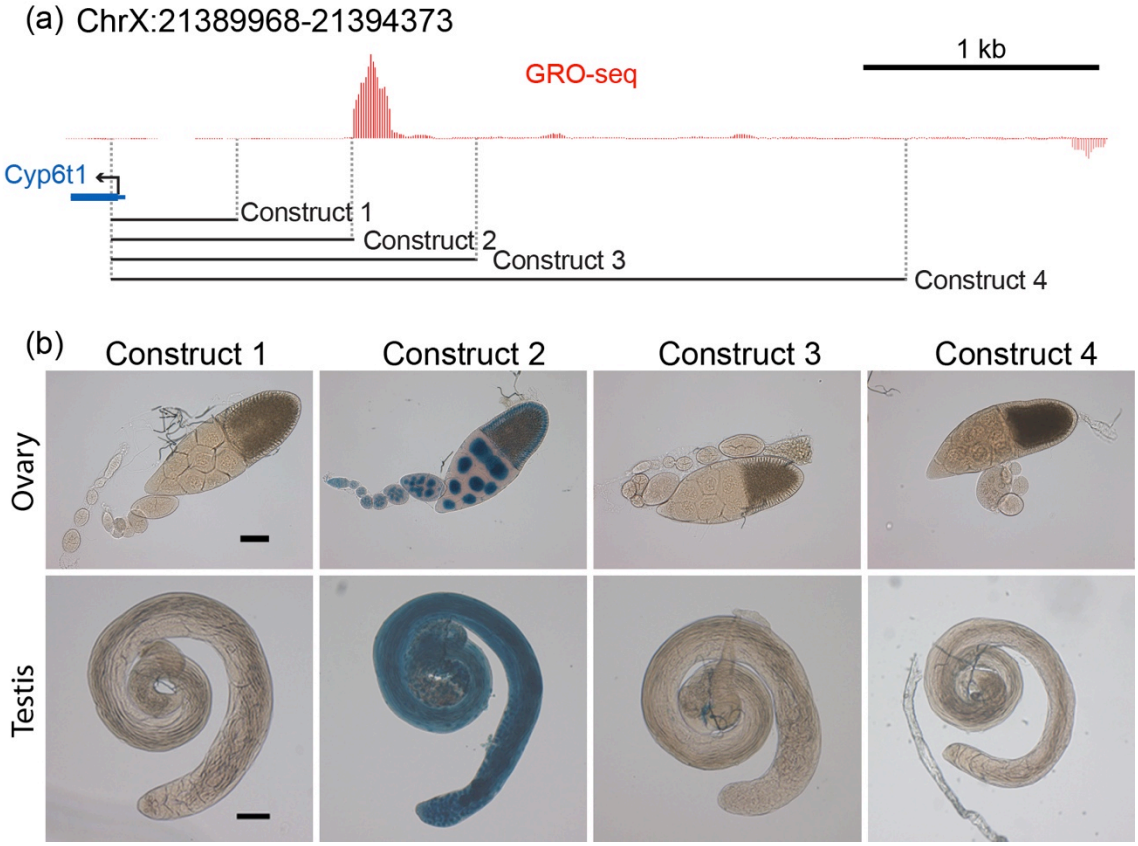


Figure 15 Design and results of promoter mapping studies of *cluster 20A*. (a) GRO-seq profile plotted at the region around a putative *cluster20A* promoter. A cartoon depicts the extents of the various promoter fragments. (b) X-gal stainings of ovaries and testis for the indicated constructs (black bar = 100 µm)

Quantification of the promoter constructs

The expression of the NLS-eGFP-LacZ reporter by putative promoter fragments has been shown in the previous sections. To get more information why certain constructs did not express the reporter I performed qPCR on total RNA of ovaries. Figure 16 shows the steady state RNA levels of the indicated reporters in comparison to the empty *nanos* sensor. Two independent experiments were conducted using either the constructs for the *flamenco* or *cluster 20A* promoter mapping studies. The expression level of *flamenco* construct 1 (93%) is almost as high as the empty *nanos* sensor (100%). *Flamenco* construct 3 shows reduction (38%) in expression by more than half of the levels of construct 1. The *flamenco* constructs 2 and 4 did not show any significant transcript levels. *Cluster 20A* construct 2 is expressed (73%) almost to levels as the empty *nanos* sensor. Compared to this, *cluster 20A* constructs 3 (17%) and 4 (6%) show only weak RNA levels, whereas construct 1 is almost not expressed. Although these results need to be confirmed, the qPCR analysis correlates with the corresponding X-gal stainings.

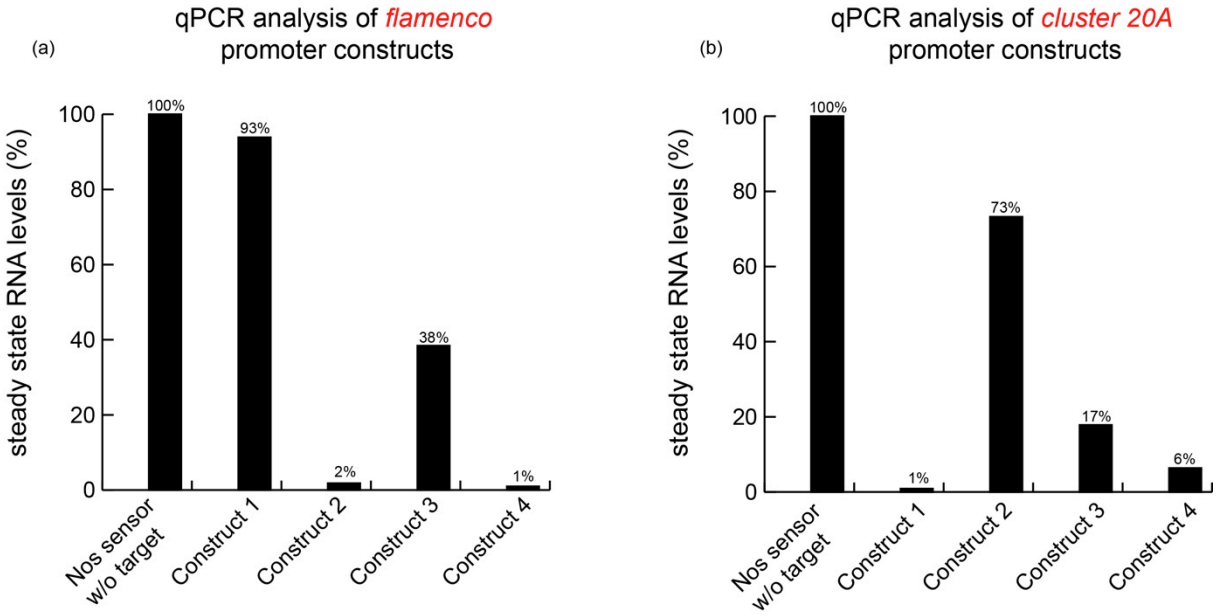


Figure 16 Quantitative analysis of the piRNA cluster promoter constructs. Comparing steady state RNA levels of the *nanos* sensor without target to (a) *flamenco* and (b) *cluster 20A* constructs.

Localization of piRNA cluster transcripts

piRNA clusters are believed to be transcribed as a long single stranded precursor molecule. Evidence for this stems from *P*-element insertions upstream of the first piRNAs mapping to the *flamenco* cluster (Robert et al., 2001; Pélisson et al., 1994; Desset et al., 2003; Prud'homme et al., 1995; Brennecke et al., 2007). Where within the cell this precursor transcript localizes to get parsed into smaller pieces is currently unknown. One would predict the cluster transcript to get transported to Yb-bodies, where essential piRNA pathway members reside (Olivieri et al., 2010; Saito et al., 2010). Another putative site for cluster transcript processing would be the nuage in germline cells. This perinuclear electrondense structure has also been shown to contain certain piRNA pathway members with unknown molecular function e. g. Avocado (Brennecke et al., 2007; Handler et al., 2011).

To follow up these hypothesis I aimed to localize cluster transcripts using different sensitive approaches (Figure 17). Conventional *in situ* hybridization should be the first method of choice. By generating RNA probes, which are either directly labeled with a fluorophore (e. g. Alexa488®) or labeled with a hapten (e . g. digoxygenin), I should be able to detect the transcript with a certain amount of specificity and sensitivity (Figure 17a). To further increase those two parameters I used an approach based on the branched DNA (bDNA or Christmas trees, see Figure 17b) hybridization method (Itzkovitz and van Oudenaarden, 2011). Besides the bDNA strategy, I also wanted to use another recently emerging technology called single-molecule FISH (smFISH or stellaris, see Figure 17c) (Itzkovitz and van Oudenaarden, 2011). Due to time constraints it was not possible to test this system properly and use it for relevant experiments.

All the following experiments were performed in cultured ovarian somatic cells (OSCs) (Niki et al., 2006; Lau et al., 2009). The RNA probes against the *flamenco* cluster were designed according to sequenced regions to exclude for SNPs not listed in the published *Drosophila* genome.

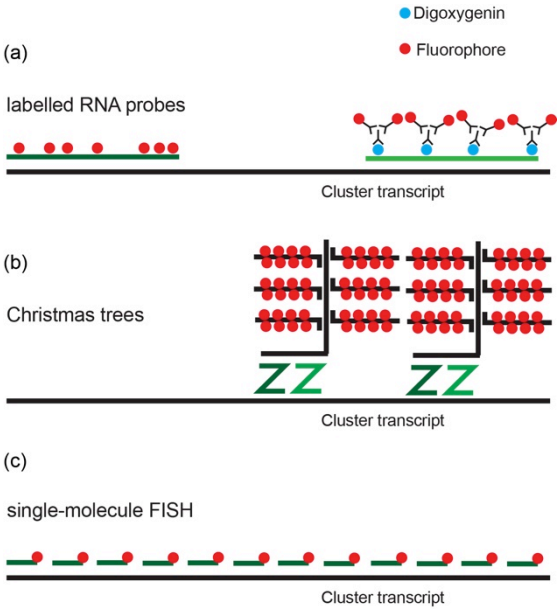


Figure 17 Overview of detection methods. (a) Conventional *in situ* probes, which are either directly labeled or labeled with digoxygenin. (b) bDNA approach: several layers of amplification increase signal intensity. No antibody background is given by the fact that only DNA molecules are used for amplification. About 20 of this so built trees bind a transcript target size of 1 kb. (c) smFISH uses about 50 small oligos (20 nt) which contain one fluorophore. These 50 probes detect a transcript target size of about 1 kb. This approach in theory should have the best signal to noise ratio, as only one hybridization step and no antibodies are used.

Conventional *in situ* reveals site of transcript localization

To optimize the protocol for *in situ* hybridization, the OSC cell line was used (Niki et al., 2006; Lau et al., 2009). These cells are easy to handle and offer a first estimation how specific the probes are and where the transcript localizes. As the OSCs represent ovarian somatic cells (e. g. they express only Piwi and have no ping-pong pathway) and *flamenco* is believed to be active only in somatic cells, I wanted to establish the *in situ* assay first using probes against *flamenco*. For this purpose I generated three probes specific to the *flamenco* transcript: a 0.85 kb probe at the beginning of the transcript (Fbeg, according to RNA seq-data), a 1.4 kb (F1) and a 0.8 kb (F2) probe in unique regions (according to piRNA density profiles). Mapped to the genome, Fbeg and F1 are about 2 kb apart and F1 and F2 are about 20 kb apart. In order to see how reliable the probes are in detecting the desired transcripts vs. unspecific background, I always combined two probes: one probe, which was directly labeled (DL) and one probe, which was labeled with a hapten (HL), that was detected using antibodies. This allowed me to use two different fluorophores, which should accumulate at the same sites. To this end the *in situ* hybridization experiments were

followed by an antibody staining for Armitage to localize Yb-bodies, the proposed piRNA biogenesis loci.

In a GFP siRNA knock-down (control) I was able to obtain weak but detectable signal appearing as dots, using either Fbeg, F1 or F2 probes. Combining HL-Fbeg with DL-F1 or DL-F1 with HL-F2 the dots coming from the corresponding channels colocalize to the same area (Figure 18). More interestingly the transcripts are most of the times close to or also overlapping with an antibody staining for Armitage (Figure 18).

The Yb-body has been shown to gain in size from small foci to a big cloud upon depletion of Zucchini using siRNAs in OSCs. This is thought to be Piwi, which is not able to enter the nucleus any longer. In order to see, what happens to the cluster transcript localization, I performed Zucchini knock-down experiments in parallel (Olivieri et al., 2010; Pane et al., 2007; Huang et al., 2011; Watanabe et al., 2011). The bright dots (HL-Fbeg, DL-F1) change their appearance accordingly to staining of Armitage upon a Zucchini siRNA knock-down: the tiny dot becomes a big cloud, which resembles accumulation of Piwi protein (Olivieri et al., 2010) (Figure 19a).

The Yb-body has already been described as the putative site for piRNA biogenesis (Olivieri et al., 2010; Saito et al., 2010). To see, what happens to cluster transcript localization after depletion of one of the Yb-body components, a Fs(1)Yb knock-down was performed in parallel in each experiment (Figure 19b).

Fs(1)Yb siRNA knock-down does not show any remarkable change in transcript localization close to nucleus (HL-Fbeg, DL-F1), but the Armitage staining is gone completely.

The conventional *in situ* hybridization method presents an useful assay to analyze cluster transcript localization. The obtained signals detected using two independent RNA probes against the *flamenco* transcript perfectly overlap in most of the experiments. The signal intensity however is not very strong. The background levels in HL vs. DL probes gives a good estimate about the specificity of the hapten antibody staining. DL probes show much less unspecific background signal, but the signal itself is weaker compared to HL probes.

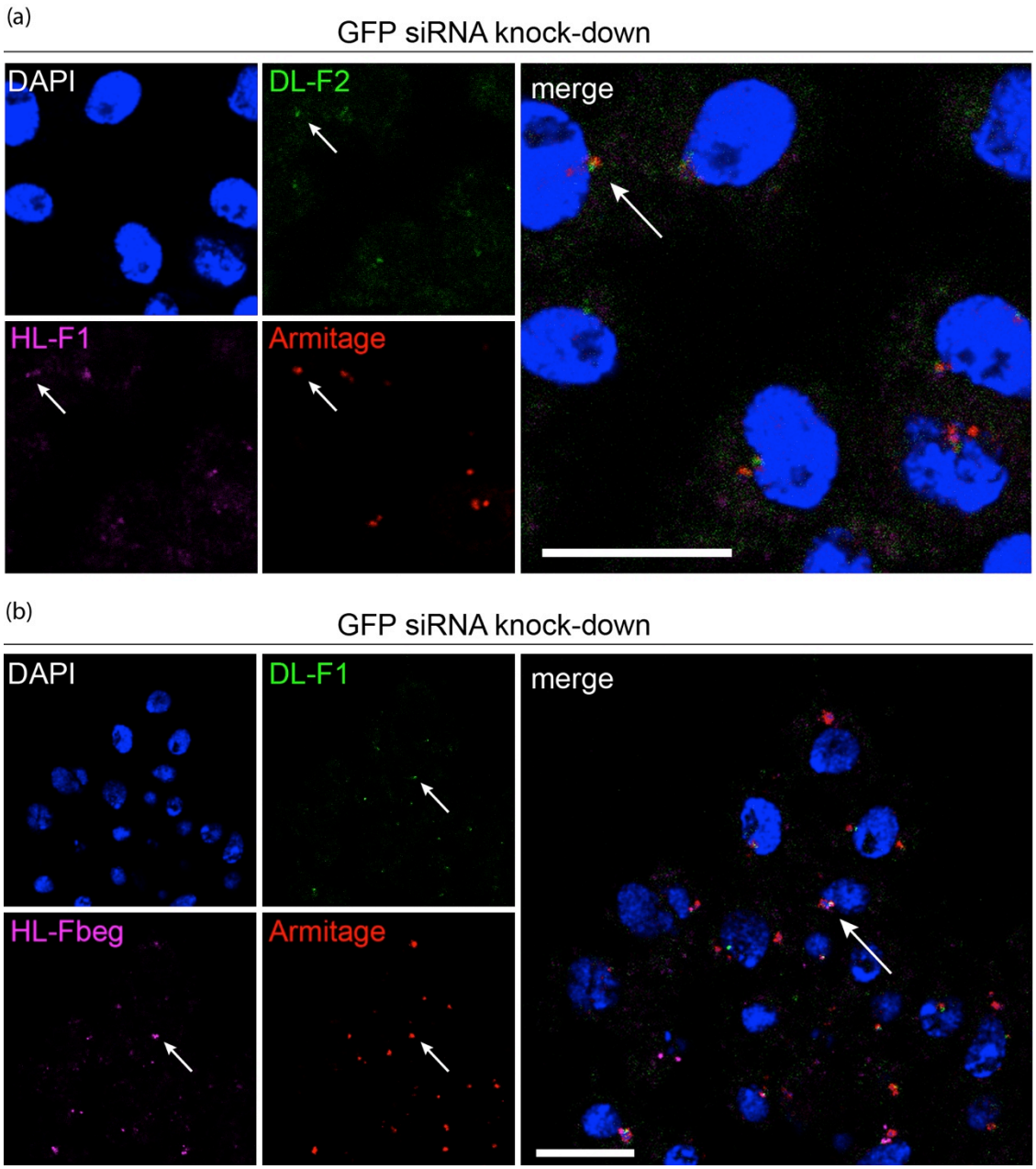


Figure 18 Testing different RNA probes to detect the *flamenco* transcript. Conventional *in situ* hybridization combining either F1 and F2 (a) RNA probes or Fbeg and F1 (b) RNA probes. The blow-up presents the merge of all channels. (white bare = 10 µm)

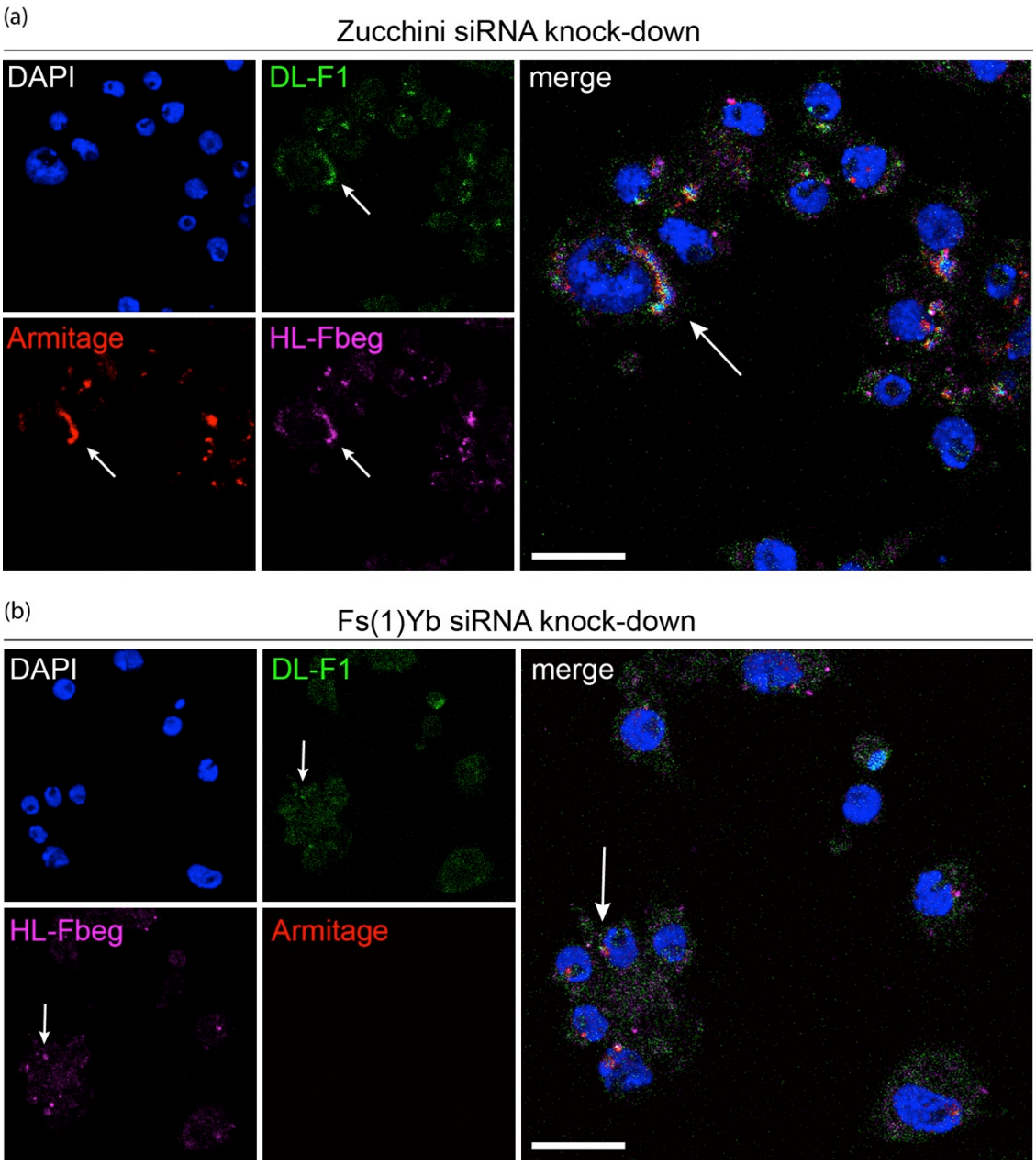


Figure 19 Conventional *in situ* hybridization upon siRNA knock-down of piRNA pathway members. (a) Zucchini knock-down combining two RNA probes against *flamenco*. (b) Fs(1)Yb knock-down using two RNA probes against *flamenco*. (white bar = 10 µm)

Christmas tree approach reveals the *flamenco* cluster transcript

The observed signal from the conventional *in situ* hybridization experiments was not optimal in terms of intensity. Therefore I followed a more specific and sensitive approach based on the bDNA technology (provided by www.panaomics.com).

According to the manufacturer this methodology is able to detect single transcripts with extremely low background levels. To test this method's usability for probing a piRNA cluster transcript, probesets were obtained to detect the *flamenco* transcript (chrX:21,506,662-21,507,875). The experiments were perfomed in control (GFP), Zucchini and Fs(1)Yb siRNA knock-downs in OSCs. Leaving away the initial probeset, which detects the cluster transcript, resulted in very low level of background signal (Figure 20a).

The detected signal using the *flamenco* probeset was very bright. The appearing dots are almost as big as the Yb-bodies (Figure 20b). The staining often showed two bigger dots within the nuclei. This could display the site of transcription. In almost every cell a big *flamenco* dot is very close to an Yb-body at the nuclear rim (this was tested by staining for lamin, data not shown). Very often smaller dots appeared within the Armitage staining.

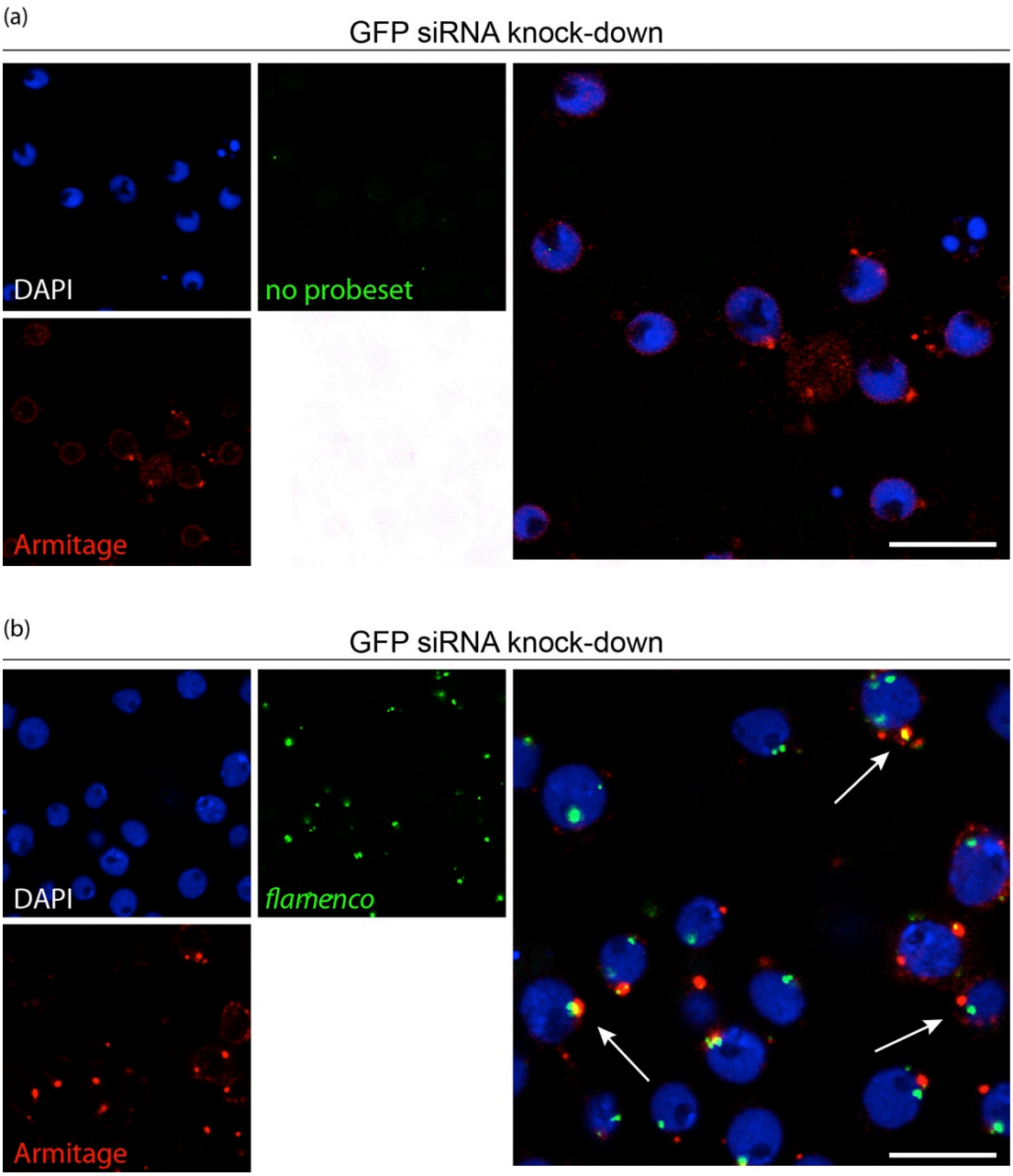


Figure 20 Localization of *flamenco* cluster transcript upon GFP control knock-down using Christmas trees. (a) no probeset control. (b) *flamenco* probeset detecting 1 kb of the cluster transcript. (white bar = 10 μm)

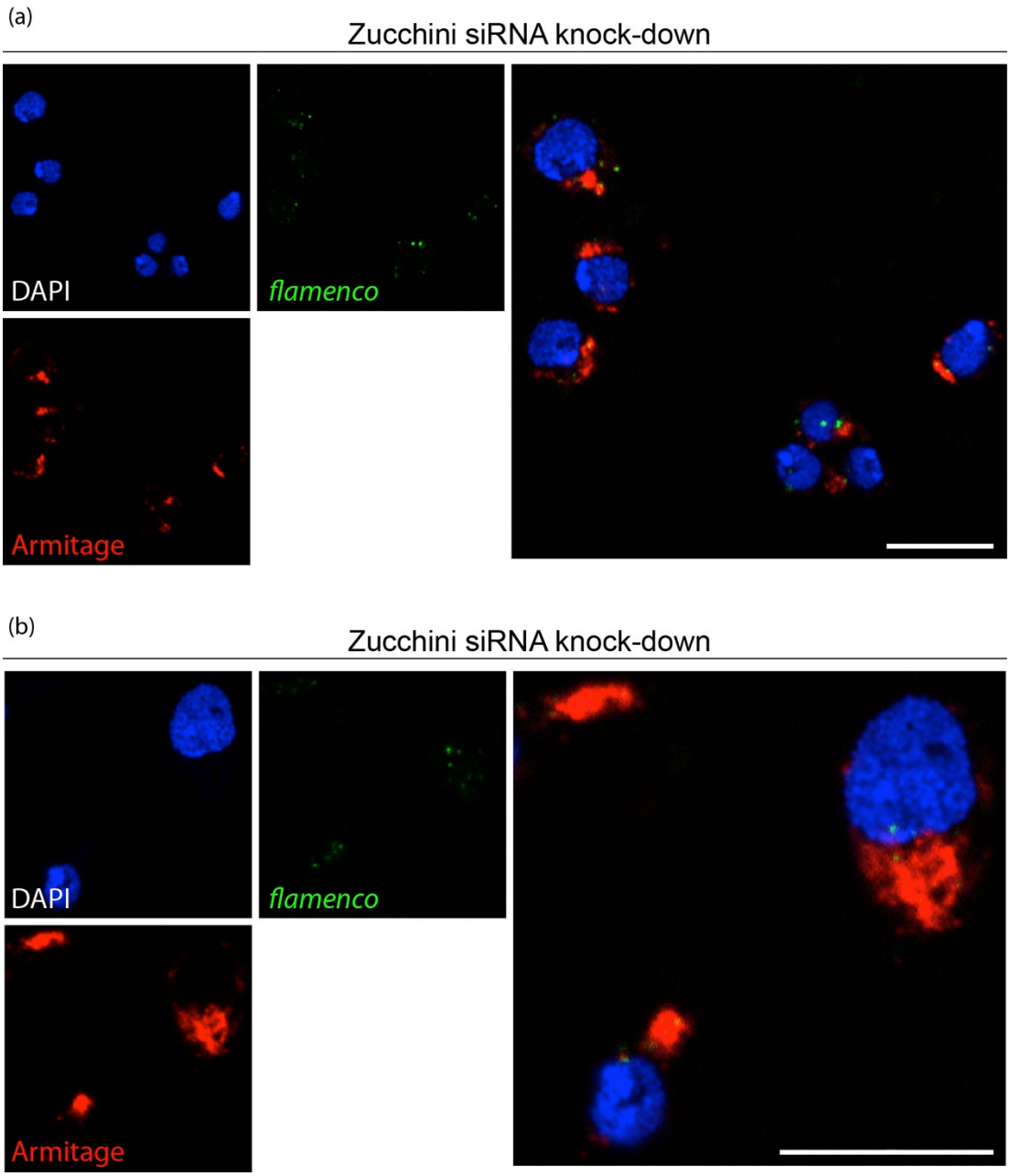


Figure 21 Christmas trees showing *flamenco* transcript localization after siRNA knock-down of Zucchini. (a) Blow-up represents merge. (b) Blow up represent zoom into a specific region. (white bar = 10 µm)

Upon Zucchini siRNA knock-down, the Armitage staining appeared as a big cloud, as expected. The *flamenco* transcript was detectable within this cloud in serveral small foci (Figure 20). The nuclear signal was still detectable in some cells, which showed again Armitage staining close to it (Figure 20).

Performing the Christmas tree approach in a Fs(1)Yb knock-down background revealed the following situation: the nuclear dots were still detectable but appeared smaller than those in the control background (Figure 19b and 22). The Armitage accumulation to Yb-bodies was lost after the siRNA knock-down. To this cytoplasmic *flamenco* signal, was observed, which was dispersed throughout the cell.

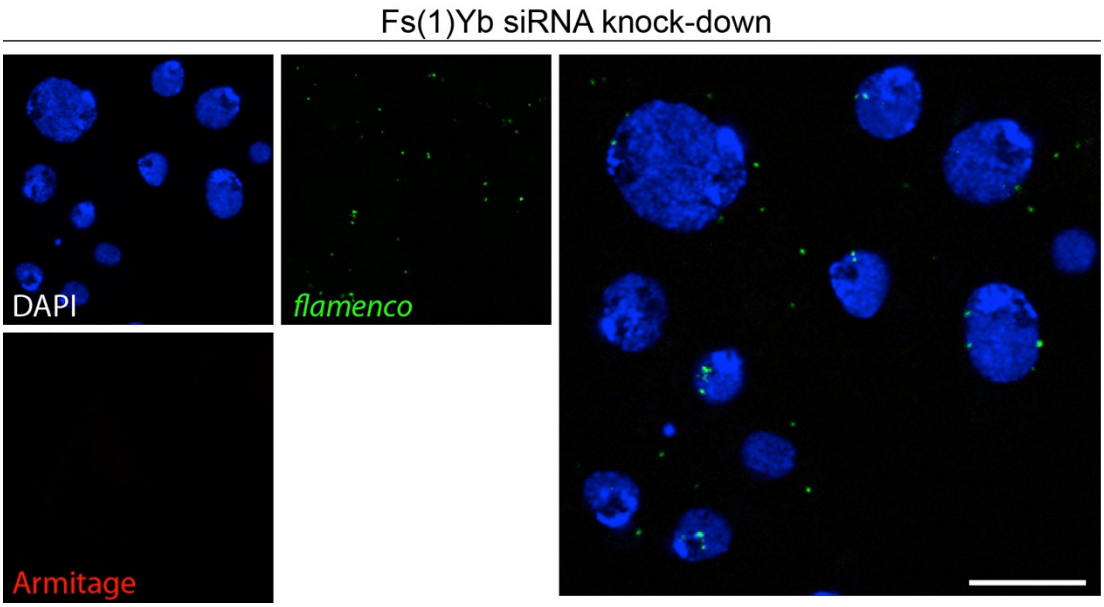


Figure 22 Christmas tree detecting *flamenco* upon Fs(1)Yb siRNA knock-down. Blow-up picture represents a merge of all channels (white bar =10 µm).

Taken together, localizing the *flamenco* transcript in OSCs using the Christmas tree method boosts the detectable signal enormously compared to the conventional *in situ* hybridization signal.

Discussion

The majority of piRNAs arise from long single stranded precursor molecules encoded by so called piRNA clusters. In *Drosophila melanogaster* piRNA clusters are genomic regions, which contain a plethora of incomplete, degenerate and immobile transposon fragments and these regions serve as a genetic memory for silencing of TEs (Brennecke et al., 2007). Important for this study are *flamenco*, *cluster 20A* and *cluster 42AB*, the three best characterized piRNA cluster in *Drosophila melanogaster*. piRNA clusters usually reside close to the euchromatin-heterochromatin boundary. It is therefore a favorable model, that piRNA cluster seem to have a special underlying epigenetic mark. One such epigenetic regulation has already been described for *cluster 42AB* and it involves a HP1 family protein encoded by the *rhino* locus (Klattenhoff et al., 2009). Besides this it has also been shown, that piRNAs from *cluster 42AB* arise from both genomic strands, whereas those from *flamenco* and *cluster 20A* map only to one genomic strand (Lau et al., 2009; Malone et al., 2009; Brennecke et al., 2007). Our current knowledge about the biology of piRNA clusters is very restricted. Hence, the aim of this study is to elucidate novel information about the three major piRNA clusters *flamenco*, *cluster 20A* and *cluster 42AB*. To do so, the establishment of new tools was essential for revealing further biological data about piRNA clusters.

A germline specific sensor is able to measure piRNA cluster activity

In this study I was able to generate a new tissue specific sensor, which is regulated by the piRNA pathway. This sensor utilizes a 0.8 kb portion of the *nanos* promoter (Ali et al., 2010; Chen and McKearin, 2003; Van Doren et al., 1998), which drives the expression of a NLS-eGFP-LacZ reporter in germline tissue in *Drosophila melanogaster* gonads. Due to the dual reporter system, the signal output is either bright eGFP or X-gal signal within nurse cell and oocyte nuclei. Very rarely the sensor is expressed in some posterior somatic follicle cells in later stages of development. It is possible that the *nanos* promoter fragment in isolation is not as

accurately repressed in non-germline cells as the neutral *nanos* locus. Nevertheless, the remarkable cell type specificity was fully sufficient for my studies.

Upon insertion of a cluster specific target sequence into the 3'UTR downstream of the NLS-eGFP-LacZ reporter, the sensor was silenced and reporter signal was lost to a certain extent. The silencing is thought to come from the piRNA pathway. To test for this hypothesis, I tested my sensor in different genetic backgrounds, which were defective in the piRNA pathway. Using the short-hairpin system (Ni et al., 2011) to generate knock-downs of certain parthway members, I could show that the sensor was desilenced again. This proved the piRNA pathway for being able to silence the sensor. With a canon of different piRNA cluster sensors I could demonstrate the activity of *flamenco*, *cluster 20A* and *cluster 42AB*.

The *flamenco* R1 sensor, which contains a 300 bp cluster specific sequence, showed only a slight downregulation. Looking at the expression levels of this sensor by qPCR analysis, the downregulation seemed to be stronger, compared to the *nanos* sensor without target insert. *Flamenco* has been reported to be active in somatic tissue only (Brennecke et al., 2007; Malone et al., 2009). Nevertheless evidence for this came from comparing piRNA sequencing data of libraries from ovaries and early embryos (which should resemble the pool of germline piRNA only) (Malone et al., 2009). From this data it cannot be excluded that *flamenco* is only expressed in somatic cells. It could be, that this cluster is expressed during early development to a certain amount but turned off in later stages of development. To test for this hypothesis, one could perform *in situ* hybridization in ovaries to see, whether a *flamenco* transcript is also detectable in germline cells as well.

The *cluster 20A* sensor mirrored an optimal sensor. The 450 bp sequence of *cluster 20A* lead to complete silencing of the sensor. Upon knock-down of an essential piRNA pathway factor (AGO3) the sensor was completely desilenced. Inserting the target region in the opposite direction after the NLS-eGFP-LacZ reporter showed the unidirectionalty. Now the sensor did not show any change in reporter levels with and without knock-down. It is important to notice that this sensor is not an optimal sensor for looking at cluster specific activity. Taking a closer look at the region selected for this cluster, one can find parts of a *ROO_LTR* and a *ROO_I* element. This may already hint the unexpected behavior of the sensor in an Armitage knock-down

background. Here only the early stages in an ovariole were desilenced, the later stages showed no reporter signal. The *roo* LTR-retroelement has been shown to respond to Spn-E mutant (Malone et al., 2009). Upon Spn-E knock-down this sensor still was not fully derepressed: the reporter signal appeared stronger compared to the Armitage knock-down but weaker compared to the AGO3 knock-down (data not shown). Whether *cluster 20A* uses a different piRNA biogenesis module and therefore Armitage is not essential for this cluster, or the *cluster 20A* sensor represents more the behavior of the *roo* element is unclear from these experiments. A new set of sensors containing a more reliable part of *cluster 20A* or parts of the *roo* element (preferable not present in *cluster 20A*) could help to shed more light on the activity of this cluster.

The sensors for *cluster 42AB* could not give many insights into the behavior of this cluster. Only one sensor, containing a target region of *cluster 42AB* in antisense orientation was downregulated to a certain extent. The *cluster 42AB* sensors clearly demonstrate, that a 300 bp region is not sufficient to silence the sensor. It is likely, that the number of piRNAs per kb of target region is the key factor. For the future I would suggest to choose a region between 700-800 bp.

All in all the *nanos* promoter drives the reporter without a target in adult ovaries and testis as well as in larval gonads. This makes this sensor valulable for adding cluster specific targets to monitor their activity. The sensor can be measured either by looking at eGFP, X-gal signal or by performing qPCR on the sensor transcript. Therefore the sensor approach will be a useful tool for investigations on the piRNA pathway in the future.

Putative cluster promoter fragments drive reporter expression in ovaries

Whether or not piRNA clusters have a single promoter, which enables the expression of a long single stranded precursor transcript is still unclear. To tackle this question I cloned sequences upstream of the first piRNAs mapping to *flamenco* and *cluster 20A* into a vector containing the NLS-eGFP-LacZ reporter. The fragment, which is able to drive the reporter in the correct tissues, will give a first hint about a cluster promoter

and can be used for further studies. This approach again allows me to get more insights in the spatiotemporal expression of piRNA clusters.

Analyzing the 2-3.7 kb *flamenco* promoter constructs it was remarkable to see, that construct 1 showed X-gal staining in the germline tissue only. This would point again into the direction that *flamenco* is partly active in germline cells in ovaries. *Flamenco* has been suggested to be soma specific (Brennecke et al., 2007; Malone et al., 2009). In contrast to construct 1, construct 3 is expressed in somatic cells only. This could mean, that construct 1 has an enhancer element, which is sufficient to drive gene expression in the germline, while construct three contains a soma specific enhancer element and a germline repressing element. Constructs 2 and 4 did not show reporter signal in ovaries, therefore I conducted qPCR analysis on the sensor transcript to proove, whether the transcript is made, but not translated properly. Both constructs did not show significant RNA levels, suggesting that there is no sensor transcript made or these are very unstable. Looking at the qPCR levels of construct 1, the expression level is almost that of the empty *nanos* sensor, while construct 3 is expressed more than half the levels of construct 1. All constructs contain a small part of the nearest upstream gene of *flamenco*, which is *DIP1*. This gene is expressed moderately in ovaries and testis (according to flybase.org). Therefore it is more likely that construct 3 contains a germline suppressing element, which prevents the germline enhancer element from acting on *flamenco* transcription.

The X-gal stainings for the *cluster 20A* promoter constructs are more traceable. Only construct 2 showed reporter signal. The ovaries contained strong signal coming from the nurse cell nuclei and detectable, albeit weaker reporter signal in the somatic cells. By looking at the eGFP signal in construct 2 one could see, whether this fragment also shows patchy reporter expression. The qPCR analysis from ovaries of the four *cluster 20A* constructs yielded the following results: construct 1 is not expressed to detectable levels, which correlates with the X-gal staining. Construct 2 is expressed close to the levels of the empty *nanos* sensor, again in agreement with the X-gal staining. Construct 3 and 4 are both expressed only to low levels revealed by qPCR. As both construct show no reporter signal in ovaries, this would hint that the transcript is made, but not translated. There could either be a premature stop

codon, which leads to degradation, or certain structural sequences prevent translation.

Nevertheless the reporter approach combined with the qPCR analysis provided first insights into the architecture of the putative *flamenco* and the putative *cluster 20A* promoter. For both clusters a sequence was found, which is potent of driving the expression of an NLS-eGFP-LacZ reporter. The reporter is able to demonstrate the activity of the corresponding cluster in a manner, which was previously described: *cluster 20A* is active in soma and germline and *flamenco* is active in the somatic tissue. The germline signal coming from *flamenco* construct 1 should not be underestimated and needs further investigations to clarify a putative expression of *flamenco* in the germline. I suggest, making constructs, which do not contain parts of the *DIP1* control regions. To this end, longer fragments or the addition of an IRES sequence could also help to test, if the reporter gets translated again. In addition, one could take the constructs, which did not show reporter expression and create new constructs by adding or removing one nucleotide. This would give hints, whether a frameshift is the reason for constructs, which do not show reporter expression. Performing 5'RACE, would predict the precise putative transcription start site of each cluster. In addition the upstream and downstream sequences of *cluster 42AB* would be of interest for this analysis, too.

The final outcome of this promoter approach would lead to a sequence, which is able to drive the expression of any sequence/reporter in a cluster specific manner. It would be highly interesting to then test the possibility, whether these transcripts would also enter piRNA biogenesis. This would mean that one could easily produce a source for unique piRNA, which could be used for further investigations on the piRNA pathway, or even as tool for silencing genes.

The *flamenco* transcript localizes close to or within Yb-bodies

Where within ovarian somatic cells (OSCs) the long single stranded piRNA cluster transcript localizes is of great interest. As we understand almost nothing about piRNA biogenesis the information on where the cluster transcript is transported from the nucleus could further help to place known pathway proteins into the right order.

In this thesis I applied highly sensitive methods to detect the *flamenco* cluster transcript. Using RNA probes labeled with fluorophores or a hapten, allowed me to get a glimpse how specific the probes are. The three RNA probes were selected to be several kb apart. Applying two probes in one experiment, the output showed two foci coming from two different channels. One of the probes was selected to be the putative beginning of the *flamenco* locus. Localizing the beginning of the transcript to the site of known putative processing bodies was one hypothesis to test. Indeed, the *flamenco* transcript was detecable close to or in Yb-bodies (stained via an Armitage antibody). Armitage is a RNA helicase, essential for the pathway and has been shown to colocalize with Fs(1)Yb to so called perinuclear Yb-bodies (Szakmary et al., 2009; Olivieri et al., 2010; Saito et al., 2010). To further increase the sensitivity and specificity, I used Christmas tree probes (www.panomics.com). This method, based on the bDNA hybridization technology, includes several layers of amplification via nucleic acid hybridization and does not use any antibody to amplify the signal. The Christmas trees turned out to have an extremely good signal to noise ratio, as there was almost no background and a very bright signal coming from foci within the nucleus or cytoplasm. In almost all experiments I could observe two very bright foci within the nucleus. These could be the sites of transcription, as sites of nascent transcription would contain multiple target transcripts for the probes. The Armitage staining, which was always done in parallel to the Christmas tree experiments showed a signal, which was in most of the cases close to one of the nuclear *flamenco* dots. In the majority of the cases a *flamenco* signal was also detectable within the Yb-body. These results would suggest, that the Yb-body and the site of *flamenco* transcription are connected somehow. There could be specialized nuclear pore proteins, which recruit the *flamenco* locus to the nuclear periphery and to this end an auxiliary molecule outside of the nucleus. The auxiliary factor could send out signals for the assembly of an Yb-body close to the site of *flamenco* transcription. This could be a mechanism how the cell selects piRNA precursor transcripts. Whether this holds true has to be tested (discussed in outlook section).

Depletion of known piRNA pathway members resulted in the following observations. After siRNA knock-down of Zucchini, the Armitage staining was significantly enlarged. This has already been described before as accumulations of Piwi protein,

which is not able to enter the nucleus (Saito et al., 2010; Olivieri et al., 2010). Zucchini in addition has been shown to localize to mitochondria and serves as a phospholipase D for lipid signaling (Saito et al., 2010; Olivieri et al., 2010; Watanabe et al., 2011; Huang et al., 2011). Upon Zucchini knock-down the *flamenco* foci within the nucleus were not as bright as in control cells, whereas several *flamenco* dots within the Armitage cloud were detectable. This could point into the direction that Zucchini is at the interface between the site of transcription and the site of processing. To this end, depletion of Zucchini could result in reduction of *flamenco* transcription, which would explain the loss of signal intensity of Christmas trees foci within the nuclei. Reduced piRNAs coming from *flamenco* have already been reported in Zucchini mutants (Malone et al., 2009).

The conventional *in situ* hybridization methods resembled a similar effect upon Zucchini knock-down. The transcript was not observed in a discrete spot, but rather dispersed within the enlarged Yb-body.

Analyzing the Fs(1)Yb knock-down with both methods, the *flamenco* transcript was still detectable but dispersed throughout the cytoplasm. At the moment it is hard to interpret this observation. The Yb-body consists of at least two proteins: Fs(1)Yb and Armitage. The dispersed localization of the *flamenco* transcript displays the importance for the Yb-body for putative piRNA biogenesis close to the site of transcription. Therefore it will be necessary to follow up these results and find factors, which might be important for transcript localization to the Yb-body.

To sum up: the conventional *in situ* approach as well as the Christmas trees approach are able to detect the *flamenco* transcripts within OSCs. Using the bDNA method localization within the nucleus resulted in a bright signal. Upon knock-down of known pathway members, both detection methods showed comparable results. As the Christmas trees showed less background and a stronger signal, I would suggest using this approach rather than the conventional *in situ* approach.

Outlook

This study included the generation of new tools and assays to study piRNA cluster biology in *Drosophila melanogaster*. These provide a starting point for asking more detailed questions on the piRNA pathway.

The sensor approach showed, that the preferable target size is around 800-900 bp to get sufficient downregulation of the reporter signal. In addition to this a sensor, which is transcribed in both (soma and germline) tissues would be the sensor of choice, as one can easily compare both tissues side by side using such a sensor.

piRNA clusters seem to have a single promoter, which drives expression of a long transcript. The exact region of the promoter has to be determined. The method of choice is 5'RACE to get a clear-cut answer.

The Christmas tree approach turned out to be an optimal assay for the OSC system to detect piRNA cluster transcripts. Whether or not this approach can also be applied in ovaries, has to be tested.

To this end the Christmas trees showed that the site of *flamenco* transcription most often has a cytoplasmic Yb-body close to it (see model in Figure x). Whether this subcellular compartmentalization between site of transcription and site of processing is real and biological relevant needs to be tested by establishing methods to untangle the two processes.

One way could be again the bDNA approach. By depletion of putative pathway members (e. g. found in genetic screens) a misslocalization of either the nuclear *flamenco* dot or the Yb-body would give a hint.

The smFISH (stellaris) has to be established for OSCs and ovaries. This approach theoretically should have even less background as the Christmas trees (only one hybridization step) and does not use any antibody for signal amplification. To this end the smFISH protocol is more economic compared to the Christmas trees.

In addition, one could use an *in vivo* tethering system for localization of genomic regions. Two systems are currently used: the LacI/LacO and the tetO/tetR systems (Belmont, 2001). The lacI/lacO system was already established in *Drosophila*. This system has been used for investigations on the meiotic pairing of homologues in *Drosophila* testis. Therefore, I wanted to see, whether this would also work in

Drosophila ovaries. Figure x shows the LacI/LacO system in testis and ovaries. The bright GFP spots represent a 256xLacO array, integrated randomly into the fly genome. The ovaries show bright foci in the germarium, but only some foci in later stages. Whether this is a problem of the HSP83 promoter used for driving the LacI-GFP fusion has to be tested. In addition, the combination of these tethering methods with the newly established MiMIC lines (Venken et al., 2011), would allow to insert the LacO/tetO array within or close to a piRNA cluster using recombinase-mediated cassette exchange (RMCE).

Using all these methods the exciting biology of piRNA clusters will become clearer. And to this end all the known factors involved in the piRNA pathway will find a position within the cascade from cluster transcription to silencing TEs.

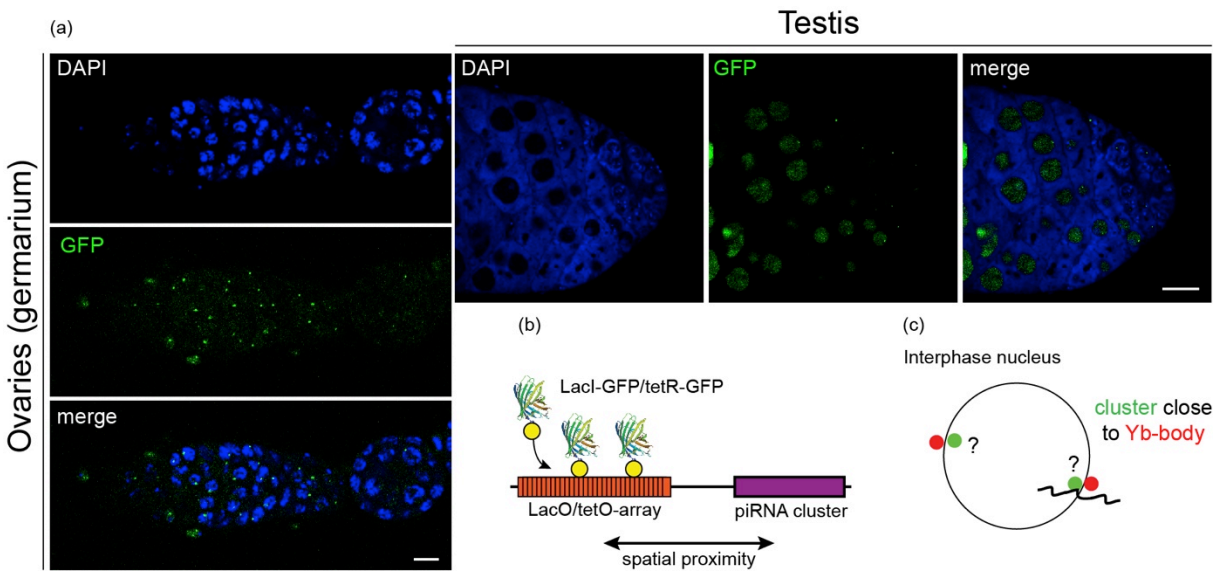


Figure 23 The LacI/LacO system and the model, which has arised from this study. (a) showing ovaries and testis using the LacI/LacO system. (b) A cartoon showing how the tethering systems would work after integration close to a cluster. (c) A model representing the site of cluster transcription in green and the processing body next to it in red. (white bar = 10 µm)

Material and Methods

Standard laboratory reagents:

Promega Wizard SV Gel and PCR purification Kit

Used according to manufacturers protocol for clean up of PCRs, restriction digests or purification of DNA from agarose gel

QIAGEN Plasmid Mini Kit

Used according to manufacturers protocol for general cloning and small injection constructs

QIAGEN Plasmid Midi Kit

Used according to manufacturers protocol for purification of larger injection constructs

QIAGEN Plasmid Maxi Kit

Used according to manufacturers protocol for purification of low copy number plasmids/PACs/BACs

QIAGEN Large-Construct Kit

Used according to manufacturers protocol for purification of high quality supercoiled DNA for injections

RNeasy Mini Kit

Used according to manufacturers protocol for clean up after in vitro transcription reaction and after in vitro labeling of RNA probes

QuantiGene® ViewRNA ISH Cell Assay for Fluorescent RNA *In Situ* Hybridization (www.panomics.com)

DNA concentration for embryo injections: 4 µg DNA in 20 µL aqua dest.

Buffers and Media

([1)] generously prepared by the IMBA media kitchen)

LB-Medium (pH 7.5)[1)]

5 g * L^{-1} NaCl
10 g * L^{-1} Tryptone
5 g * L^{-1} Yeast extract
pH adjusted with NaOH

LB-AMP plates (pH 7.5)[1)]

15 g * L^{-1} Agar
5 g * L^{-1} NaCl
10 g * L^{-1} Tryptone
5 g * L^{-1} Yeast extract
pH adjusted with NaOH

2 x YT (pH 7.5)[1)]

16 g * L^{-1} Tryptone
10 g * L^{-1} Yeast extract
5 g * L^{-1} NaCl
pH adjusted with NaOH

Flyfood[1)]

7.5 g * L Agar
80 g * L^{-1} corn meal
18 g * L^{-1} dried yeast
80 g * L^{-1} Malzym
0.5 mL * L^{-1} O/phosphoric acid
8.4 mL* L^{-1} propionic acid
10 g * L^{-1} soya meal
22 g * L^{-1} sugar beets syrup

Apple juice plates[1)]

17.5 g * L^{-1} Agar
250 mL * L^{-1} Applejuice
10 mL * L^{-1} Nipagin
25 g * L^{-1} sugar

Yeast Paste: dry yeast, ddH2O

5 M NaCl[1)]

292.2 g * L^{-1} NaCl

1x PBS[1)]

137 mM NaCl
2.7 mM KCl
10 mM Na_2HPO_4
2 mM KH_2PO_4

1x TE[1)]

10 mM Tris/HCl pH 7.4
1 mM EDTA pH 8
1 M MgCl2
203.3 g * L^{-1} MgCl2.6H2O

0.5 M EDTA pH8[1)]

186.1 g * L^{-1}

1 M Tris pH 7.5[1)]

21.1 g TRIZMA® base

700 mL aqua dest.
pH adjusted with conc. HCl

10x TAE buffer[1)]
0.4 M Tris
10 mM EDTA
pH adjusted with acetic acid

20x SSC[1)]
3 M NaCl
300 mM sodium citrate

Solution A
0.1 M Tris HCl pH 9
0.1 M EDTA pH 8
1 % (v/v) SDS

Squishing Buffer
10 mM Tris HCl pH 8
1 mM EDTA
25 mM NaCl
+ 200 µg * mL^{-1} of Proteinase K before use

Buffer B
1 mM MgCl2
150 mM NaCl
in PBS

Ferricyanide ($[Fe(CN)_6]^{3-}$ also hexacyanidoferrate(III) or hexacyanoferrate(III))
0.3 M Potassium ferricyanide(III) solution, protect from light

Ferrocyanide ($[Fe(C N)_6]^{4-}$)
0.3 M Potassium hexacyanoferrate(II) trihydrate solution, protect from light

10 % Triton X-100
prepared in PBS, stored at 4 °C

X-Gal solution
10 % 5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid prepared in DMF or DMSO, stored at – 20 °C

Fixation Buffer (FAX)
4 % paraformaldehyde solution (EM grade)
0.1 % Triton X-100
store at – 20°C

PBX
0.1 % Triton X-100
in PBS, store at 4 °C

BBX
0.1 % Triton X-100
0.1 % Bovine serum albumin (BSA)
in PBS, store at 4 °C

in Glycerol, store at – 20 °C, protect from light

Blocking solution

100 mM Tris pH 7.5

150 mM NaCl

0.5 % Blocking reagent (Roche)

Heat up (e.g. microwave) to dissolve completely

Hybridization solution

50 % Formamide

5x SSC

100 µg * mL^{-1} Heparin

100 µg * mL^{-1} Salmon sperm DNA

0.1 % Tween-20

Filter sterilize after preparation

Mounting solution

0.1 M Tris pH 9

0.5 % (w/v) n-propyl gallate

Antibiotics

Substance	Stock concentration	Working concentration	Solvens
Ampicillin	50 mg * mL^{-1}	1:1000	EtOH abs.
Chloramphenicol	30 mg * mL^{-1}	1:2000	EtOH abs.
Kanamycin	30 mg * mL^{-1}	1:2000	Aqua dest.
Streptomycin	50 mg * mL^{-1}	1:500-1:1000	Aqua dest.
Tetracycline	10 mg * mL^{-1}	1:3000	EtOH abs.

Cloning procedures

All cloning procedures (restriction digest, ligation) were performed based on (Sambrook et al., 1989). For analytic PCR in house taq polymerase was used, for standard high fidelity PCR KOD hot start polymerase (Novagen®) was used.

Primer design

All PCR primers were made either by hand or primer3 (http://frodo.wi.mit.edu/primer3/)

general considerations for primers:

T_m = 55-65 °C

%GC = 40-60

length = 20-25 bp

GC-clamp = 1

ΔT_m = 0.1-0.9

product size for qPCR primers (150-300 bp)

Single fly genomic DNA preparation for PCR

take up 50 µL of Squishing buffer (with ProtK) in pipette tip

mash fly with pipette tip, without expelling the buffer

expel remaining buffer

incubate in PCR machine for 30' @ 37 °C

inactivate ProtK by heating up to 95°C for 3'

store at 4°C

Before proceeding with PCR, spin down, only use supernatant

Use 2-3 µL of DNA Prep for PCR

Genomic DNA preparation of *Drosophila melanogaster*

collect 20-30 flies in an Eppendorf (keep on ice, not longer than 1 h)

add 200 µL solutionA

homogenize using a Pellet pestles Z359971 (Simga-Aldrich)

add 200 µL solutionA

incubate for 30' @ 70 °C

add 14 µL of KOAc (8 M) per 100 µL of solutionA

incubate for 30' on ice

spin for 15' at max. rpm

take supernatant to a new tube

extract with 400 µL Phenol/Chlorophorm

spin 2' at max. rpm

take supernatant and repeat extraction

add 200 µL isopropanol

mix by inversion, no vortex

spin 10’ at max. rpm

wash pellet with 70 % EtOH

air dry pellet

redissolve in 100 µL of nuclease free water

A-tailing for TOPO® cloning (40 µL reaction):

x µL PCR product

4 µL 10 x buffer

1 µL dATP (10 mM)

0.5 µL Taq Polymerase (in house)

up to 40 µL with aqua dest.

incubate 10' @ 72 °C

TOPO® cloning for sequencing:

1 µL A-tailed PCR product

1 µL salt solution

0.25 µL pCR™4-TOPO® TA vector

up to 6 µL with aqua dest.

incubate 30' @ RT

Transformation to TOP10 chemically competent cells:

thaw competent bacteria on ice

add 1 µL of Topo reaction to bacteria and incubate for 15-30’ on ice

heat shock at 42 °C for 30"

incubate for 1’ on ice

add 250 µL SOC medium or LB medium (w/o antibiotics)

and plated on LB-AMP plates

Total RNA preparation

dissect ovaries of 5 flies into ice cold PBS (keep on ice for max 30’)

add 200 µL TRIzol®
dounce ovaries
add 800 µL TRIzol®
vortex 1’
incubate for 10’ @ RT
meanwhile spin down peqGOLD PhaseTrap tubes for 30’’ @12,000 g
vortex 1’
transfer solution to peqGOLD PhaseTrap tubes
add 200 µL Chloroform
mix by inverting gently 8-10x (phases should be mixed thoroughly) no vortexing!
spin for 5’ @ 12,000-16,000 g @ RT
pour aqueous phase into new Eppendorf tube

Precipitation of RNA

add 550 µL Ispropanol
invert several times
optional: incubate for 10’ @ RT
invert again
spin full speed for 20’ @ 4 °C
remove supernatant
wash by adding 200 µL 80 % EtOH
spin full speed 5’ @ 4 °C
remove supernatant
centrifuge again 20’’ (helps to remove residual supernatant, as pellet might not stick to wall of the tube any more)
remove residual EtOH
air dry pellet for 5-10’ @ RT
resolve pellet in x µL RNAse free water (depending on the concentration wanted 20-50 µL)
measure concentration with NanoDrop
store at -20 °C

cDNA preparation

DNA digestion

250-500 ng RNA

1 µL 10x DNase I Reaction Buffer

1 µL RNase free DNAse (Invitrogen or Promega)

aqua dest. up to 10 µL

incubate for 15-30' @ RT

add 1 µL 25 mM EDTA and incubate for 10' @ 65 °C for inactivation

Reverse transcription using Invitrogen SuperScript™ II reverse transcriptase

50–250 ng random primer

11 µL of RNA sample (digested with DNAse)

1 µL dNTP Mix 10 mM

incubate for 5' @ 65 °C, quick chill on ice

4 µL First-Strand Buffer 5x

2 µL DTT 0.1 M

1 µL RiboLock™ (40 U * μL^{-1}, Fermentas)

incubate for 2' @ 25 °C

add 0.5 - 1 µL SuperScript™ II RT and mix by pipetting gently up and down

incubate tube at 25 °C for 10'

incubate at 42 °C for 1 hour

inactivate by incubating for 15' @ 70 °C

qPCR

0.5 µL of cDNA

0.7 µL Primer 1

0.7 µL Primer 2

10 µL 2x iQ SYBR Green Supermix (Bio-Rad)

up to 20 µL with aqua dest.

Program:

95 °C	10'	
95 °C	15''	40x
55 °C	30''	
72 °C	30''	

Melting curve: Temp. range (65-95 °C); Increment: 0.5 °C; Time per step: 5''

qPCR results were analyzed using the method described in (Schmittgen and Livak, 2008).

X-Gal staining of ovaries

dissect ovaries in PBS (keep on ice)
fix with 0.5% Glutaraldehyde in PBS (10-15' @ RT)
rinse twice with PBS
incubate in staining solution ((37 °C faster than RT, time has to be determined)
wash twice with PBS
store in PBS at 4 °C
or
mount with 80% Glycerol in PBS on glass slides

Probe generation for *in situ* hybridization

DNA template was generated using primers, which amplify the region of interest, containing a T7 minimal binding site TAATACGACTCACTATAGG (www.ambion.com), gel and column purified.

Generation of RNA probes with aminoallyl-UTP, after (Glotzer et al., 1997)
5-10 µg DNA template
10 µL $MgCl_2$ 140 mM
10 µL Buffer for T7 polymerase 10x
25 mM ATP, GTP, CTP
25 mM aminoallyl-UTP (Biozym) : UTP ratio of 1:1-1:5
1 µL RiboLock™ (total 40 Units)
2 µL T7 polymerase (total 40 Units)
up to 100 µL with aqua dest.
mix by pipetting up and down (important)
incubate for 1-2 h @ 37 °C
add another 20 Units of T7 polymerase (optional)
incubate again for 1-2 h @ 37 °C (solution gets turbid after a while)

remove DNA by adding DNAse (see cDNA preparation for protocol)

Clean up of reaction

Phenol/Chloroform/Isoamylalcohol extraction

add 100 µL of P/Chl/I

vortex 10''

spin for 2' @ max. rpm

take aqueous upper phase and repeat extraction

take again upper phase and extract once with 100 µL Chloroform

take upper aqueous phase and clean RNA using RNeasy kit (follow suppliers protocol for RNA clean up)

Check probes on 1 % agarose gel

Conjugating the aminoallyl-RNA probes with Digoxygenin/Alexa488 (Molecular Probes)

Concentrate RNA probes by using SpeedVac or alcohol precipitation to small volume (5-10 µL)

add 14 µL of either Digoxigenin (3-amino-3-deoxygigoxigenin hemisuccinamide, SE) or Alexa Fluor® 488 carboxylic acid, succinimidyl ester *mixed isomers*

add 75 µL 0.1 M sodium tetraborate pH 8.5 (prepare fresh, set pH correctly with 10 M NaOH)

up to 100 µL if needed with H_2O

Incubate 24 h, shaking @ RT

Clean up reaction using RNeasy kit (QUIAGEN)

store probes at -20 °C

in vitro transcription for ISH probe generation

(using ROCHE DIG labeling kit)
according to manufacturers protocol and
(http://pmm.umassmed.edu/theurkauf/Protocols.html)

1-5 µg of DNA template

3 µL of Label mix

1 µL RiboLockTM

2 µL T7 polymerase

up to 30 µL

incubate for 2-4 h @ 37 °C

+3 µL DNAse buffer 10x

+ 2 µL DNAse

15’ @ 37 °C

purify over RNeasy column

Dot blot

to check how efficient and whether probes (Digoxygenin) are labeled, after (Weiszmann et al., 2009)

Make dilution series of choice

Using a multi-channel pipette, drop 1 µL of dilution series on a Hybond™-N+ (Amersham)

UV-cross link using AUTO function

block membrane in PBTM (0.1 % Tween-20, 1 % milk powder) for 30’ @ RT

incubate with α-dig-FAB antibody conjugated with alkaline phosphatase (1:5000) in PBTM, for 30’ @ RT

wash 4x 10’ with PBT (PBS, 0.1 % Tween-20)

wash 2x 5’ with PBS

incubate with BM-purple substrate (Roche), wait until signal has developed satisfyingly (30-60’, reaction is faster @ 37 °C, protect from light)

wash once with PBS

The probes labeled with a fluorophore (e. g. Alexa488) can be checked on a NanoDrop for UV absorption at according wavelength.

Culturing OSCs

performed after ‘Culturing Ovarian Somatic and Germline Stem Cells of *Drosophila* (Yuzo Niki)’

http://www.currentprotocols.com/protocol/sc02e01

In situ hybridization of OSCs

Protocol for small coverslips in 24-well plates (250 µL volume needed)

The signal was significantly enhanced, if the initial hybridization temperature was set to 85 °C for 5 minutes and then set to 40 °C, so that the temperature goes down with time. In general lower hyb-temperatures seem to be favorable (37 – 40 °C).

Proteinase treatment was performed using Proteinase QS, supplied in the Quantigene kit. This proteinase can be acquired separately from Affymetrix. As an alternative one could titrate out the concentration of Protenase K need for *in situ* experiments followed by antibody staining.

Preparation of coverslips

wash in 70 % EtOH

dip into Concanavalin A

air dry slides and store at 4 °C in a petri dish

Conventional *in situ* hybridization

add cells directly on a coverslip and incubate for 3 h at 27 °C

wash once with PBS

fix with 4% paraformaldehyde (EM grade) for 30' @ RT

wash 2x with PBS

incubate with PBT (PBS, 0.01 % Tween-20) for 10' @ RT

wash 2x with PBS

incubate 5' with Proteinase QS (1:4000 in PBS) @ RT

wash 2x with PBS

incubate cells with probe (300 ng of probe) in hybridization buffer for 5' @ 85 °C, set the temperature to 37-40 °C afterwards and incubate o/n.

add 200 µL PBT to the well, incubate for 5' (can be prolonged up to 30')

wash 2x for 5' with PBT (can be prolonged up to 30' each)

add primary ABs in blocking buffer

incubate o/n @ 4 °C

wash 2x 10' with PBT

incubate with secondary ABs in blocking buffer (3-5 h @ RT or o/n @ 4 °C)

wash 2x 10’ with PBT

incubate with DAPI (1:2000-5000) in PBS for 5-10’

wash once with PBS

mount coverslips on slides

Christmas tree *in situ* hybridization (using the Quantigene assay)

Used according to the manufacturers protocol (http://panomics.com/downloads/UM18801_QGViewRNA_ISH_CellAssay_RevA_110525.pdf) except for two changes: Proteinase QS (1:4000 in PBS) digest for 5 minutes. The hybridization temperature was performed 85 °C for 5‘ then the temperature was set to 40 °C. Slides were then incubated for 3 h for the first step. In addition, the washing steps were performed for 5-10 minutes each.

smFISH protocol (stellaris)

According to the manufacturers protocol

http://www.biosearchtech.com/assets/bti_custom_stellaris_adherent_cell_protocol.pdf

and

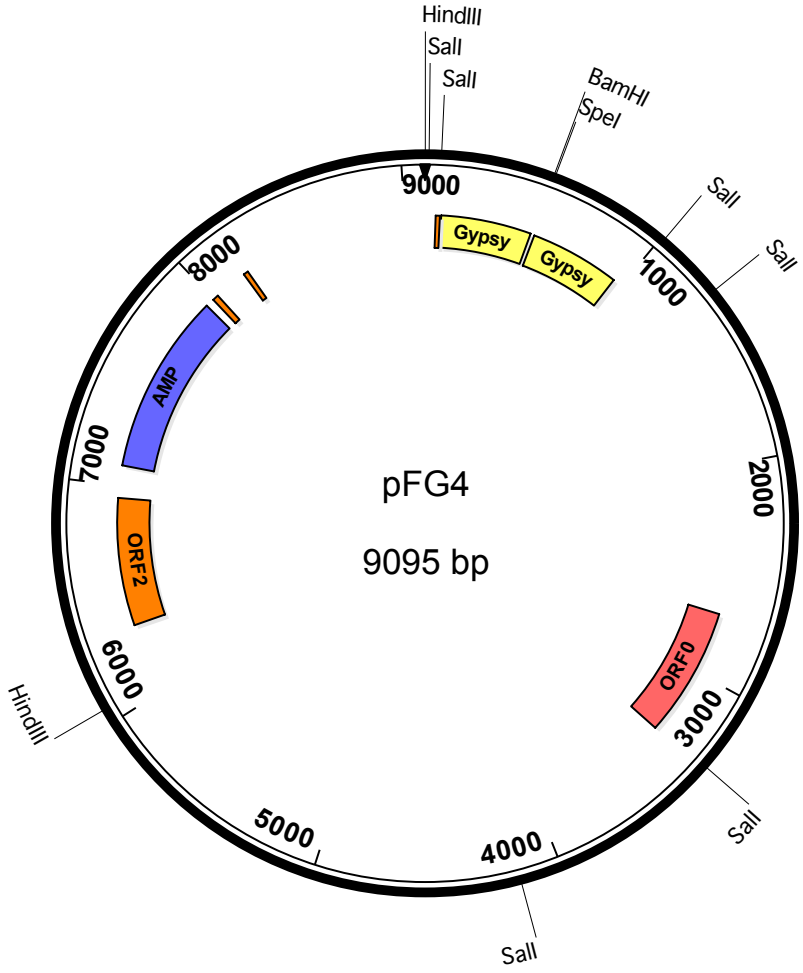
http://www.biosearchtech.com/assets/bti_custom_stellaris_general_protocol.pdf

Primer sequences and vector maps

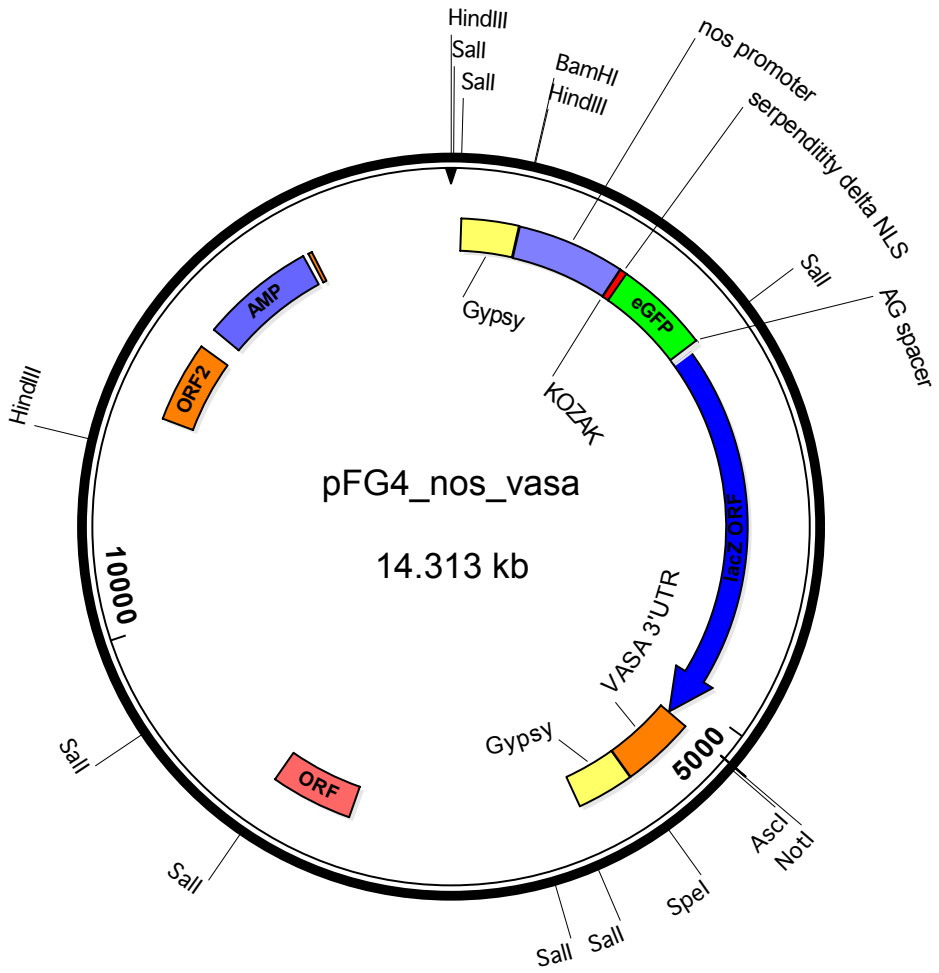
Fla-r1-Not-fw	ATTAgcggccgcCAGATTACCATTTGGCTATGAG
Fla-r1-Asc-rev	attaGGCGCGCCTGATACCGATTAGATTAGGTTGTC
Fla-r2-Not-fw	ATTAgcggccgcAAGGAAAACGTGGGAAAGTTG
Fla-r2-Asc-rev	attaGGCGCGCCCCCCTCTACATACACAGAAAAAGG
42-R1-Not-fw	ATTAgcggccgcAACATCCTCGTCGATTTCCTC
42-R1-Asc-rev	attaGGCGCGCCCTGCCACCGTCCTATATTCG
42-R2-Not-fw	ATTAgcggccgcATATAGGACGGTGGCAGATCC
42-R2-Asc-rev	attaGGCGCGCCCTTCGGAAACTTTGTGAAAAGG
42-R1as-Not-fw	attaGGCGCGCCAACATCCTCGTCGATTTCCTC
42-R1as-Asc-rev	ATTAgcggccgcCTGCCACCGTCCTATATTCG
42-R2as-Not-fw	attaGGCGCGCCATATAGGACGGTGGCAGATCC
42-R2as-Asc-rev	ATTAgcggccgcCTTCGGAAACTTTGTGAAAAGG
Fla-r1-Not-fw	ATTAgcggccgcCAGATTACCATTTGGCTATGAG
Fla-r1-Asc-rev	attaGGCGCGCCTGATACCGATTAGATTAGGTTGTC
Fla-r2-Not-fw	ATTAgcggccgcAAGGAAAACGTGGGAAAGTTG

Fla-r2-Asc-rev	attaGGCGCGCCCCCCTCTACATACACAGAAAAAGG
42-R1-Not-fw	ATTAgcggccgcAACATCCTCGTCGATTTCCTC
42-R1-Asc-rev	attaGGCGCGCCCTGCCACCGTCCTATATTCG
42-R2-Not-fw	ATTAgcggccgcATATAGGACGGTGGCAGATCC
42-R2-Asc-rev	attaGGCGCGCCCTTCGGAAACTTTGTGAAAAGG
42-R1as-Not-fw	attaGGCGCGCCAACATCCTCGTCGATTTCCTC
42-R1as-Asc-rev	ATTAgcggccgcCTGCCACCGTCCTATATTCG
42-R2as-Not-fw	attaGGCGCGCCATATAGGACGGTGGCAGATCC
flam1_ish_for	GTGACTCGAAATGGTTGGTTTG
flam1_ish_rev	TAATACGACTCACTATAGGGGTACGACCATCCAAA CAGATG
flam2_ish_fw	TCCTTAACAAACATTCGGTTCG
flam2_ish_rev	TAATACGACTCACTATAGGATATTTAGGACCGGCC AACTCC
nos-pro-fw	attaGGATCCaagcttcgaccgttttaacc
nos-pro-rev	cagctcctcgcccttgctcaccatGCACTCCTGCTTGACGCG CTTCTTCGTCGGCCTGTCCTCGCCATCGGAGGGC ATTTTGCATGTAACTAAACTCGC
nos-seq-fw	gtcacagtgcgcgaaattcg
Vasa-GFP-rev	attaAAGCTTgccggccttgtacagctcgtccatgccgag
vasa-lacz-fw	attaAAGCTTgccggcATGGCTCGCGATGATCCCGTC G
vasa-lacz-rev	attaGCGGCCGCTTATTTTTGACACCAGACCAACTG G
piwi-GFP-rev	attaGTCGACgccggccttgtacagctcgtccatgccgag
piwi-lacz-fw	attaGTCGACgccggcATGGCTCGCGATGATCCCGTC G
LacZ_seq_fw	GATCCCGTCGTTTTACAACG
osk-ISH-fw	AAAAATGCCAGTACCCATCAAC
osk-ISH-rev	TAATACGACTCACTATAGGTGACCTTTAGGTGACA GCATTC
vasa-3utr-fw	ttaGCGGCCGCaatattaattGGCGCGCCAATGTATGGA CATAGATTTC
vasa-3utr-rev	AATactagtTTCTGCAAAGACAGCAAACG
vasa-3utr-seq	CACTTGGCTCTTGCGCTCTC
pCasper_2gypsy_fw	CGTTAACGTTCGAGGTCGAC
pCasper_2gypsy_fw2	CCAACAACTCTAGAGGATCC
Fla1-qPCR-fw	CCATTTGGCTATGAGGATCAGAC
Fla1-qPCR-rev	GCCCACAGACAAGCTACACAAA
Fla2-qPCR-fw	GGAGAGGAAATTTTTCGATTGC
Fla2-qPCR-rev	CGACTTTCAGATACCCGTTACTCA
Fla3-qPCR-fw	GAAGTCTTGGGACACTCATAGGT
Fla3-qPCR-rev	CCAGAAAATTAAGCGGAAGC
flam-pro-fw	ATTAtctagaACACTCCTGCACACACTTGC
flam-pro-rev1	ATTAggatccTTCCTTTTTCTTGCGTCCATAC
flam-pro-rev2	ATTAggatccAACTTCAATCGAACCACATCG
flam-pro-rev3	ATTAggatccTGAAATGTAAAACGCCACACA
flam-pro-rev4	ATTAggatccATTGAACCTTACCCCGACAAT
Burdock-Not-fw	ATTAgcggccgccaaattcgctattgtgcaacc
Burdock-Asc-rev	attaGGCGCGCCtttcaaatcattcgggtctagc
tj-not-fw	ATTAggcgcgccGTTTGTTAAAGCGTTTCCAAGG
tj-asc-rev	ATTAgcggccgcGGAAACCGCCTAATTTTTCAG
fla-seq-1	GCACGACGATGCTCTTTGG
fla-seq-2	GTTACTTAGATAGTGGGAG

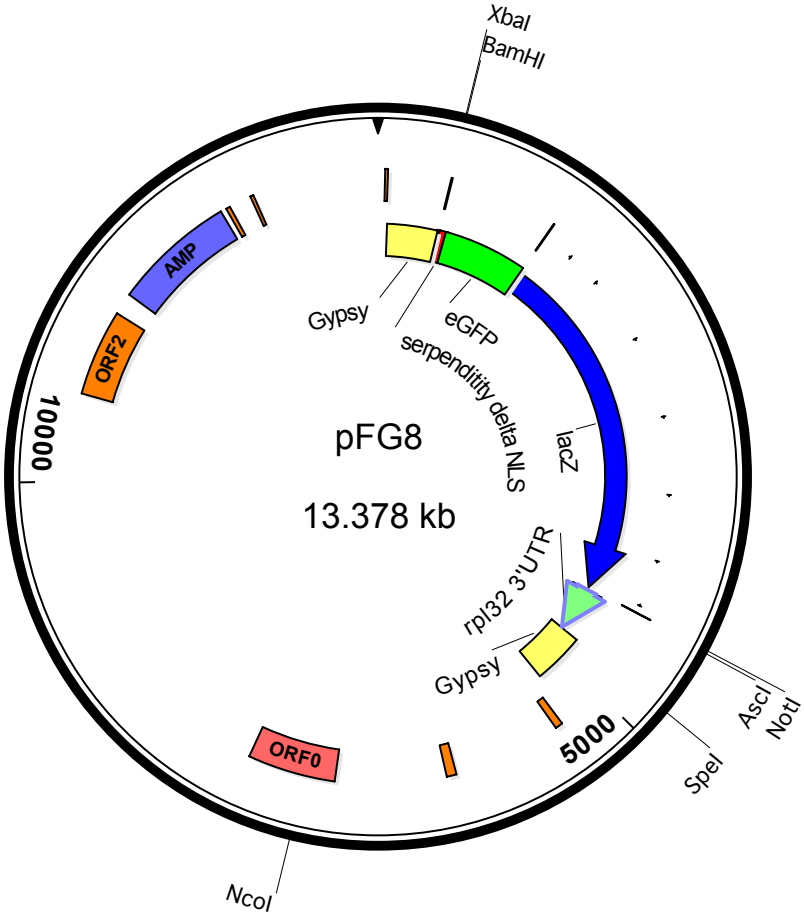
fla-seq-3	CAACCCACTTGCAGAAAATC
fla-seq-4	CATGTGCCAGTTTGAATTTG
fla-seq-5	CAGCTCAGCAGCAGTGTAG
fla-seq-6	AGAATAGGCAGGCTATAAC
fla-seq-7	TGTTAGCAGTTTTACTTGG
gfp-nls-seq-rev	ctgaacttgtggccgtttacg
Fla_ISH_2_fw	ATATTTAGGACCGGCCAACTCC
Fla_ISH_2_rev	TAATACGACTCACTATAGGTCCTTAACAAACATTC GGTTCG
fla1_short_as_fw	attaGGCGCGCCCAGATTACCATTTGGCTATGAG
fla1_short_as_rev	ATTAgcggccgcTGATACCGATTAGATTAGGTTGTC
fla2_short_as_fw	attaGGCGCGCCAAGGAAAACGTGGGAAAGTTG
fla2_short_as_rev	ATTAgcggccgcCCCCTCTACATACACAGAAAAAGG
ePFG-fw	tgtacagccctgaaaaagggctcgagacgtaaacggccacaagttcag
ePFG-BamHI-rev	attaGGATCCgggtgctcaggtagtggttgtc
BoxB-Fw	tgtacagccctgaaaaagggctc
BoxB-EcoRI-rev	attaGAATTCccgcggtaccgccctttttc
Kan-EcoRI-fw	attaGAATTCGCAAGCGAACCGGAATTGCCAGC
Kan-NotI-rev	attaGCGGCCGCTCAGAAGAACTCGTCAAGAAGG
Kan-seq1	GCAAAGTAAACTGGATGGCTTTC
Kan-seq2	CAGGATCTCCTGTCATCTC
Kan-seq3	CCGCTTTTCTGGATTCATCG
flamenco_reg2_seq_fw	CAAATCTCGATCCCGTATAACC
flamenco_reg2_seq_rev	CAACCAATTTCTAATCGGAAGC
flamenco_reg2_seq_rev2	AATATTTCCACGTCTCCGCTAC
sensor-gfp-qpcr-fw	taccccgaccacatgaagcag
sensor-gfp-qpcr-rev	cttgaagtcgatgcccttcagc
Sensor-LacZ-qPCR-rev	AATGTTGATGAAAGCTGGCTAC
Sensor-LacZ-qPCR-fw	GCTCAGGTCAAATTCAGACG



pFG4 is a derivative of pCa4B2G (Markstein et al., 2008), a NotI site has been filled up.



pFG4 with *nanos* promoter and vasa 3' UTR, containing nls-eGFP-LacZ reporter. This vector was the basis for all the cluster sensors containing a target sequence, which was cloned into NotI/AscI sites.



Another derivative of the pFG4 vector is pFG8, which was used for cloning in the promoter constructs using XbaI (compatible with NheI or SpeI) and BamHI (compatible with BglII). This vector has 400 bp 3'UTR of the rpl32 gene together with the NLS-eGFP-LacZ reporter.

References

Ali, I., ur Rehman, M., Rashid, F., Khan, S., Iqbal, A., Laixin, X., ud din Ahmed, N., and Swati, A. Z. (2010). Cis-regulatory elements affecting the Nanos gene promoter in the germline stem cells. J. Biotechnol. *145*, 323–329.

Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., and Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr. Biol. *11*, 1017–1027.

Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev. Cell *5*, 337–350.

Arbouzova, N. I., and Zeidler, M. P. (2006). JAK/STAT signalling in Drosophila: insights into conserved regulatory and cellular functions. Development *133*, 2605–2616.

Bate, M., and Arias, A. M. (1993). The Development of Drosophila melanogaster M. Bate and A. M. Arias, eds. (Cold Spring Harbor Laboratory Press).

Bellen, H. J., Levis, R. W., He, Y., Carlson, J. W., Evans-Holm, M., Bae, E., Kim, J., Metaxakis, A., Savakis, C., Schulze, K. L., et al. (2011). The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics *188*, 731–743.

Belmont, A. S. (2001). Visualizing chromosome dynamics with GFP. Trends Cell Biol. *11*, 250–257.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Bregliano, J. C., Picard, G., Bucheton, A., Pélisson, A., Lavige, J. M., and L'Heritier, P. (1980). Hybrid dysgenesis in Drosophila melanogaster. Science *207*, 606–611.

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

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

Castro, J. P., and Carareto, C. M. A. (2004). Drosophila melanogaster P transposable elements: mechanisms of transposition and regulation. Genetica *121*, 107–118.

Chen, D., and McKearin, D. M. (2003). A discrete transcriptional silencer in the bam gene determines asymmetric division of the Drosophila germline stem cell. Development *130*, 1159–1170.

Cox, D. N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development *127*, 503–514.

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

Eddy, E. M. (1974). Fine structural observations on the form and distribution of nuage in germ cells of the rat. Anat. Rec. *178*, 731–757.

Ghildiyal, M., and Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. *10*, 94–108.

Glotzer, J. B., Saffrich, R., Glotzer, M., and Ephrussi, A. (1997). Cytoplasmic flows localize injected oskar RNA in Drosophila oocytes. Curr. Biol. *7*, 326–337.

Goodier, J. L., and Kazazian, H. H. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23–35.

Griffiths, A. J. F., Wessler, S. R., Carroll, S. B., and Doebley, J. (2010). Introduction to Genetic Analysis Tenth Edition. (W. H. Freeman).

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

Hamilton, A. J., and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950–952.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011a). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J.

Huang, H., Gao, Q., Peng, X., Choi, S.-Y., Sarma, K., Ren, H., Morris, A. J., and Frohman, M. A. (2011). piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. Dev. Cell *20*, 376–387.

Hurst, G. D., and Werren, J. H. (2001). The role of selfish genetic elements in eukaryotic evolution. Nat. Rev. Genet. *2*, 597–606.

Itzkovitz, S., and van Oudenaarden, A. (2011). Validating transcripts with probes and imaging technology. Nat. Methods *8*, S12–9.

Kazazian, H. H. (2011). Mobile DNA: Finding Treasure in Junk (FT Press Science) 1st ed. (FT Press).

Kazazian, H. H. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626–1632.

Khurana, J. S., and Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. J. Cell Biol. *191*, 905–913.

Kidwell, M. G., Kidwell, J. F., and Nei, M. (1973). A case of high rate of spontaneous mutation affecting viability in Drosophila melanogaster. Genetics *75*, 133–153.

Kidwell, M. G., Kidwell, J. F., and Sved, J. A. (1977). Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. Genetics *86*, 813–833.

King, R. C. (1970). Ovarian Development in Drosophila Melanogaster (Academic Press).

Kirino, Y., Kim, N., de Planell-Saguer, M., Khandros, E., Chiorean, S., Klein, P. S., Rigoutsos, I., Jongens, T. A., and Mourelatos, Z. (2009). Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat. Cell Biol. *11*, 652–658.

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

Lau, N. C., Robine, N., Martin, R., Chung, W.-J., Niki, Y., Berezikov, E., and Lai, E. C. (2009). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. Genome Res. *19*, 1776–1785.

Leblanc, P., Desset, S., Giorgi, F., Taddei, A. R., Fausto, A. M., Mazzini, M., Dastugue, B., and Vaury, C. (2000). Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. J. Virol. *74*, 10658–10669.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Levin, H. L., and Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. Nat. Rev. Genet. *12*, 615–627.

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

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat. Genet. *40*, 476–483.

McCLINTOCK, B. (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U.S.A. *36*, 344–355.

Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A. R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615–1624.

Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat. Methods *8*, 405–407.

Niki, Y., Yamaguchi, T., and Mahowald, A. P. (2006). Establishment of stable cell lines of Drosophila germ-line stem cells. Proc. Natl. Acad. Sci. U.S.A. *103*, 16325–16330.

Nishida, K. M., Okada, T. N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., Huang, H., Chen, D., Kodama, T., Siomi, H., et al. (2009). Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. EMBO J *28*, 3820–3831.

Noselli, S., and Vincent, A. (1991). A Drosophila nuclear localisation signal included in an 18 amino acid fragment from the serendipity delta zinc finger protein. FEBS Lett. *280*, 167–170.

Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J *29*, 3301–3317.

Pane, A., Wehr, K., and Schüpbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev. Cell *12*, 851–862.

Pardue, M. L., and Debaryshe, P. G. (2000). Drosophila telomere transposons: genetically active elements in heterochromatin. Genetica *109*, 45–52.

Pélisson, A., Song, S. U., Prud'homme, N., Smith, P. A., Bucheton, A., and Corces, V. G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J *13*, 4401–4411.

Picard, G. (1976). Non-mendelian female sterility in Drosophila melanogaster: hereditary transmission of I factor. Genetics *83*, 107–123.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697–711.

Prudhomme, S., Bonnaud, B., and Mallet, F. (2005). Endogenous retroviruses and animal reproduction. Cytogenet. Genome Res. *110*, 353–364.

Ravi Ram, K., and Wolfner, M. F. (2007). Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. Integr. Comp. Biol. *47*, 427–445.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Jackson, R. B. (2010). Campbell Biology (9th Edition) 9th ed. (Benjamin Cummings).

Robert, V., Prud'homme, N., Kim, A., Bucheton, A., and Pélisson, A. (2001). Characterization of the flamenco region of the Drosophila melanogaster genome. Genetics *158*, 701–713.

Robine, N., Lau, N. C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., Blower, M. D., and Lai, E. C. (2009). A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr. Biol. *19*, 2066–2076.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science *218*, 348–353.

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M. C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature *461*, 1296–1299.

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H., and Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. *24*, 2493–2498.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (3 Volume Set) 2nd ed. (Cold Spring Harbor Laboratory Pr).

Sano, H., Nakamura, A., and Kobayashi, S. (2002). Identification of a transcriptional regulatory region for germline-specific expression of vasa gene in Drosophila melanogaster. Mech. Dev. *112*, 129–139.

Sarot, E., Payen-Groschêne, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics *166*, 1313–1321.

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc *3*, 1101–1108.

Senti, K.-A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet. *26*, 499–509.

Shiga, Y., Tanaka-Matakatsu, M., and Hayashi, S. (1996). A nuclear GFP/*β*-galactosidase fusion protein as a marker for morphogenesis in living *Drosophila*. Develop. Growth Differ. *38*, 99–106.

Shpiz, S., Kwon, D., Uneva, A., Kim, M., Klenov, M., Rozovsky, Y., Georgiev, P., Savitsky, M., and Kalmykova, A. (2007). Characterization of Drosophila telomeric retroelement TAHRE: transcription, transpositions, and RNAi-based regulation of expression. Mol. Biol. Evol. *24*, 2535–2545.

Siomi, H., and Siomi, M. C. (2009). On the road to reading the RNA-interference code. Nature *457*, 396–404.

Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246–258.

Szakmary, A., Reedy, M., Qi, H., and Lin, H. (2009). The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. J. Cell Biol. *185*, 613–627.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320–324.

Vagin, V. V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G. J., and Aravin, A. A. (2009). Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes Dev. *23*, 1749–1762.

Van Doren, M., Williamson, A. L., and Lehmann, R. (1998). Regulation of zygotic gene expression in Drosophila primordial germ cells. Curr. Biol. *8*, 243–246.

Venken, K. J. T., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A., et al. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat. Methods *8*, 737–743.

Watanabe, T., Chuma, S., Yamamoto, Y., Kuramochi-Miyagawa, S., Totoki, Y., Toyoda, A., Hoki, Y., Fujiyama, A., Shibata, T., Sado, T., et al. (2011). MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. Dev. Cell *20*, 364–375.

Weiszmann, R., Hammonds, A. S., and Celniker, S. E. (2009). Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. Nat Protoc *4*, 605–618.

Wolpert, L., and Tickle, C. (2010). Principles of Development 4th ed. (Oxford University Press, USA).

Zamparini, A. L., Davis, M. Y., Malone, C. D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G. J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development *138*, 4039–4050.

Zhang, P., and Spradling, A. C. (1993). Efficient and dispersed local P element transposition from Drosophila females. Genetics *133*, 361–373.

Curriculum Vitae

Franz Sebastian Gruber

Date and place of birth: 07.12.1985, Schärding am Inn
Nationality: Austria
franzsebastiangruber@gmail.com

Education

2011
Diploma thesis under supervision of Julius Brennecke at the IMBA Vienna

10/2006 – 12/2011
Studies of Molecular Biology at the University of Vienna

10/2005-09/2006
Community service at the general hospital, Schärding am Inn

2005
General qualification for university entrance („Matura"), BORG Ried im Innkreis

Research experience

11/2010-12/2010internship in the lab of Julius Brennecke, IMBA Vienna
07/2010-09/2010
internship in the lab of Francesca Peri, EMBL Heidelberg

04/2010
internship in the lab of Carrie Cowan, IMP Vienna

12/2009
internship in the lab of Prof. Ulrich Technau, Department of Molecular Evolution and Development, University of Vienna

08/2009-09/2009
internship in the lab of Christine Hartmann, IMP Vienna

02/2009
internship in the lab of Prof. Gustav Ammerer, MFPL Vienna