Animal microRNAs and the evolution of tissue identity

Foteini Christodoulou

October 2009, Heidelberg

Joint PhD between European Molecular Biology Laboratory and
University of Crete, School of Sciences, Department of Biology

Joint PhD EMBL- University of Crete
Animal microRNAs and the evolution of tissue identity
Foteini Christodoulou

PhD supervisor:
Detlev Arendt, Professor, European Molecular Biology Laboratory

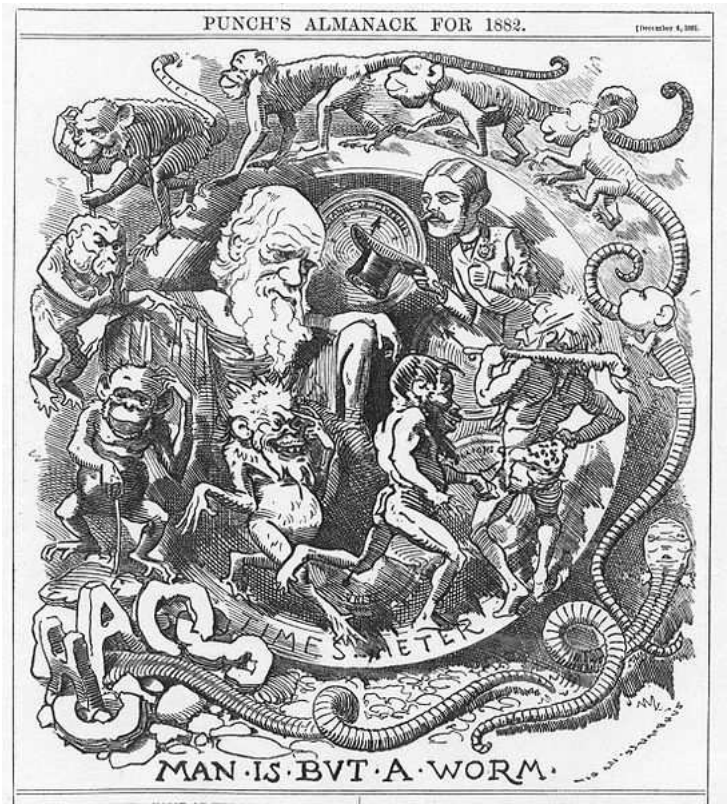
PhD Thesis Defense Committee:
Detlev Arendt, Professor, European Molecular Biology Laboratory
Michalis Averof, PhD, Institute of Molecular Biology and Biotechnology in Crete
Christos Delidakis, Professor, University of Crete, Department of Biology
Matthias Hentze, Professor, European Molecular Biology Laboratory
Ioannis Iliopoulos, Lecturer, University of Crete, Medical school
Kriton Kalantidis, Assistant Professor, University of Crete, Department of Biology
Domna Karagogeos, Professor, University of Crete, Medical school

This study was carried out at the European Molecular Biology Laboratory, Heidelberg

Στη Χριστίνα και τους γονείς μας

"Είμαστε σκουληκάκια μικρά μικρά [...] απάνω σ' ένα φυλλαράκι γιγάντιου δέντρου. Το φυλλαράκι αυτό είναι η γης μας. Τ' αλλα φύλλα είναι τ' αστέρια που βλέπεις να κουνιούνται μέσα στη νύχτα. Σουρνόμαστε απάνω στο φυλλαράκι μας, και το ψαχουλεύουμε με λαχτάρα τ' οσμιζόμαστε, μυρίζει, βρωμάει, το γευόμαστε, τρώγεται, το χτυπούμε, αντηχάει και φωνάζει σαν πράμα ζωντανό. Μερικοί άνθρωποι, οι πιο ατρόμητοι, φτάνουν ως την άκρα του φύλλου. Από την άκρα αυτή σκύβουμε, με τα μάτια ανοιχτά, τα αυτιά ανοιχτά, κάτω στο χάος. Ανατριχιάζουμε. Μαντεύουμε κάτω μας το φοβερό γκρεμό, ακούμε ανάρια ανάρια το θρο που κάνουν τα φύλλα του γιγάντιου δέντρου, νιώθουμε το χυμό ν' ανεβαίνει από τις ρίζες του δέντρου και να φουσκώνει την καρδιά μας. Κι έτσι σκυμμένοι στην άβυσσο, νογούμε σύγκορμα, σύψυχα, να μας κυριεύει τρόμος. Από τη στιγμή εκείνη αρχίζει..."

Βίος και πολιτεία του Αλέξη Ζορμπά -Του Νίκου Καζαντζάκη



Contents

Acknowledgements

I wish to express my gratitude to my PhD supervisor, Detlev Arendt, for the excellent guidance and inspiring mentoring he provided me with in the past 4 years. I feel lucky because he never discouraged my affinity towards the RNA field although it was not his primary scientific focus. By giving me the freedom to study what I was most passionate about and by being daring and open to integrate my findings in his field of expertise, evo-devo, we could bridge the two fields and assemble the exciting evolutionary scenario this thesis presents.

The microRNA project would have never spin off had I not been supervised by Alexandra Boutla in the beginning of my PhD. Alexandra taught me most things I know about microRNAs and continuously supported me even from distance (after permanently moving to Crete). Therefore, one of the best things about my PhD project was the opportunity to become good friends with her.

All members of the Arendt lab are especially thanked for an amazing atmosphere and for being so collaborative and open to discuss ideas and help me whenever needed. Especially Antje Fischer (although academically younger) has supported me so many times (with the patience and skill of a post-doc) and took time to brainstorm whenever I reached dead-ends in my project or went through phases of demotivation. Her strong zoological and evolutionary background complemented my predominantly molecular expertise. Raju Tomer was always keen to help me with any computational challenge I faced in my project and has on several occasions helped me analyze my data. Together with Alexandru Denes we had many interesting and inspiring scientific and philosophical discussions which always stimulated and challenged my mind. Maria-Antonietta Tosches has been a delightful desk/bench neighbour in the lab with whom I had several brainstorming sessions with exciting conclusions and creative moments. Oleg Simakov has contributed to the target analysis of this thesis and after our pleasant collaboration I am especially happy he is following up the microRNA proeject. Heidi Snyman, our great techy, has been there to support me both as yet another PhD student but mostly as a good friend from day 1 of my PhD. Without her, the lab atmosphere would not be as cheerful. Patrick Steinmetz taught me how to take good white light pictures using Nomarski optics and has been my *Nematostella* collaborator and expert who constantly helped me with protocols and guidance on understanding the animal's morphology. Kristin Tessmar-Raible is the person who first taught me most of what I know about *Platynereis* and I especially thank her for her patience and mentoring over the first months of my PhD. Many thanks for protocols, discussions, group meetings feedback and overall support to Mette Handberg-Thorsager, Antonella Lauri, Elia Benito Gutierrez, Daria Gavriouchk-

ina, Keren Guy, Tomas Larsson and Diana Bryant. It has been a great pleasure to spend 4 years in a tiny little lab with all of you Arendt lab crew!

I also wish to thank people outside the laboratory in pretty much every floor of EMBL who made my life easier with the project. I especially thank Vladimir Benes for being so supportive from the beginning of my PhD. Vladimir has been catalytic for the success of my PhD project both by providing me with his expertise in molecular biology but also for acting as a second academic mentor independent to Detlev. His initiative to involve me in so many courses organized by genecore contributed to my development of teaching and presentation skills and I really enjoyed participating. Jens Stolte from Genecore taught me everything I know about qPCR and I sincerely thank him for this. Sabine Schmidt has also kindly offered her help and taught me about microRNA microarrays. From the Bork lab, Mani Arumugam and Kallia Trachana were always there when I was stuck with lots of data and little computational ability to analyze them. From the Hentze lab, Bruno Galy, Dunja Ferring, Agnieszka Zdanowicz, Francesca Moretti have on several occasions helped me with protocols and given me the latest insider info from the microRNA field. Also members of the Cohen lab, Natasha Bushati and Valerie Hilgers significantly helped me with protocol troubleshooting.

Outside EMBL, I wish to thank collaborators who contributed to my PhD project. Julius Brennecke helped me clone all small RNAs from *Platynereis* and has since been very supportive also by giving us useful feedback on the manuscript. Greg Hannon is especially thanked for prioritizing my solexa libraries sequencing over a long pipeline and for being especially enthusiastic about my project. Enrique Arboleda has been my sea urchin collaborator and aside from providing me with material he was additionally helpful by patiently explaining to me the plutei morphology. His PhD supervisor, Ina-Maria Arnone was also very kind to provide me with more material when needed.

My thesis advisory committee has been very supportive by suggesting experiments and constructively criticizing my project plans on the planned annual meetings but also anytime I have seeked out for help. For this, I cordially thank Anne Ephrussi, Lars Steinmetz, Christos Delidakis, Domna Karagogeos and Matthias Hentze.

This thesis has been revised after useful comments from Detlev Arendt, Antje Fischer, Maria-Antonietta Tosches, Raju Tomer, Diva Tomei, Paul Bertone, Francesca Moretti and Sevil Yavuz. Special thanks to all of you.

Aside from work-related acknowledgements I wish to thank my friends at EMBL without whom the PhD would be less fun: Sevil Yavuz was the "Turkish delight" of my time at EMBL. Together we spent endless time discussing, watching movies, taking walks, eating in great restaurants (or not when our salary was zero) but also keeping each other

company on boring endless microscopy nights. We spent Xmas 2008 in the confocal room and nevertheless had lots of fun enjoying each other's company!

The "greek mafia" was also a delight of my PhD with lots of cooking and eating and giggling with Evangelia Petsalaki, Melpi Platani, Piyi Papadaki, Anastasios Koutsos, Georgios Pavlopoulos, Thodoris Soldatos, Stella Lamprinaki, Tina Diamantara, Yannis Legouras, and above all, my dearest flatmate Kallia Trachana. With Kallia we became very good friends after sharing the apartment and I am especially grateful to her because she supported me in a non-stop/mother-like manner in the hardest last two years of the PhD. I also thank Kallia for translating this thesis summary in greek.

Mani Arumugam, my favorite "talented" predoc has been a great friend throughout my PhD and I especially thank him for making Heidelberg a far more interesting place than it really is. Together we organized the famous jazz-bar movie nights and spent a great deal of time watching movies, chatting and travelling around Heidelberg and India. Having mentioned "talented" predocs, I have to thank Reaka Aiyar for being so cheerful and socially active and for all the fun moments we had together.

Last but not least, I wish to thank all my friends and family from Greece who have been so patient with me in the last 4 years. I thank them because although I have been very absorbed by my work, they never doubted how much I love them and likewise they never stopped supporting me.

Abstract

This thesis addresses the emerging link between the evolution of microRNAs and the evolution of complex bilaterian body plans. Recent deep sequencing of various metazoan animals revealed that early metazoans possessed at least one conserved microRNA, miR-100 (plus an unknown number of non conserved microRNAs), growing to 36 microRNAs in the last common ancestor of protostomes and deuterostomes. To further explore the putative link between the evolution of stem bilaterians and microRNAs, I set out to determine the ancient sites of activity of conserved microRNA families in a comparative approach.

I investigated the full set of 36 conserved bilaterian microRNAs in two slow-evolving protostome animals, the marine annelids *Platynereis* and *Capitella*, in a slow-evolving deuterostome, the sea urchin *Strongylocentrotus* and in a basal metazoan, the sea anemone *Nematostella*. The resulting comparative expression data showed that in these animals, other than in the fast-evolving fly *Drosophila* and nematode *Caenorhabditis*, microRNAs largely retain their ancient expression sites.

The oldest animal microRNA, miR-100, together with the co-transcribed let-7 and miR-125, was found expressed in a small group of neurosecretory cells located around the mouth, in cnidarians, annelids and sea urchin. This is where the conserved role of let-7 and miR-125 in developmental timing must have evolved. Other sets of ancient microRNAs were first present in locomotor ciliated cells, specific brain centres, or, more broadly, one of four major organ systems: central nervous system, sensory tissue, musculature and gut.

Insights into the contribution of the step-wise acquisition of microRNA families towards bilaterian complexity are given in the second part of this thesis. Using *Platynereis* and *Capitella* developing annelids, I localized the expression of 7 microRNAs specific to the protostome, 6 to the lophotrochozoan, 2 to the annelid and 2 to the *Platynereis* lineage. In most cases lineage specific microRNAs appeared to reinforce the regulatory signature of ancient bilaterian microRNAs by joining their expression in the above stated tissues. However few of them were highly restricted to subsets of these ancient bilaterian tissues.

In conclusion, the expression data of this comparative study suggest that both ancient bilaterian and "younger" lineage -specific microRNAs evolved in a tissue -specific context and conferred developmental robustness to an ancient set of animal cell types and tissues. They also imply that these microRNA-defined tissues were in place already in the last common ancestor of protostomes and deuterostomes. By this property, microRNAs provide a new tool for reconstructing ancient animal body plans at important evolutionary nodes, focussing here on the protostome-deuterostome divergence.

Περίληψη

Η εξέλιξη της περιπλοκότητας του σώματος των σύγχρονων αμφισυμμετρικών ζώων και πιο συγκεκριμένα το πως αύτη επηρεάστηκε από την εξέλιξη των μικρών-RNA αποτελούν σκοπό αυτής της διδακτορικής εργασίας. Μετά από πρόσφατη εις βάθος αλληλούχηση (deep sequencing) των μικρών-RNA ζώων που κατατάσσονται σε γενεαλογικές γραμμές πριν και μετά τον διαχωρισμό των πρωτοστόμιων-δευτεροστόμιων, βρέθηκε ένα κοινά διατηρημένο μικρό-RNA, το miR-100.

Για την περαιτέρω μελέτη των μικρών-RNA στην εξέλιξη των αμφισυμμετρικών ζώων, και πιο συγκεκριμένα εαν τα μικρά-RNA έπαιξαν σημαντικό ρόλο στην διαμόρφωση του σώματος του κοινού προγόνου των προτεοστόμιων και δευτεροστόμιων, προσπάθησα να χαρακτηρίσω και να συσχετίσω την δράση 36 αρχαίων μικρών-RNA -τα οποία έχουν διατηρηθεί και στις δύο γενεαλογικές γραμμές- σε οργανισμούς που εξελίσσονται αργά.

Οι θαλάσσιες αννελίδες, *Platynereis dumerilii*, *Capitella spI*, και ο αχινός *Strongylocentrotus purpuratus* εκπροσωπούν τα αργά εξελισσόμενα πρωτοστόμια και δευτεροστόμια, αντίστοιχα. Σημαντικός οργανισμός σε αυτή την σύγκριση είναι η θαλάσσια ανεμώνη, (*Nematostella*), που εκπροσωπεί έναν προδρομικό κλάδο των σύγχρονων αμφισυμμετρικών ζώων, τα κνιδάρια.

Συγκρίνοντας την δράση των αρχαίων μικρών-RNA στα παραπάνω ζώα φάνηκε ότι τα πρώτα έχουν διατηρήσει την έκφραση τους κατά την διάρκεια της εξέλιξης, ενώ ταυτόχρονα επιτράπηκε η πρόβλεψη τις δράσης τους στον πρόγονο των πρωτοστόμιων-δευτεροστόμιων. Αντίθετα, το παραπάνω συμπέρασμα δεν ήταν εμφανές από την σύγκριση των πρότυπων έκφρασης των αρχαίων μικρών-RNA σε ζώα που εξελίσσονται γρήγορα, όπως είναι τα είδη *Drosophila* και *Caenorhabditis*.

Ενδιαφέρον είχε το πρότυπο έκφρασης του αρχαιότερου μικρού- RNA, miR-100, το οποίο συν-μεταγράφεται με τα let-7 και miR-125. Όλα μαζί εκφράζονται σε μία μικρή ομάδα νευρικών-εκκριτικών κυττάρων που εντοπίζονται γύρω από το στόμα. Αυτό το πρότυπο έκφρασης ήταν κοινό στα κνιδάρια, στις αννελίδες και στον αχινό. Πιθανώς να υποδεικνύει τον αρχικό ρόλο των let-7 και miR-125 στην επιλογή της χρονικής στιγμής για την μεταμόρφωση κατά την διάρκεια της ανάπτυξης του ζώου, μια διατηρημένη λειτουργία στην εξέλιξη των μεταζώων. Τα υπόλοιπα αρχαία μικρά-RNA εντοπίστηκαν σε κινητικά βλεφαριδωτά κύτταρα σε συγκεκριμένα τμήματα του εγκεφάλου και σε ένα από τα τέσσερα οργανικά συστήματα: κεντρικό νευρικό σύστημα, αισθητήρια όργανα, μυϊκό σύστημα και πεπτικό σύστημα.

Το δεύτερο μέρος της διδακτορικής διατριβής εξετάζει την σχέση της κλιμακούμενης απόκτησης των μικρών-RNA και της περιπλοκότητας του σώματος των ζώων. Σε αυτό το σημείο εστιάζομαι, κυρίως, σε μικρά-RNA που είναι διατηρημένα σε μια γενεαλογική γραμμή και πως αυτά συσχετίζονται με σωματικά χαρακτηριστικά αυτής. Η μελέτη αυτή περιορίστηκε

στις δυο αννελίδες (*Platynereis* και *Capitella*), στις οποίες μελετήθηκε η έκφραση των μικρών-RNA που εμφανίζονται αποκλειστικά σε πρωτοστόμια, λοφοτροχοζώα, αννελίδες και *Platynereis*, αντίστοιχα. Στις περισσότερες περιπτώσεις τα γενεαλογικά-εξειδικευμένα μικρά-RNA συν-διαμορφώνουν με τα αρχαία μικρά-RNA την κυτταρική ταυτότητα των προαναφερθέντων οργανικών συστημάτων.

Συνοψίζοντας, η παραπάνω συγκριτική μελέτη προτείνει ότι τόσο τα αρχαία, όσο και τα νεότερα γενεαλογικά-εξειδικευμένα μικρά-RNA προσφέρουν αναπτυξιακή σταθερότητα σε μια ομάδα αρχέγονων κυτταρικών τύπων και οργάνων, οι οποίοι φαίνεται να ήταν παρόν στον κοινό πρόγονο των πρωτοστόμιων-δευτεροστόμιων. Επιπλέον, τα αρχαία μικρά-RNA διατηρούν το πρότυπο έκφρασης τους σε αργά-εξελισσόμενα ζώα αποτελώντας ένα νέο εργαλείο για να την έρευνα της εξέλιξη του σχεδίου σώματος των μεταζώων και πιο συγκεκριμένα του προγόνου των πρωτοστόμιων-δευτεροστόμιων.

1 Introduction

1.1 The chronicle of microRNA discovery

In 1993, a forward genetic screen for genes involved in developmental timing of *Caenorhabditis elegans* revealed for the first time the regulatory role of a small RNA in animal development (Lee et al., 1993; Wightman et al., 1993). The reported defects in *C. elegans* early embryonic development in lin-4 null mutants were attributed to the alleviated downregulation of lin-14 protein whose temporal gradient specifies the temporal order of cell lineages (Ruvkun and Giusto, 1989). Despite the observation that the 5' end of the 22nt long lin-4 microRNA had a number of complementary sites in the 3' untranslated region (UTR) of the lin-14 mRNA, the finding did not attract enough attention because at the time lin-4 orthologs were non identifiable in other animals and the finding was considered a nematode specific peculiarity (Ambros, 2008).

Seven years later, the discovery of yet another regulatory small RNA of *C. elegans* , let-7, this time with a conserved sequence across bilaterally symmetrical animals (Reinhart et al., 2000; Pasquinelli et al., 2000) followed by that of the *Drosophila* microRNA bantam (Brennecke et al., 2003) caught the eyes of the RNA community, which joined the small RNA field in search for more such regulators and their mode of action. From then on, animal microRNA discovery has been a subject of molecular cloning and sequencing in both invertebrates (Lau et al., 2001; Lee and Ambros, 2001; Ruby et al., 2006) and vertebrates (Lagos-Quintana et al., 2001; Berezikov et al., 2006; Landgraf et al., 2007) as well as bioinformatics predictions (Sandmann et al., 2007; Stark et al., 2007). Recent development of sequencing technologies (Lu et al., 2005; Margulies et al., 2005) has allowed the discovery of less abundant microRNAs leading to a total of at least 6930 microRNA sequences deposited for animals and their viruses in the latest version of the microRNA repository; miRBase (Griffiths-Jones et al., 2008). Defining characteristics for all animal microRNAs deposited in miRBase are their short size (~23nt), their non-coding nature and biogenesis and their repressive function by imperfectly binding to target mRNAs 3' UTRs.

Up to date, thousands of articles about microRNAs have contributed to growing knowl-

edge on biological processes in which microRNAs seem to play a role. Currently, microRNAs are computationally predicted to control at least 20-30% of animal transcripts which bear at least one or more microRNA binding sites in their 3' UTRs (Lewis et al., 2005; Krek et al., 2005; Ruby et al., 2007).

1.2 microRNA biogenesis

1.2.1 Transcription of microRNA genes

Depending on their genomic position with respect to protein coding genes, microRNA genes can be either intergenic or intronic (Kim and Nam, 2006). Intergenic microRNAs, such as lin-4 and let-7 in *C. elegans*, do not overlap with other genes and form their own transcriptional units with a promoter and a poly-A terminator sequence (Saini et al., 2008). Of the intronic microRNAs, many are hosted in protein coding genes' introns and for such cases co-expression of microRNA and its host gene mRNA has been demonstrated (Baskerville and Bartel, 2005). Non-coding RNA (ncRNAs) transcription units can also host microRNAs in both their introns and exons (Rodriguez et al., 2004).

Most microRNA genes' transcription relies on the activity of RNA Pol II (Lee et al., 2004). Several Pol-II-dependent transcription factors have been linked to microRNA transcriptional regulation (Lee and Dutta, 2009) indicating that microRNA expression can be regulated in different cell types and developmental time points.

The primary microRNAs (pri-miRNAs) that Pol-II transcribes are capped and polyadenylated (Lee et al., 2004; Rodriguez et al., 2004). Their length can vary between 1-10kb (Saini et al., 2008)with local stem loop structures protruding as candidates for further processing by the Microprocessor complex (Figure 1.1;Kim et al., 2009).

Involvement of RNA polymerase III in the transcription of microRNA genes has also been shown for few cases such as microRNAs which lie among Alu repeats (Borchert et al., 2006) and for microRNAs encoded by viruses (Pfeffer et al., 2005; Andersson et al., 2005).

1.2.2 RNA editing on primary transcripts

Modifications of adenosine (A) into inosine (I) by adenosine deaminases acting on RNA (ADARs) within the stem loop of a precursor microRNA (pre-miR) can influence subsequent processing steps. Since the base pairing properties of inosine are similar to guanosine (G), a single nucleotide change can affect the structural properties of the pre-miR stem loop (by introducing an additional bulge or eliminating one). Single nucleotide

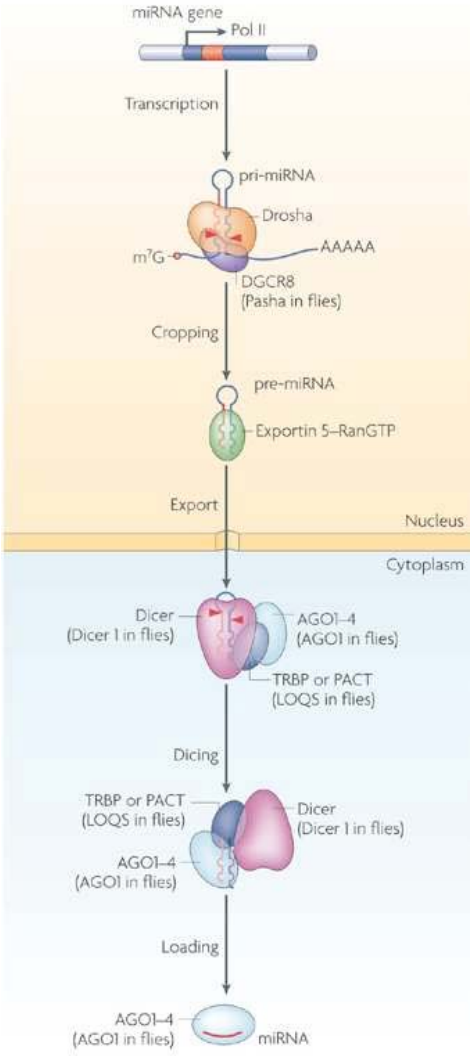


Figure 1.1: **The canonical microRNA biogenesis pathway.** microRNAs are transcribed by Pol II into primary transcripts (pri-miRNAs) which get cleaved in the nucleus by Drosha/DGCR8 to yield a shorter precursor (pre-miRNA). The latter get exported from the nucleus to the cytoplasm where they get further processed by Dicer to generate a shorter microRNA::microRNA* duplex. Once loaded into AGO1, the microRNA strand can direct translational repression and mRNA destabilization of its target mRNAs. Figure adapted after Kim et al. (2009).

polymorphisms (SNPs) have been shown to drastically affect the processing eligibility of pri-miRNAs (Duan et al., 2007).

Like with SNPs, there are examples where RNA editing led to more efficient nuclear processing of pri-miRs (Kawahara et al., 2008) as well as examples where the new structural properties inhibited processing (Yang et al., 2006). Aside from affecting microRNAs road towards maturity, ADAR modification has the power to redirect a microRNA's target set of mRNA transcripts and therefore affect its function (Ohman, 2007; Kawahara et al., 2007).

1.2.3 Nuclear processing of primary transcripts

In this first step towards microRNA maturation, primary trancripts (pri-miRNAs) get cleaved at the base of microRNA stem loop structures which protrude from the transcript owing to their secondary structure. An average human pri-miRNA, has a hairpin stem of 33 base-pairs with a terminal loop and single-stranded flanking sequences upstream and downstream of the hairpin (Winter et al., 2009).

The cleavage reaction takes place in the nucleus by the nuclear RNaseIII type endonuclease named Drosha (Lee et al., 2003). Together with its cofactor, DGCR8 (in humans) or Pasha (in *Drosophila* and *C.elegans*), which contains two dsRBD domains, they form the Microprocessor complex (Denli et al., 2004). DGCR8/Pasha recognizes the junction between flanking ssRNA and dsRNA at the base of the stem loop where it anchors to form a "cleavage complex" which facilitates the positioning of Drosha so that it cuts 11bp from the base of the stem loop (Figure 1.1; Han et al., 2006). The product of this reaction is a precursor microRNA (pre-miRNA) which is 60 to 80nt long with a 5' monophosphate and a 3' 2nt overhang (Kim et al., 2009).

Recent cloning and sequencing of small RNAs in *Ciona intestinalis* revealed a previously unnoticed class of small RNAs which are slightly shorter than microRNAs (19-20nt long) and are mapped to flank the predicted stem loop of the pre-miRNA at both 5' and 3'(Shi et al., 2009). Such "microRNA -offset RNAs" (moRs) have also been detected in human small RNA libraries albeit less abundantly (Langenberger et al., 2009). Their processing still remains to be elucidated.

There is growing evidence that pri-miRNA processing takes place co-transcriptionally. Pri-miRNAs have been found enriched in the chromatin-associated nuclear fraction (Pawlicki and Steitz, 2008) and chromatin immunoprecipitations revealed enriched Drosha and Pol-II binding on the same microRNA loci in human cell lines (Morlando et al., 2008). In addition, the nuclear exosome (3'-5' exonuclease) and XRN2 (5'-3' exonuclease) were found co-localised with Drosha on the introns of microRNA host genes, suggesting that

after Drosha processing the introns get excised and degraded (Morlando et al., 2008). The emerging model suggests that Drosha cleaves after the transcript associates to the early spliceosome complex but before the intron gets excised (Kim et al., 2009).

1.2.3.1 Mirtrons skip nuclear processing

Certain introns which host microRNAs can be so small in size, that upon completion of splicing and debranching they fold into hairpin structures resembling those of pre-miRNAs and can thus by-pass canonical processing by Drosha/Pasha (Ruby et al., 2007). Mirtrons have been identified in both invertebrates and vertebrates but unlike canonical microRNAs they appear to be phylogenetically restricted in the phyla where they are discovered (Okamura et al., 2007; Berezikov et al., 2007).

1.2.4 Nuclear export of precursor microRNAs

Transport of pre-miRNAs into the cytoplasm is mediated by Exportin-5 (EXP5), a member of the karyopherin family of transport receptors, in a RanGTP-dependent manner (Bohnsack et al., 2004; Lund et al., 2004). EXP5 can recognize dsRNA stems longer than 14bp with a short 3' overhang of 1-8nt (Gwizdek et al., 2003; Zeng and Cullen, 2004) on which it directly binds without the need for any adaptor proteins (Lund et al., 2004). Once in the cytoplasm, GTP gets hydrolyzed to GDP and EXP5/Ran-GDP then releases the pre-miRNA (Figure 1.1).

1.2.5 Cytoplasmic processing of precursor microRNAs

The second processing step towards microRNA maturation involves the excision of pre-miRNAs terminal loop and takes place in the cytoplasm (Figure 1.1). Dicer, a cytoplasmic RNase III endonuclease, can recognize and bind the 3' 2nt overhang that Drosha cleavage leaves behind, through its PAZ domain which preferentially binds single-stranded 3' ends of dsRNAs (Lingel et al., 2003; Ma et al., 2004). Upon PAZ binding, Dicer cuts 20bp from the base of the stem to remove the loop and leave an additional 2nt 3' overhang (Zhang et al., 2004). Imprecise or alternative processing by Dicer could account for the observed length heterogeneity at 3' ends of mature microRNAs (Landgraf et al., 2007; Ruby et al., 2007; Seitz et al., 2008).

However, similar to Drosha, Dicer activity requires association with proteins which bear dsRNA binding domains such as Loquacious (LOQS) in *Drosophila* (Forstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005) and its ortholog TRBP in addition to PACT in human (Figure 1.1;Chendrimada et al., 2005; Lee et al., 2006). The final

product, is a ~22nt miRNA:miRNA* duplex with 2nt overhangs at both 3' ends (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001).

1.2.6 Loading mature microRNAs to Argonautes

To exert their regulatory role, microRNAs must associate with the argonaute proteins (AGOs) and form the RNA-induced silencing complex (RISC) which will cause translational repression and/or destabilization of target mRNAs (Figure 1.1; Carthew and Sontheimer, 2009).

The mature miRNA:miRNA* duplex product of Dicer enters miRISC assembly (while still bound by Dicer and LOQ/TRBP) which involves duplex unwinding and association of one of the two strands with AGO (Gregory et al., 2005; Maniataki and Mourelatos, 2005). Strand retention is influenced by thermodynamic criteria similar to those which determine which strand of an siRNA duplex will be the guide and which the passenger (Schwarz et al., 2003; Khvorova et al., 2003). MicroRNAs can mature from either arm of the pre-miR stem loop and in some cases both arms produce mature microRNAs (Ruby et al., 2007; Okamura et al., 2008; Ro et al., 2007).

During RISC assembly, there is a choice of AGO proteins on which a mature microRNA can be loaded. Five, eight and twenty-seven paralogs of the argonaute superfamily exist in *Drosophila*, human and *C. elegans* respectively (Carthew and Sontheimer, 2009). Among paralogs, some have been shown to be functionally specialized for the miRNA and siRNA silencing pathways in each species (Okamura et al., 2004; Yigit et al., 2006; Diederichs and Haber, 2007). In invertebrates, the basic criterion over which small RNAs are sorted/loaded to different AGOs is the structure of the pre-miRNA stem (Tomari et al., 2007).

In *Drosophila*, Dicer/LOQ bound microRNA duplexes with disturbed base pairing (by central mismatches) preferentially associate with Ago1 (Tomari et al., 2007; Forstemann et al., 2007). Recent high throughput sequencing of AGO1 associated small RNAs confirmed that the vast majority were microRNAs (Czech et al., 2008). Only a small fraction of AGO2 associated small RNAs corresponded to microRNA sequences while the majority comprised siRNAs which originate from perfectly matching duplexes which are specifically bound by Dicer/R2D2 complex (Czech et al., 2008; Tomari et al., 2004). Similarly, the *C. elegans* argonaute family member ALG-1 gets loaded with mature microRNAs originating from missmatched precursors while RDE-1 associates with small RNAs unwound from perfect matched dsRNA precursors (Steiner et al., 2007). In humans, which have four AGO proteins (AGO1-4), binding preferences for microRNAs have been reported for AGO3 and AGO2 after sequencing all AGOs associated small RNAs

(Azuma-Mukai et al., 2008).

1.2.7 MicroRNA delivery to neighbouring cells through exosomes

Investigation for the presence of nucleic acids in exosomes of mouse and human mast cell lines identified substantial amounts of RNA, in particular mRNA from ~1300 genes but also large amounts of small RNAs, including 121 different microRNAs (Valadi et al., 2007). Similar, recent findings were reported for embryonic stem cell microvesicles which were found to carry protein and RNA of size <2kb among which several abundant microRNAs (Yuan et al., 2009). There is therefore growing evidence suggesting that in cell types which have the capacity to release exosomes, a hypothetical final step in microRNA biogenesis could be their packaging and export. Such a model would assign an additional role to microRNAs in intercellular regulation of gene expression.

1.3 microRNA mediated silencing mechanisms

Following the assembly of the silencing complex, the mature microRNA guides miRISC to its target mRNA. The mechanism through which miRISC will exert gene silencing depends on the degree of complementarity between the microRNA and its target sequence. Up to now, three different mechanisms have been reported: mRNA cleavage, mRNA degradation and translational repression.

1.3.1 mRNA cleavage

Of the three known silencing mechanisms, mRNA endonucleolytic cleavage is the best understood. This is because of the thorough characterization of AGO2, the only AGO protein which can catalyze endonucleolytic cleavage of a base-paired (perfectly matching) miRNA::target duplex through an RNaseH-like fold of its PIWI domain (Liu et al., 2004; Meister et al., 2004; Okamura et al., 2004). However, such "slicing" repression activity against target transcripts has only been reported for a couple of mammalian microRNAs; miR-198 which guides the enonucleolytic cleavage of HOXB8 mRNA (Yekta, 2004) and miR-122 that guides cleavage of Transition protein 2 (Tnp2) mRNA (Yu et al., 2005). In contrast to this, "slicing" is a very widespread target transcript repression mechanism in plants where microRNAs bind their targets with full complementarity (Jones-Rhoades et al., 2006).

1.3.2 mRNA degradation

In the last 5 years, work from several groups has established deadenylation followed by decapping and degradation of target transcripts as one of the important silencing mechanisms by microRNAs (Bagga et al., 2005; Lim et al., 2005; Giraldez, 2006; Rehwinkel et al., 2006; Eulalio et al., 2009). In eukaryotes, mRNA degradation can either be exerted by a 3'-5' exonuclease in the endosome after deadenylation (Chen et al., 2001; Hilleren et al., 2001; Wang and Kiledjian, 2001; Tourrière et al., 2002)or by a 5'-3' exonuclease XRN1P/LSM1P after deadenylation and decapping by DCP1/DCP2 (Muhlrad et al., 1995; Parker and Song, 2004).

The aforementioned mRNA destabilization components have been found enriched (Rehwinkel et al., 2005; Bagga et al., 2005) together with microRNAs, AGO proteins and target messengers in cytoplasmic foci, called P-bodies, known as sites of catabolism and/or storage of non-translated mRNAs (Liu et al., 2005a; Sen and Blau, 2005; Liu et al., 2005b; Jakymiw et al., 2005). Despite the fact that P-body integrity was shown to be non-essential for microRNA function (ying Chu and Rana, 2006), P-body components do play an important role in microRNA-mediated silencing and their formation was shown to be a consequence rather than the cause of RNA-mediated silencing (Eulalio et al., 2007).

MicroRNA mediated mRNA decay requires AGO proteins, the P-body component GW182, the CAF1-CCR4-NOT deadenylaze complex, the decapping enzyme DCP2 and decapping activators like DCP1, Ge-1, EDC3 and RCK/p54. GW182 recruits the CAF1-CCR4-NOT complex and induces deadenylation of the target mRNA which subsequently gets decapped by DCP1-DCP2 complex and degraded by the 5'-3' exonuclease XRN1 (Behm-Ansmant et al., 2006; Eulalio et al., 2007). Recently, it has been shown that GW182 additionally interacts with the poly(A)-binding protein (PABP) which is a co-activator of deadenylation (Fabian et al., 2009). The P-body component GW182 (which acts downstream of microRNA processing and loading to AGO) appears to be of primary importance in this silencing mechanism not only for recruiting the mRNA destabilization machinery but primarily for associating with AGO proteins at the target mRNA's 3'UTR (Behm-Ansmant et al., 2006; Till et al., 2007).

The latest state of the art on deadenylation and degradation as a microRNA-mediated silencing mechanism indicates that although deadenylation is not crucial for silencing, it does contribute by enhancing the inhibitory effect of microRNAs which primarily repress translation (Eulalio et al., 2009; Zdanowicz et al., 2009; Fabian et al., 2009).

1.3.3 Translational repression

The mechanisms by which miRISC exerts translational repression are still an ongoing subject of debate. Many independent experiments suggest that microRNAs block translation at the initiation step. However, some evidence is pointing towards post-initiation steps. It is still not clear whether the lack of consensus is a product of "artefactual noise" due to the different experimental systems used. Below, I shall present the proposed models:

1.3.3.1 Inhibition of ribosomal subunit joining or cap recognition

The observation that mRNAs with a non-functional cap or an IRES cannot effectively be repressed by microRNAs indicated that repression possibly happens at the initiation step of translation and might involve the recognition of the m^7GpppN cap (Pillai et al., 2005; Humphreys et al., 2005; Thermann and Hentze, 2007; Mathonnet et al., 2007).

Several studies on cell free extracts reason that microRNAs repress translation during initiation by inhibiting association of target mRNA with the small ribosomal subunit 40S of *Drosophila* (Thermann and Hentze, 2007), the 80S initiation complex of mouse (Mathonnet et al., 2007) but also the 60S ribosomal subunit (Chendrimada et al., 2007; Wang et al., 2008). Relief of translational repression after addition of purified initiation factor eIF4F (which includes eIF4E cap-binding subunit) highlighted and further supported the importance of the m^7GpppN cap recognition process in miRNA-mediated repression (Mathonnet et al., 2007). Very recent work by Zdanowicz et al. (2009) confirmed the importance of the m^7GpppN cap, which was shown to be the primary target of miR-2/RISC as its substitution with chemically modified structural analogs (which were more accessible for RISC and translational inhibitor 4E-BP binding) increased susceptipility to microRNA-mediated repression.

At the level of whole organism study, work in *C. elegans* studying the repressive effect of let-7 and lin-4 on several endogenous targets reported their decreased association with polysomes consistent with the model of repression at the initiation step (Ding and Grosshans, 2009). Taken together, a model emerges whereby a microRNA induced barrier prevents the essential for translation synergy between the 5' cap and the 3' poly(A) tail and can exert repression of translation at the initiation step.

1.3.3.2 Cotranslational degradation of nascent protein

Translational repression via recruitment of proteases to polysomes was the suggested mechanism of microRNA silencing after failing to immunoprecipitate the repressed mRNA

with antibodies against the reporter growing polypeptide (Nottrott et al., 2006). Aside from basing conclusions on negative instead of positive evidence, this model came in contradiction with earlier findings that silencing is not affected after inhibition of the proteasome or targeting the protein to the Endoplasmic Reticulum (Pillai et al., 2005). These observations argue against a mechanism that would involve degradation of the nascent polypeptide.

1.3.3.3 Premature termination of translation

Investigation of translational repression using a microRNA mimic (a synthetic microRNA reporter with a 3' UTR which bares six identical partially complementary binding sites for cotransfected siRNA), showed association of the reporter to polysomes in addition to premature termination of translation (Petersen et al., 2006). These findings led the authors to the conclusion that microRNAs cause premature ribosome drop-off to actively translated mRNAs. However, earlier work had already shown that microRNA-repressed mRNAs do not co-sediment with polysomes (Pillai et al., 2005) and these complexes were later shown to be "pseudo-polysomes" which formed even after ribosome complex formation or elongation of translation were inhibited (Thermann and Hentze, 2007).

1.4 Target recognition by microRNAs

1.4.1 Target binding rules

As increasing numbers of microRNA targets were validated it became apparent that the 5' end of microRNAs, particularly nucleotides 2-8, conferred the most important functional role in target recognition (Lai, 2002). This core functional element binds to the 3' Untranslated Region (UTR) of mRNAs, repressing their expression either post-transcriptionally or at the level of translation. This particular region was termed the "seed" sequence (Lewis et al., 2003) in view of the later finding that this portion of the microRNA sequence is the most conserved across metazoan microRNAs (Lim et al., 2003). A number of single-nucleotide mutation studies monitoring the effect of repression of known microRNA targets confirmed the biological importance of the seed region (Doench and Sharp, 2004; Kiriakidou et al., 2004; Kloosterman et al., 2004; Brennecke et al., 2005). The first rule of microRNA target recognition was therefore established: as little as seven base-pairs of complementarity between the 3' UTR of mRNA targets and the 5' end of targeting microRNAs is sufficient to exert a repressive effect.

This observation raises the question as to why the first nucleotide of a microRNA's 5'

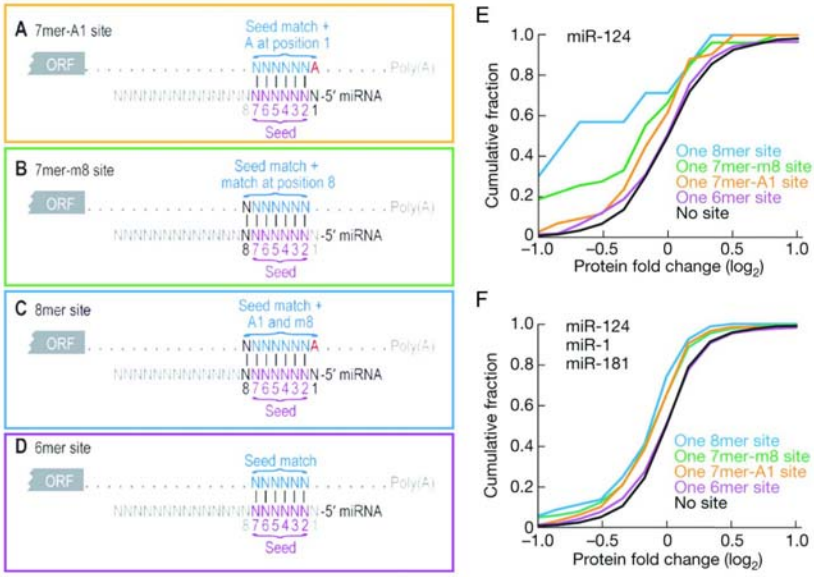


Figure 1.2: **Different types of target sites and their impact on protein output. a-c,** canonical 7 and 8mer target sites with vertical dashes indicating base pairing. **d,** marginal 6mer target site. **e-f,** protein output repressive efficacy of each target site type matching miR-124 and averaged data from matching miR-124, -1 and -181 respectively. Figure adapted by Baek et al. (2008)

end is not included in the seed -particularly when adenosines are common at position t1 (adjacent to microRNA seed matches) and most microRNA sequences contain a uracil at position m1 (first position from 5' end of microRNA). It has been shown that even when base pairing occurs between t1 (A) and m1 (U), this feature alone does not contribute to a repressive effect on the target (Nielsen et al., 2007). Interestingly, structural studies of the AGO-siRNA complex demonstrated that the 5' most nucleotide of the guide RNA (equivalent to the m1 base of microRNAs) is not base-paired but instead bound by AGO (Ma et al., 2005; Parker et al., 2005). Evolutionary conservation across components of the siRNA and microRNA pathways may give insight into the lack of t1-m1 recognition at the 5' end of the seed region.

Sites with an adenine at t1 position are named "7mer-A1". They lie at the bottom of the targeting efficacy hierarchy (Figure 1.2), with "7mer-m8" sites out-performing them and "8mer" sites demonstrating the strongest repressive effect (Grimson et al., 2007; Nielsen et al., 2007; Baek et al., 2008). Many microRNA targets exhibit recognition sites with predicted bulges or G:U matches (Ha et al., 1996; Reinhart et al., 2000; Brennecke et al., 2003), but surprisingly these have been demonstrated not to inhibit the target gene's function in vitro (Doench and Sharp, 2004; Kiriakidou et al., 2004). The number of bulges or G:U base pairs which can be tolerated in the seed region without precluding

target binding has been a research topic of its own. In vivo experiments have shown that more than one G:U base pair within the seed region could compromise the efficiency of microRNA-mediated target repression, a finding recently confirmed by large-scale whole proteome microRNA impact analyses (Brennecke et al., 2005; Selbach et al., 2008; Baek et al., 2008).

Target sites having either a single-nucleotide bulge or mismatch within the seed,can remain functional, given extensive binding complementary between the target and the 3' end of the microRNA (Yekta, 2004; Brennecke et al., 2005). Such "3'-compensatory sites" can supplement a 7mer seed match and additional target pairing to microRNA nucleotides 13-16 was also shown to boost target site efficacy (Grimson et al., 2007; Bartel, 2009). However, unlike the 7mer seed, 3' compensatory or supplementary sites do not appear to be under selection pressure (Brennecke et al., 2005; Lewis et al., 2005). These findings establish the second rule of target binding: the quality and extent of microRNA seed:target binding is directly linked to its repressive efficacy.

Of equal importance to the quality of sequence match between microRNA and target is the overall accessibility of the target site (Kertesz et al., 2007). RNA-binding proteins have been shown to interfere with microRNA-target interaction by occupying the target recognition site (Kedde et al., 2007). Even in the absence of interfering bound complexes, the secondary structure of the 3'UTR may be such that for microRNA binding to occur, unpairing of local secondary structures flanking the target site might be necessary (Kertesz et al., 2007). Recent work by Iovino et al. (2009) further underscores the importance of 3'UTR context by experimentally validating a predicted target with a (poor) 6mer site, not evolutionarily conserved (see below for details) but very accessible in terms of flanking 3' UTR secondary structure. This was computationally measured using the PITA algorithm (see below for details). It therefore appears that as a complement to purely sequence-based approaches, site accessibility should also be considered as a rule for microRNA target prediction.

1.4.2 Computational prediction of microRNA targets

Employing an experimental approach to validate each target gene annotated in miRBase would be a very laborious and time-consuming task. As an alternative, several target prediction algorithms have been developed to apply the above criteria to search a set of transcript sequences for potential target sites. A fundamental challenge in this field has been to successfully predict the biologically functional targets while excluding false predictions (Bartel, 2009).

To date, there are at least seven different target prediction tools publicly available, and

each one is based on distinct criteria with varying false-positive rates. Some specialize on a particular organism's transcriptome sequences, while others include data from related transcriptomes. These tools include: Targetscan (Lewis et al., 2005), EMBL (Stark et al., 2005), PicTar (Krek et al., 2005), EIMMo (Gaidatzis et al., 2007), Miranda (Betel et al., 2008), miRBase Targets (Griffiths-Jones et al., 2008) and PITA (Kertesz et al., 2007).

The performance of the above algorithms has been continuously assessed with a growing number of target validation experiments (Baek et al., 2008; Selbach et al., 2008). An additional prediction criterion appears to be target site evolutionary conservation, which is detected through alignment of orthologous 3' UTR sequences (Lewis et al., 2003). This indicator has been further confirmed by single-nucleotide polymorphism (SNP) genotyping in the human transcriptome, which reveals SNP density to be significantly lower in conserved target sites (Chen and Rajewsky, 2006b). Whole-transcriptome and -proteome analyses have reinforced the importance of target site evolutionary conservation by demonstrating a stronger down-regulation of the mRNAs having conserved sites (Farh, 2005; Nielsen et al., 2007; Baek et al., 2008; Selbach et al., 2008).

Despite the observed conservation of target sites within vertebrates (which surprisingly share high 3'UTR homology) very few target sites are conserved among drosophilid and nematode lineages (Chen and Rajewsky, 2006a). A recent study focusing on the function of miR-7 in *Drosophila* development identified only 9 orthologous targets of miR-7 shared between *Drosophila* and human (as an intersection from 97 *Drosophila* and 581 human predicted targets) (Li et al., 2009).

In addition to the evolutionary conservation filters widely used in target prediction, some methods factor in predicted secondary structure accessibility (Kertesz et al., 2007). In this approach, targets are scored by computing the difference ($\Delta\Delta$G) resulting from the subtraction of the free energy gained by forming the miRNA-mRNA duplex (ΔG_{duplex}) and the energetic cost of unpairing the target site in order to make it accessible for the miRNA (ΔG_{open}). A recent work highlighted the very low value of $\Delta\Delta$G (less than -12) as the only link able to discriminate between favorable (8mer and evolutionarily conserved) and unfavorable (6mer, non conserved and only predicted by one algorithm out of several tested) sites (Iovino et al., 2009). It would therefore seem that target prediction algorithms would benefit from the addition of filters designed to reduce the false negative rate.

With predictions in the range of 300 evolutionarily conserved targets per mammalian microRNA family, it appears that half of human protein-coding genes are potentially under the control of microRNAs, with a strong selective pressure to maintain their con-

served target sites in the 3' UTRs (Friedman et al., 2009). Conversely, the depletion of target sites in exceptionally short 3' UTRs of "housekeeping" genes (Stark et al., 2005), together with the observation that in cancer cells oncogene isoforms having shorter 3' UTRs escape miRNA-mediated repression (Mayr and Bartel, 2009), reveals target avoidance.

Based on target predictions it appears that microRNAs have the potential to modulate the expression of nearly all the mammalian mRNAs (Bartel, 2009). To support this notion with experimental evidence, target validation followed by functional analysis of a given microRNA's repressive effect is necessary.

1.4.3 Target validation

MicroRNA target prediction analysis eventually faces a bottleneck at the step of target validation. Several validation methods have been employed, ranging from traditional genetic studies (Lee et al., 1993; Reinhart et al., 2000; Brennecke and Cohen, 2003), rescue assays (Brennecke and Cohen, 2003), reporter-gene constructs (Lewis et al., 2003; Kiriakidou et al., 2004; O'Donnell et al., 2005; Johnson et al., 2005) and mutation studies (Doench and Sharp, 2004; Kiriakidou et al., 2004; Kloosterman et al., 2004; Brennecke et al., 2005). High-throughput approaches have also been developed, involving over-expression of microRNAs in cell lines followed by microarray profiling to detect down-regulated targets (Lim et al., 2005), as well as the reverse approach of depleting microRNAs to identify up-regulated targets (Rehwinkel et al., 2006).

Each of the above methods offers both clear benefits and disadvantages. For example, reporter construct assays provide high-resolution information about both mRNA and protein repression but are very time consuming and laborious. Measuring the effect of microRNA targeting on a whole transcriptome basis is of great value, but yields an incomplete picture since microRNAs have been seen to act primarily by blocking translation without additional target mRNA degradation (Baek et al., 2008; Selbach et al., 2008; Zdanowicz et al., 2009). This has been confirmed by studies in which immunoprecipitations of Ago1 pulled down predicted targets that were marked unchanged in the mRNA profiling experiments and thus considered false positive predictions while in reality being true targets (Easow et al., 2007; Karginov et al., 2007).

A recent method combining quantitative-mass-spectrometry with stable isotope labeling of amino acids in cell culture (SILAC) allowed sensitive measurements of genome-wide changes in protein output as a response to microRNA expression or depletion (Baek et al., 2008; Selbach et al., 2008). These studies showed that a single microRNA can repress production of hundreds of proteins by down-regulating mRNA levels or directly repress-

ing translation of hundreds of target mRNAs. Of the total number of computationally predicted targets, one-third were not detected in SILAC-based mass spectrometry data (Baek et al., 2008). This observation supports earlier (and somewhat controversial) prediction data identifying potentially numerous microRNA targets. Additionally, the false positives identified through experimental analyses will be of excellent value in developing more accurate prediction methods.

1.5 The role of microRNAs in animal development

What other role than the one of a genetic switch could have been attributed to *lin-4* and *let-7*, when the context of their discovery was a forward genetic screen? In their absence, transition of their de-repressed targets protein levels from inconsequential to "toxic" gave a mutant phenotype to be characterized. Had *lin-4* and *let-7* repressive efficacy not been this strong against lin-14 and lin-28 target sites respectively, then neither of the two microRNAs would have been discovered using such an approach.

With *lin-4* and *let-7* setting the initial paradigm of microRNA targeting interactions as being on/off "switches" with a role in animal development, follow up studies were up to surprises with most microRNA mutants exhibiting "moderate" defects. Still, these moderate defects appear to have a significant biological role which will be described in the following sub-sections.

1.5.1 The effect of global microRNA shut off

A straightforward experiment which gave insight into the role of microRNAs in early animal development involved the global depletion of all mature microRNAs. Generation of mutants for key enzymes of the microRNA bio genesis pathway such as Dicer or Loquacious, gave a panoramic view over the role of microRNAs during early development of both invertebrate and vertebrate model organisms.

In *C. elegans*, inactivity of both maternal and zygotic *dicer-1* (*dcr-1*) caused embryonic lethality (Grishok et al., 2001). *Drosophila* Loquacious mutants were viable but female sterile (Forstemann et al., 2005). Zebrafish maternal-zygotic *dicer1* mutants (*Mzdicer*) exhibited late embryonic lethality without severe impairement in patterning, axis formation and differentiation. Nevertheless, zebrafish morphogenesis, somitogenesis, heart and brain development were severely affected in the absence of microRNAs (Giraldez et al., 2005). In mouse, *Dicer1* mutants displayed severe defects as they lacked Oct4+ pluripotent stem cells and their development arrested early in development, during gastrulation, at stage E7.5 (Bernstein et al., 2003). Global microRNA depletion can therefore be em-

bryonic lethal in *C.elegans* and mouse in contrast to zebrafish where phenotypes were relatively mild. These studies demonstrated that the role of Dicer and microRNAs in early embryonic developmental patterning as well as germline stem cells can vary among different organisms. Further genetic studies on microRNA mutants came to shed some more light over their functional role in animal development.

1.5.2 Functions inferred from genetic studies of microRNA mutants

Several studies on microRNA mutants have implicated microRNAs in a range of biological processes such as developmental timing (Lee et al., 1993; Reinhart et al., 2000), signalling pathways (Hornstein et al., 2005; Li and Carthew, 2005; Flynt et al., 2007), circadian rythmicity (Cheng et al., 2007; Kadener et al., 2009a), myogenesis (Sokol and Ambros, 2005; Zhao et al., 2007; Chen et al., 2006), neurogenesis (Johnston et al., 2005; Schratt et al., 2006; Li et al., 2006; Karres et al., 2007; Makeyev et al., 2007), hematopoesis (Chen et al., 2004; Thai et al., 2007), energy homeostasis (Teleman et al., 2006), germline development (Iovino et al., 2009), maternal mRNA clearance during transition from maternal to zygotic transcription (Giraldez, 2006; Tang et al., 2007a; Bushati et al., 2008) and cell proliferation and apoptosis (Brennecke et al., 2003).

Judging from the number of mRNAs a single microRNA can target, it would be difficult to find a biological process in which microRNAs do not play a role. As a result, clustering microRNAs into groups based on the biological processes in which they are involved shall only produce a long list of diverse systems where microRNAs are involved without necessarily giving an insight into their mode of action as a regulatory class of molecules. A more conclusive categorization of microRNAs could use as a basis the type of interaction between microRNA and its target/s by considering the degree and consequences of dampening the target's protein levels as well as the extent of how often can this happens in the organismal/biological context.

For example, a microRNA might have a predicted target which *in vitro* gets very efficiently repressed but in reality, the two never "meet" as the microRNA is expressed in one cell type and the target in another. In fact there are studies which compared the expression sites of microRNAs and their predicted targets and could report a mutual exclusion between the two (Stark et al., 2005; Visvanathan et al., 2007). Given that microRNA and target never meet, the microRNA's role in relation to this particular target becomes equivalent to that of a fail-safe mechanism which protects the microRNA-expressing cell from leaky target's transcripts and the potential accumulation of target's protein up to consequential levels (left panels of Figure 1.3). In this type of interaction, the microRNA functions to reduce target's protein output down to inconsequential levels

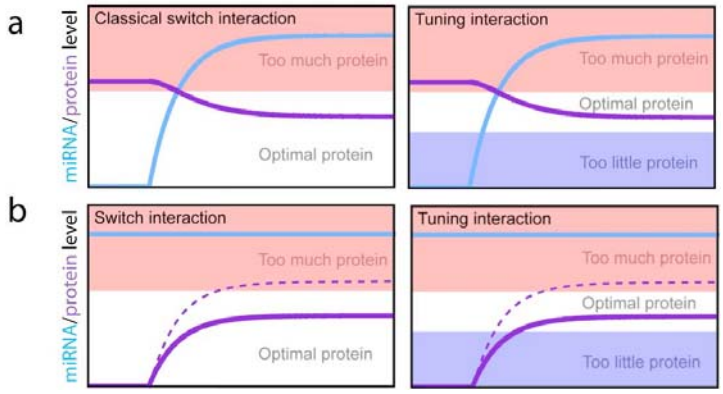


Figure 1.3: **Types of microRNA regulatory interactions. a,** a microRNA can act as a binary switch or may fine tune target protein's output at optimal levels in response to a developmental or environmental cue (microRNA expression level in blue and protein levels in purple). **b,** pre-existing microRNA expression in a cell/tissue may act to restrict target's protein production to optimal levels. Accumulation of target's protein levels in the absence of the microRNA is consequential for the cell in both switch and tuning type interactions (depicted by dashed purple line). For tuning type interactions, excessive microRNA repression can be as critical as the absence of microRNA because it bring the target's protein levels down to consequential. Figure modified after Bartel (2009).

(Bartel, 2009).

In the opposite scenario, where a microRNA and its predicted target are co-expressed in the same cell/tissue-as reported by (Farh, 2005; Shkumatava et al., 2009), a "tuning" role is suggested for the microRNA because a "binary off-switch" would render the target gene functionless in the microRNA-expressing cell (right panels in Figure 1.3). In tuning interactions, the microRNA functions to reduce protein levels to an optimal rather than off (consequential for the particular cell type) state (Karres et al., 2007; Bartel, 2009).

By classifying studied microRNAs based on their apparent switch or tuning role I will present a collection of key genetic studies which have contributed towards the better understanding of microRNAs functions and have underlined their involvement in a multitude of systems.

1.5.2.1 microRNAs with identified switch-type interactions

***lin-4*, a switch in developmental timing** Induction of *lin-4* and *let-7* occurs in a stepwise manner during *C.elegans* development and coincides with the time points when their targets must be repressed in order to allow transition to the next larval stage of development. More specifically, the early target of *lin-4* is a transcription factor essential

for the completion of the first larval stage L1 (Hristova et al., 2005). Worms deficient for transcription factor *lin-14* skip L1 and transit into L2, while gain of function mutants lacking *lin-14* 3' UTR go over L1 -specific lineages again (Ruvkun et al., 1989; Ambros and Horvitz, 1987). In wild type flies, *lin-4* becomes active during L2 transition and subsequently turns off the expression of pre-existing *lin-14* (Lee et al., 1993; Wightman et al., 1993).

The second target of *lin-4*, *lin-28* is yet again a heterochronic gene but important for the L2 to L3 transition. Accordingly, *lin-28* rescuing trangene with mutated *lin-4* binding sites causes developmental delays and specifically reiteration of L2 -specific lineages (Moss et al., 1997).

For both lin-14 and lin-28, their interaction with their targeting microRNA lin-4 is a paradigm of a switch interaction whereby the microRNA onset shuts down the expression of a pre-existing target.

***let-7*, a switch in developmental timing** *let-7* induction occurs in the late larval stages of *C.elegans* at the L4-adult transition when the TRIM protein levels encoded by *lin-41* must be repressed to allow the developmental transition (Reinhart et al., 2000). The late onset of *let-7* seems to be conserved across bilaterian animals (Pasquinelli et al., 2000) and recent experiments in *Drosophila* let-7 gene cluster, which additionally hosts miR-100 and the lin-4 ortholog miR-125, revealed a conserved role in developmental timing (Sokol et al., 2008; Caygill and Johnston, 2008).

In the wing imaginal disc of the fly, *let-7* and *miR-125* mediate cell-cycle exit just like *C.elegans let-7* promotes cell-cycle exit of hypodermal cells to allow transition into adulthood. In *let-7*, mutant wing discs persist their cell divisions 24h after puparium formation as opposed to ceased divisions in wild type. Thus the onset of *let-7* expression during puparium formation controls the period of wing disc cell divisions (Caygill and Johnston, 2008).

A parallel study uncovered a second role for *let-7* during *Drosophila* development which induces the destruction of larval specific dorsal internal oblique muscles (Sokol et al., 2008). During adult maturation these larval muscles get destroyed but in *let-7* mutants they persisted and neuromuscular junctions associated to adult-specific dorsal muscles appeared immature (Sokol et al., 2008; Caygill and Johnston, 2008). Intriguingly, a closer look in *let-7 C.elegans* mutants reveals equivalent "unfinished work" as the programmed cell death of the "linker cell", connector of the male reproductive system to the exterior, meant to happen during the L4-adult transition is skipped (Abraham et al., 2007).

The *Drosophila let-7* target is a transcription factor, *Abrupt* (*Ab*), whose protein levels

drop in synchrony with the up-regulation of *let-7*. Partial loss of *Ab* function "mildens" the retarded dorsal muscle and neuromuscular junction phenotype of *let-7* mutants, indicating a role for *Ab* in developmental timing by blocking adult fates until repressed by let-7 (Caygill and Johnston, 2008). Interestingly, expression of *Ab* and the *Drosophila lin-41* ortholog, brat, gets induced by 20E, the active form of ecdysone hormone, which peaks at the late third instar larvae and triggers puparium formation and adult differentiation, in other words, metamorphosis (Beckstead et al., 2005).

Just like *lin-4*, the onset of *let-7* is necessary for the down-regulation of pre-existing *lin-41* in *C.elegans* and the subsequent developmental transition this allows. The exciting finding with *let-7* is that this switch interaction reported for *C.elegans* is also identifiable in *Drosophila* (see above). Despite the fact that the *Drosophila* target is not conserved across evolution, the switch mode of action is the first example of a conserved target interaction reported for a microRNA.

***miR-9*, a fail-safe switch in PNS development** In *Drosophila*, during development of the peripheral nervous system (PNS) of the young larva, sensory organ precursors (SOPs) derive from epidermal cells in highly stereotypical numbers and positions. Expression of *miR-9* was detected in epithelial cells of the young embryo in a non-overlapping/mutually exclusive manner to its predicted and validate target *Senseless* (*sens*) (Li et al., 2006). The transcription factor *sens* is dynamically regulated within proneural clusters so as to be highly expressed in SOP cells (where it up-regulates and maintains proneural gene expression necessary for SOP specification) and kept at low levels in non-SOP adjacent cells which makes *sens* a repressor of proneural genes transcription (Nolo et al., 2000; Jafar-Nejad et al., 2003).

miR-9 mutants were viable and fertile indicating that the microRNA's role is not essential for survival, but the mutant flies displayed ectopic and more than average bristles in their notum and the anterior wing margin. These extra bristles were linked to extra emerging SOPs due to de-repression of *miR-9* target *sens* (Li et al., 2006). As opposed to the *lin-4* and *let-7* interactions with their targets which get temporally switched off, the interaction between *miR-9* and *sens* has a spatial basis. *miR-9* expression in all epithelial cells except SOP cells acts as a "fail-safe" mechanism whereby leaky *sens* originating from stochastic cell-to-cell noise gets switched off down to inconsequential levels for non-SOP cell fate (Cohen et al., 2006; Bartel, 2009).

The premise of a fail-safe switch interaction is therefore a mutually exclusive expression of microRNA and its target (Bartel, 2009). Many such examples of fail-safe switch interactions have been reported for *Drosophila* in a large in situ screen investigating the

localisation of microRNAs and their predicted targets (Stark et al., 2005). However, it is noteworthy that few of the target sites which show mutual exclusion to their targeting microRNAs are evolutionarily conserved (Bartel, 2009).

miR-430, a temporal switch in maternal-to-zygote transition Temporal switch interactions of microRNAs and their targets have also been described for vertebrates. In zebrafish, a large family of microRNAs (miR-430-family) which is highly abundant in the early fish zygote(Giraldez et al., 2005), accelerates the decay of a large set of maternally deposited mRNAs during the activation of zygotic transcription (Giraldez, 2006). MicroRNA-depleted fish embryos accumulate maternally deposited mRNAs which in wild type decline after maternal-zygotic transition (MZT) exactly when miR-430 family starts to accumulate and deadenylates its targets (Giraldez, 2006).

In this example of temporal mutual exclusion, miR-430-family facilitates the temporally sharp MZT by clearing pre-existing maternal transcripts. Similar temporal switch interaction as that of miR-430-family in fish, has been reported to facilitate the MZT of *Drosophila*, involving however another microRNA family (miR-309) which is not evolutionarily conserved outside the insects lineage (Bushati et al., 2008).

1.5.2.2 microRNAs as fine tuners of developmental programs

***miR-8*, a fine tuner which prevents neurodegeneration** Despite the extensive and complex expression pattern miR-8 exhibits in *Drosophila* embryos and larvae, miR-8 mutants manage to complete larval development but show reduced survival rate in pupal and early adult stages (Karres et al., 2007). The surviving mutants exhibit morphological malformations and behavioral defects which were shown to be a consequence of extensive CNS apoptosis stemming from elevated levels of the atrophin gene product. miR-8, was shown to target atrophin with 4 binding sites at the mRNA's 3' UTR, two of which are conserved from *Drosophila* to human as of the very few deep evolutionarily conserved targets (Karres et al., 2007; Chen and Rajewsky, 2006a).

The co-expression of miR-8 with its widely expressed target atrophin in the CNS suggested a tuning role for this microRNA which was experimentally challenged. Surprisingly complete elimination of atrophin in miR-8+ cells (such that could mimic a switch hypothetical role of miR-8), gave a phenotype identical to what was previously described for clones of wing cells mutant for atrophin (Zhang et al., 2002; Erkner et al., 2002). It was therefore clear that miR-8+ cells require atrophin function and thereby the microRNA does not act like a switch (Karres et al., 2007). This work has set a clear example whereby a microRNA can have a tuning role to maintain a given co-expressed

target's protein output at favorable levels because either excess or less than favorable can compromise cell viability. Accordingly, the premise of this type of tuning interaction between microRNA and target is their co-expression (Bartel, 2009).

miR-7, a tuner which stabilizes developmental programs against environmental perturbations Recent work has highlighted the involvement of miR-7 in *Drosophila* photo- and proprioreceptor determination during the course of development (Li et al., 2009). miR-7 participates in two coherent feedforward loops and targets YAN transcription repressor within the network which controls photoreceptor determination (Li and Carthew, 2005; Li et al., 2009). In this network, the role of miR-7 is to stabilize the developmental decision after EGF signalling which leads to YAN protein degradation by reinforcing Yan down-regulation and ensuring a stable change of cell fate from progenitor to photoreceptor (Li and Carthew, 2005; Bushati and Cohen, 2007). In the network controlling SOP determination, miR-7 participates by targeting Enhancer of Split (E(spl)) in an incoherent feedforward loop as *Atonal* (*Ato*) induces both miR-7 and E(spl) with the latter repressing *Ato* (Li et al., 2009).

The exciting finding over the role of miR-7 in these two gene networks was only made once each developmental process was challenged by environmental perturbations. Although Ato expression appears normal in miR-7 mutants, a temperature perturbation in the environment led to a strong decrease of Ato levels. A concominant reduction in the number of SOPs was observed. In contrast, such environmental perturbation did not affect Ato levels in wild type flies. Therefore in the absence of miR-7, E(spl) is no longer suppressed and can then itself repress atonal uncontrollably (Li et al., 2009). A similar picture emerged upon temperature perturbation of the photoreceptor gene network, with Yan expression being abnormally high and irregular eyes forming in miR-7 mutants (Li et al., 2009).

This pioneering work uncovered the biological importance of a fine-tuner microRNA in a realistic experimental design which took into account the environmental context and its effects on gene expression (which can get stochastic upon perturbations). Further studies of this kind might discover more microRNAs which fine tune target levels to keep developmental programs stable and robust in conditions of environmental influx.

miR-375 a fine tuner of insulin secretion in mice In mice, miR-375 expression is restricted to MIN6, TC1 cells and the pancreatic islets and has been shown to control insulin secretion independent from Ca2+ signalling or glucose metabolism (Poy et al., 2004). Over-expression of miR-375 in MIN6 cells was shown to cause a 40% reduction of

insulin secretion. The same effect was reported upon silencing of the miR-375 predicted target, Myotrophin (Mtpn), which is co-expressed in MIN6 pancreatic cells. The regulation of Mtpn by miR-375 in pancreatic cells was validated and seems to be another example of tuning interaction between microRNA and target whereby Mtpn levels are kept constant by miR-375 for functional insulin secretion (Poy et al., 2004; Bartel, 2009).

1.5.3 Roles inferred from large-scale studies

Many important insights into microRNA target recognition and function come from large-scale studies of microRNA regulation. Such studies involve computational analysis of microRNA binding site conservation or depletion among predicted targets (Lewis et al., 2005; Brennecke et al., 2005; Farh, 2005; Stark et al., 2005; Grimson et al., 2007; Nielsen et al., 2007), followed by profiling of target mRNA expression to find out which targets overlap in expression with their targeting microRNA (Stark et al., 2005; Farh, 2005; Sood et al., 2006; Shkumatava et al., 2009) or experimental identification of targets which respond to changes in microRNA regulation either at the mRNA (Lim et al., 2005; Giraldez, 2006; Easow et al., 2007; Karginov et al., 2007) or protein level(Selbach et al., 2008; Baek et al., 2008). The advantage with large-scale studies is in that they offer a panoramic view on a microRNA's interaction with all its predicted targets.

The two first large-scale analyses compared the expression of microRNAs with that of their predicted targets. A large *in situ* screen performed in *Drosophila* for the identification of the spatial localisation of both microRNAs and their predicted targets during development, showed that microRNAs tend to target mRNAs which are not expressed in the same tissue (Stark et al., 2005). In the second study, microarray data from mammals gave a similar picture with mutual exclusion between the expression of microRNAs and their targets but only the targets with non-conserved binding sites (Farh, 2005). Mammalian targets with conserved binding sites were found co-expressed with their targeting microRNA albeit at lower levels (Farh, 2005; Sood et al., 2006).

These conflicting results were attributed to the poor dynamic range of *in situ* data for those who favored Farh (2005) and to poor cellular resolution of the arrays for those who favored Stark et al. (2005). The latter study was rather influential in the microRNA community and inspired the adaptation of the canalization theory(Waddington, 1942) to suggest a role of animal microRNAs in the 'canalization' of development (Hornstein and Shomron, 2006). It was therefore suggested that by clearing unwanted transcripts from preexisting mRNAs during developmental transitions (Bushati and Cohen, 2007), or by repressing 'leaky' transcripts that should be expressed in neighbouring tissues but not in the cells where the microRNA is expressed(Stark et al., 2005), microRNAs provide

a fail-safe mechanism, ensure accuracy and confer robustness to tissue differentiation (Hornstein and Shomron, 2006).

However, latest studies which analyzed the molecular profile of microRNA-expressing cells purified to cellular resolution support the observation from earlier arrays in that expression of conserved targets tends to overlap with that of their targeting microRNA, and expression of messages with nonconserved target sites is less prone to overlap (Shkumatava et al., 2009). The confirmed observation for nonconserved target sites supports the idea of anti-targeting. As 3' UTRs accumulate mutations in their sequences during the course of evolution, microRNA binding sites continuously "sample" matches to co-expressed microRNAs. Depending on whether the resulting microRNA-mediated repression is selectively favorable, neutral or "toxic" to the cell the newly emerged target site will accordingly be conserved, neutral or selectively avoided (Bartel, 2009). In anti-targeting, nonconserved target sites are lost during the course of evolution from mRNAs highly expressed in cells expressing their targeting microRNA because they are toxic. On the contrary, these sites can persist in mRNAs expressed in cells that do not express the microRNA (Bartel, 2009).

Recent large-scale proteomic analyses studying the influence of microRNAs on the protein output of their targets revealed that the repressive effect on individual proteins was modest, with some exceptional proteins being repressed down to 50%-80% (Baek et al., 2008; Selbach et al., 2008). Taking these latest results into account, a picture emerges whereby most microRNA-target interactions appear to be of the tuning type, at least for conserved and hence important target sites (Karres et al., 2007; Baek et al., 2008), and switch type interactions (sometimes acting as extreme as fail-safe) are seen more often for nonconserved target sites (Li et al., 2006; Giraldez, 2006; Bartel, 2009). Still, with so many conserved targets for each microRNA and considering all the biological roles they may be involved in, the proportions of tuning, switch or fail-safe interactions might vary from one microRNA to another (Bartel, 2009). A combination of more large scale studies followed up by single protein analyses will give a clearer impression on the role of microRNAs in animal development.

1.5.4 MicroRNAs ensure precision of protein output

With the current state of the art, a role of microRNAs as stabilizers and "micromanagers" of protein output can easily be inferred. Their involvement in rendering developmental decisions and programs more robust has been demonstrated in a number of different organisms. With special emphasis on latest reports which identified microRNAs buffering complex gene regulatory networks (Li and Carthew, 2005; Li et al., 2009; Hammell et al.,

2009), their role in conferring robustness to gene networks and developmental programs can hardly be argued.

Considering the importance of gene networks in development and evolution (Davidson, 2006) together with the inherent noise of gene expression (Elowitz et al., 2002; Raser, 2005) raises an antithesis: How can the output of gene networks with intrinsic noise exhibit minimal phenotypic consequences? Phylogenetically basal and statigraphically "early" species show high variability in polymorphic traits (Webster, 2007; Hughes, 2007), underlining the evolutionary implications of genetic noise.

Interestingly, a cross species comparison on the expression levels of mRNA orthologs (between fly, mouse, chimpanzee and human) in a number of homologous tissues pointed out a role for microRNAs in reducing noise. MicroRNA targets showed similar levels of expression across species as opposed to mRNAs not targeted by microRNAs which showed significant variations in their levels (Cui et al., 2007). Therefore, an effective way to reduce genetic noise during the course of evolution is by introducing a post-transcriptional layer of regulation like microRNAs. In light of recent proposals whereby microRNAs are suggested to be instrumental for the evolution of metazoa (Peterson et al., 2009), I shall dedicate the next chapter to present current hypotheses on how microRNAs evolved, their phylogenetic distribution across metazoa and hypotheses that stem from these data.

1.6 Evolution, diversification and phylogeny of animal microRNAs

1.6.1 microRNA gene birth during the course of evolution

1.6.1.1 De novo microRNA gene emergence

Large scale analyses of genomic transcription have revealed that virtually the entire non-repeat portion of animal genomes gets transcribed into RNA (Kapranov et al., 2007; Affymetrix, 2009). Such broad transcription of the genome could assist the birth of novel microRNA genes as computational approaches predict hundreds of thousands candidate hairpins which could get processed by Drosha/Pasha (Liu et al., 2008). In *Drosophila* genome, 100,000 candidate hairpins get predicted (Lu et al., 2008b) while in mammalian genomes predictions reach 1,000,000 of such hairpins (Bentwich et al., 2005).

The above facts together with recent evidence that Drosha/Pasha microprocessor complex does not strictly process stem loops from microRNA primary transcripts but also from random transcripts (Kadener et al., 2009b) assigns big evolutionary implications in

this large availability of hairpins that genomes offer. Across evolutionary time, the microprocessor complex could sample a wide range of such incidental hairpins which upon maturity adopt a set of targets (Figure 1.4c). This is probably what happened with miR-220 whose sequence encodes for a tubulin gene which was most likely competent for Drosha/Pasha processing after antisense strand transcription and stem loop availability (Shomron et al., 2009).

Depending on a microRNA's beneficial or toxic effect on the newly acquired set of targets, "new-born" microRNA genes can be positively selected and fixed in the genome or not (Chen and Rajewsky, 2007; Lu et al., 2008a).

1.6.1.2 Transposon-assisted microRNA gene birth

Transposons and their no-longer functional remnants can be found widely interspersed in eukaryotic genomes (Britten and Kohne, 1968). With the ability to translocate anywhere in the genome, transposable elements (TE) can mutate genes by jumping into their open reading frame (ORF) or by disrupting their cis-regulatory elements. Flanking regulatory elements sequences carried by TEs can even alter the expression of the gene which newly hosts them. Lastly, TEs can incorporate into host genes' introns, exons and UTRs (Liu et al., 2008).

TEs often carry inverted repeats and the extent of their palindromic sequence maybe such that allows intramolecular folding into a hairpin loop (as seen with miniature inverted-repeat transposable element -MITEs) eligible for processing by Drosha/Pasha (Piriyapongsa and Jordan, 2007; Feschotte, 2008). Even TEs without inverted repeats can lead to the formation of a hairpin as they often jump proximally to other TEs resulting in repeat arrangements prone to secondary structure formation (Liu et al., 2008). In mammals, the microRNAs which derive from LINE transposable elements or other repetitive elements are not few (Smalheiser and Torvik, 2005), with at least 55 human microRNAs predicted to derive from LINE, SINE elements or MITEs and 85 from novel TE-derived candidates (Piriyapongsa et al., 2006).

A fascinating speculation about transposon-assisted microRNA birth concerns the potential assembly of a microRNA network from one TE family. Given that a "newly born" microRNA derives from the same TE family which has also been inserted in the 3'UTRs of a number of genes, the newly born microRNA immediately encompasses a set of targets with complementary TE sequences in their 3' UTRs (Feschotte, 2008). Such a scenario remains to be experimentally tested.

1.6.2 microRNA diversification upon fixation

As opposed to protein coding genes which were in some cases lost extensively in different (derived) evolutionary lineages (Raible et al., 2005), microRNAs rarely get lost from a genome once they get fixed (Heimberg et al., 2008). During the course of evolution, a microRNA's sequence and transcriptional output may diversify in a number of ways listed in the folowing sub-sections.

1.6.2.1 microRNA gene duplication and cluster formation

Just like protein coding genes (in which microRNAs are often hosted), microRNAs can be subject to local (tandem) gene duplication (Ohno, 1999) or whole genome duplication events (GDE) as reported in vertebrates (Hertel et al., 2006). Tandem duplications of a microRNA gene followed by mutational accumulation (Zhang et al., 2007) can give rise to clusters of paralogous microRNAs which form a microRNA family if mutational accumulation does not alter the seed 7mer (Chen and Rajewsky, 2007; Shabalina and Koonin, 2008).

Segment duplications of entire clusters by GDEs followed by species specific loss of some microRNAs, gain of new microRNAs through subsequent tandem duplications or/and enhanced substitution rates in some paralogs (Tanzer and Stadler, 2004; Hertel et al., 2006; Zhang et al., 2007), can give rise to paralogous clusters which further subfunctionalize. Depending on the new paralogous cluster's genomic position and regulatory elements which control its expression, microRNAs of the cluster can be deployed in distinct biological systems (Zhang et al., 2009).

The importance of tandem and segment duplications in microRNA evolution is not negligible. A large fraction of microRNA genes appear clustered, forming uninterrupted arrays in intergenic regions of the genome with possible operon like organization (Lai et al., 2003). A comparative study on four vertebrate species (Human, mouse, rat, chicken), identified ~30% of microRNA genes clustered (Megraw et al., 2007) while in *Drosophila* the fraction goes up to ~50% (Bartel, 2004). Clustered microRNAs often get co-transcribed as long polycistronic transcripts (Lee et al., 2002; Mourelatos, 2002; Lai et al., 2003) and have been found co-expressed with neighbouring genes (Baskerville and Bartel, 2005; Wang et al., 2009).

For duplicated genes to be fixed over evolutionary time, distinct functions should be acquired compared to their paralog, otherwise mutational accumulation can render the paralog nonfunctional and subsequently a pseudogene (Lynch and Conery, 2000; Prince and Pickett, 2002). In the following subsections I present the evolutionary modes

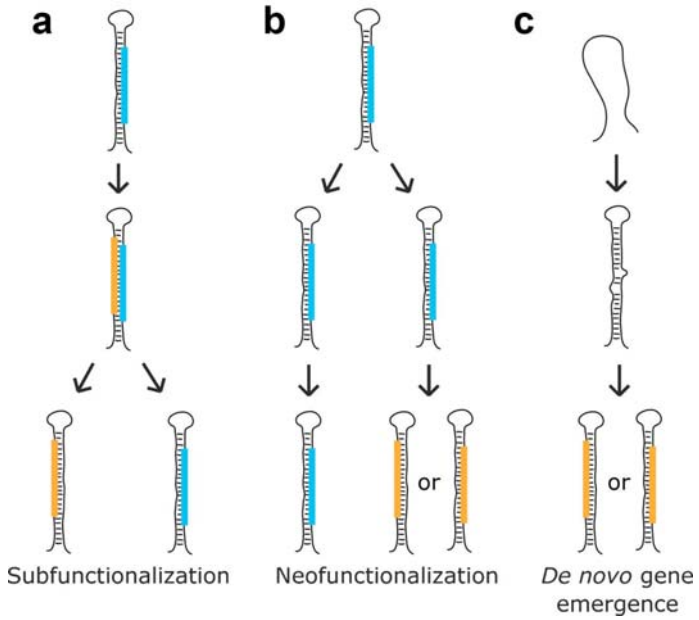


Figure 1.4: **Scenarios of microRNA evolution. a,** subfunctionalization can occur when a miRNA* (orange) acquires function following microRNA gene duplication or when the daughter copy becomes subject to heterologous processing (while the other daughter copy does not). **b,** neofunctionalization can happen when a daughter copy of the duplicated microRNA gene accumulates mutations which confer novel functionality to the microRNA. **c,** incidental hairpins which occur in the genome are subject to Drosha/Dicer processing and if their product is selectively maintained then de novo microRNA genes may emerge. Adapted figure by Ruby et al. (2007).

through which paralogous microRNAs can "survive" by acquiring novel functions upon their emergence in the genome.

1.6.2.2 Subfunctionalization

Subfunctionalization is a process that may follow gene duplication, whereby each daughter gene can specialize unique functions so as to complement each other and additively reach the functional output of their ancestral gene (Lynch and Conery, 2000). Such an evolutionary process would require the ancestral gene to have multiple functions and several mechanisms of the microRNA pathway can assign multiple functions to an "ancestor" microRNA hairpin (Ruby et al., 2007). Upon duplication of a microRNA hairpin, "division of labor" between the two arms among the two copies shall complement the ancestral hairpin's function (Figure 1.4a).

Heterologous processing by Drosha or Dicer As mentioned earlier in introduction on page 21, imprecise processing by Drosha and Dicer can occur (Ruby et al., 2007; Landgraf et al., 2007). The biological implications of such an event are subject to natural selection because a frameshift in the seed region of a microRNA immediately introduces a radical new set of target interactions which could be beneficial or toxic for the cell.

Close examination of large microRNA sequencing datasets revealed that some seemingly identical microRNA loci give rise to mature microRNAs with entirely distinct seeds because of heterologous processing. For example, among the "K box" microRNAs which constitute the largest family of microRNA paralogs in *Drosophila* with four genomic clusters (Lai et al., 2003), the previously considered identical miR-2a-1 and miR-2a-2 were found to have a 2nt offset in their 5' ends and seeds (Ruby et al., 2007; Liu et al., 2008).

Thus, mutational accumulation is not necessarily the only way to subfunctionalize a microRNA copy, as stochastic processing mechanisms alone can generate subfunctional microRNA paralogs.

Bringing the microRNA* into service A readily available mechanism for microRNA subfunctionalization relies on the bipartite nature of the microRNA stem loop. Although at a generally less frequency, microRNA* species are loaded into miRISC and regulate target sets of their own (Okamura et al., 2008). Some microRNA* species even show evolutionary conservation across bilateria further underlining their functional importance (see results Table 2.1 andWheeler et al., 2009).

Following a microRNA gene duplication, one daughter microRNA hairpin can maintain the original function of the ancestor while the other daugther microRNA hairpin can maintain the function of microRNA* (Ruby et al., 2007). In this way acquisition of microRNA* functionality allows subfunctionalization.

1.6.2.3 Neofunctionalization

Neofunctionalization is a process that may follow gene duplication, whereby one of the daughter genes mutates into a selectable function which was not present in the ancestral gene (Rastogi and Liberles, 2005). Although mutational accumulation can be a driving force towards neofunctionalization of microRNA genes, post-transcriptional RNA editing poses a second mode of action.

Changes in microRNA sequence Alterations in a microRNA's sequence certainly affect its target recognition and binding capacity (Brennecke et al., 2005). Therefore, mutational accumulation in newly emerged microRNA paralogs can directly alter their target

groups, especially when mutations accumulate in the seed region (Figure 1.4b). If newly emerged target interactions are neutral or beneficial for the microRNA expressing cell, then the neofunctionalized microRNA can be fixed in the genome (Liu et al., 2008). This is indeed what is commonly seen in many microRNA families, members of which are sequence related with some mismatches (Bentwich et al., 2005; Ruby et al., 2006; Berezikov et al., 2006; Ruby et al., 2007). The sequence of founding genes of these families are often evolutionarily conserved and each evolutionary lineage displays its own series of related family members (Wheeler et al., 2009).

RNA editing As discussed earlier in introduction (on page 20) RNA editing on primary-microRNA transcripts can have a big impact on target recognition and binding (Ohman, 2007). A newly acquired set of targets, renders novel functions to the edited mature microRNA, distinct from its ancestor's, and can therefore be a driving force of neofunctionalization. A simple comparison of the genomic and cloned microRNA sequences can highlight potential editing events like for miR-376 which varies by one nucleotide in the middle of the seed when compared to the genomic sequence. Indeed, human and mouse miR-376 was experimentally shown to get postranscriptionally edited by adenosine deaminases acting on RNA (ADARs) (Kawahara et al., 2007).

1.6.2.4 Transcriptional control over newly emerged microRNAs

Knowing that a single microRNA can repress the protein output of hundreds of targets and having in mind several examples of antitargeting, a paradox emerges when considering microRNA evolution: How can a microRNA arise in a genome without seriously impairing the fitness of the organism? Simply by chance, some of the target interactions will have a deleterious effect (Chen and Rajewsky, 2007).

A proposed mechanism is through tight transcriptional control of newly emerging microRNAs (Chen and Rajewsky, 2007). An overall weak transcriptional output, spatial restriction of transcription in a cell type or temporal restriction of transcription in a particular developmental stage could allow natural selection to eliminate deleterious target sites over time (Chen and Rajewsky, 2007). This hypothetical scenario becomes more plausible after comparing expression data between recently acquired human microRNAs which are overall weakly expressed to strongly expressed ancient conserved microRNAs (Baskerville and Bartel, 2005; Berezikov et al., 2006).

This hierarchical component in expression levels of a microRNA according to its "acquisition age" is illustrated with the example of two microRNAs with a role in heart development. miR-1, an ancient microRNA already present in Urbilateria (Wheeler et

al., 2009) is strongly expressed in the heart (and overall musculature) together with miR-208, a heart specific microRNA which arose in the vertebrate lineage (Heimberg et al., 2008) and is therefore "younger". Expression of miR-1 in the heart is not only stronger than that of miR-208 but elimination of miR-1 gives a much more severe cardiac phenotype (Zhao et al., 2007) when compared to the mild miR-208 knockout phenotype (van Rooij et al., 2007). Such examples support the notion that transcriptional control over newly emerged microRNAs can allow their gradual functional maturation over evolutionary time.

1.6.3 The phylogenetic distribution of microRNAs across bilateria

MicroRNAs have been available to regulate gene expression since at least very early in animal evolution (Grimson et al., 2008). *Monosiga brevicollis*, a unicellular choanoflagelate, which poses the closest sister group to metazoa (King et al., 2008), lacks Drosha/Pasha and accordingly microRNAs (Grimson et al., 2008). The earliest branching lineage of metazoa in which microRNAs could be detected is that of sponges, with *Amphimedon queenslandica* possessing at least 8 microRNAs (Figure 1.6;Grimson et al., 2008). As closer relatives to bilateria, cnidaria were also surveyed in search for microRNAs and their repertoire contained 40, a higher number than what was identified in simpler sponge, yet smaller than what is known for the more complex bilateria (Grimson et al., 2008). It is noteworthy that no sequence homology is shared between any of these "early" metazoan microRNAs and bilaterian microRNAs with only one exception, miR-100, found in the cnidarian *Nematostella vectensis* (Grimson et al., 2008).

1.6.3.1 Bilaterian microRNA repertoires and animal complexity

The above findings complemented a series of studies which had put forward the idea of linking the size of microRNA repertoire to animal complexity (Hertel et al., 2006; Sempere et al., 2006; Prochnik et al., 2007; Heimberg et al., 2008; Wheeler et al., 2009). While the absolute number of transcription factors did not rise substantially during metazoan evolution (Putnam et al., 2007), the number of microRNAs identified in animals that are morphologically simple (such as sponges and cnidarians) appears considerably lower than those reported for more complex animals, such as nematodes, insects or vertebrates (Figure 1.5; Grimson et al., 2008; Heimberg et al., 2008).Very simple metazoans do not possess microRNAs at all (Grimson et al., 2008; Wheeler et al., 2009).

Recent comparisons of microRNA inventories revealed that at least one microRNA, miR-100, must have been present in the last common ancestor of bilaterians and cnidar-

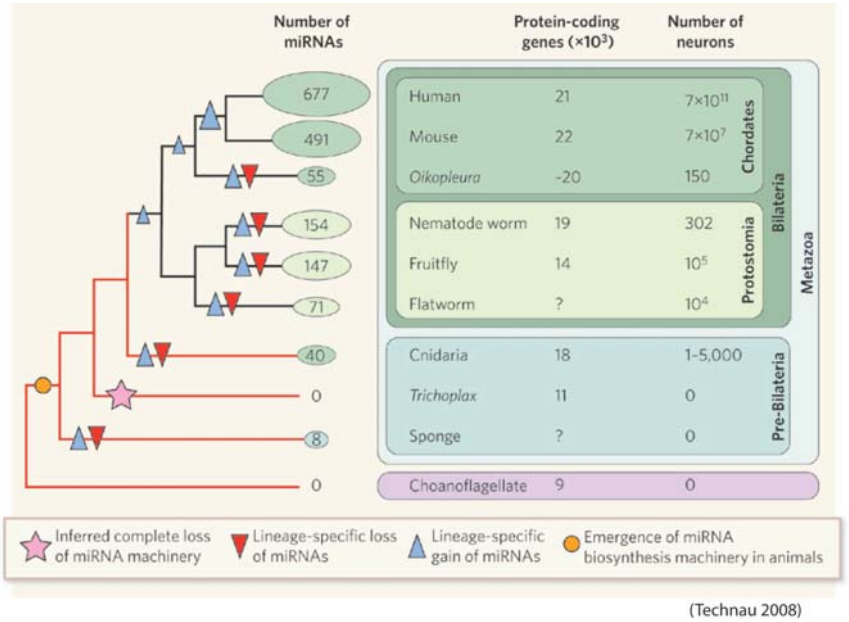


Figure 1.5: **The size of microRNA repertoires grows together with animal complexity.** While the number of protein coding genes does not rise substantially in line with escalating morphological complexity, microRNA number correlate with the organism's total number of neurons. Adapted after Technau (2008).

ians (plus an unknown number of non-conserved ones). This compares to 34 conserved bilaterian microRNAs (Table 2.2;Wheeler et al., 2009) that necessarily existed in the last common ancestor of protostomes and deuterostomes (which represent the two major superphyla of the bilaterians including insects, nematodes, annelids and sea urchins, vertebrates, respectively, among others). From that evolutionary node onwards the sizes of microRNA repertoires increase proportionally to organismal complexity as illustrated in Figure 1.5, 1.6.

1.6.3.2 The rate of microRNA acquisition: a second link to morphological complexity

Aside from the stem line of Bilateria, other important lineages which subsequently emerged within bilateria (such as the vertebrate and mammalian lineages; Figure 1.6) were also accompanied by new acquisitions of microRNAs (Heimberg et al., 2008; Wheeler et al., 2009). Of special interest concerning the acquisition of new microRNA families at important evolutionary nodes of metazoan evolution, is the temporal context . It appears that in all lineages which were accompanied by dramatic increases in morpho-

logical complexity (such as the stem line of Bilateria, the vertebrate and mammalian lineages), the rate (not only the number) of microRNA acquisition was anomalously high (Heimberg et al., 2008; Peterson et al., 2009). Originally, the large microRNA family expansions observed for vertebrates were associated to whole genome duplication events (GDE) (Hertel et al., 2006). However, subsequent studies of the microRNA repertoire of basal vertebrates which preceded GDE such as lamprey, identified verterbrate specific microRNAs occurring in single copies in laprey genome therefore supporting that vertebrate microRNAs were acquired prior to vertebrate GDE (Heimberg et al., 2008).

In the time frame during which vertebrates acquired 40 new microRNAs, crustaceans, annelids and echinoderms acquired no more than 10 novel microRNAs. Similarly, in the time needed for primates to acquire 84 new microRNAs, rodents gained just 16 new microRNAs (see Figure 1.6; Peterson et al., 2009). To comprehend the immense rate of microRNA acquisition in primate evolutionary history, one just has to compare the total repertoire of cephalochordate *Amphioxus floridae* (84 microRNAs) which comparable to primates, took 10 times longer to reach its final number (Peterson et al., 2009).

1.6.3.3 microRNAs and cambrian explosion

Approximately 600 Mya, a spectacular escalation in animal complexity took place in the stem line of the Bilateria, known as the "cambrian explosion". Body plans which predated those of the cambrian explosion are poorly documented in the precambrian fossil record with very few "accepted" fossils of sponges (Love et al., 2009) and *Kimberella*, a mollusc-like bilaterian fossilized organism from the late pre-cambrian (Fedonkin and Waggoner, 1997). This early sudden morphological disparity combined with the temporal asymmetry of emerged complexity has posed two great mysteries for evolutionary biologists: first, on what grounds was this complexity built upon? In other words how complex was the last common ancestor of bilaterian animals? Second, which was the driving force towards the vast morphological disparity seen in emerged phyla?

Intriguingly, the cambrian explosion coincides with a strong increase in microRNA number (Figure 1.5, 1.6; Wheeler et al., 2009). To further explore the putative link between the evolution of stem bilaterians and microRNAs, I set out to determine the ancient sites of activity of conserved microRNA families in a comparative approach. Is there a link between the birth of a given pan-bilaterian microRNA and specific cell types, tissues or organs? And if yes, what do we learn about the evolution of the bilaterian body plan?

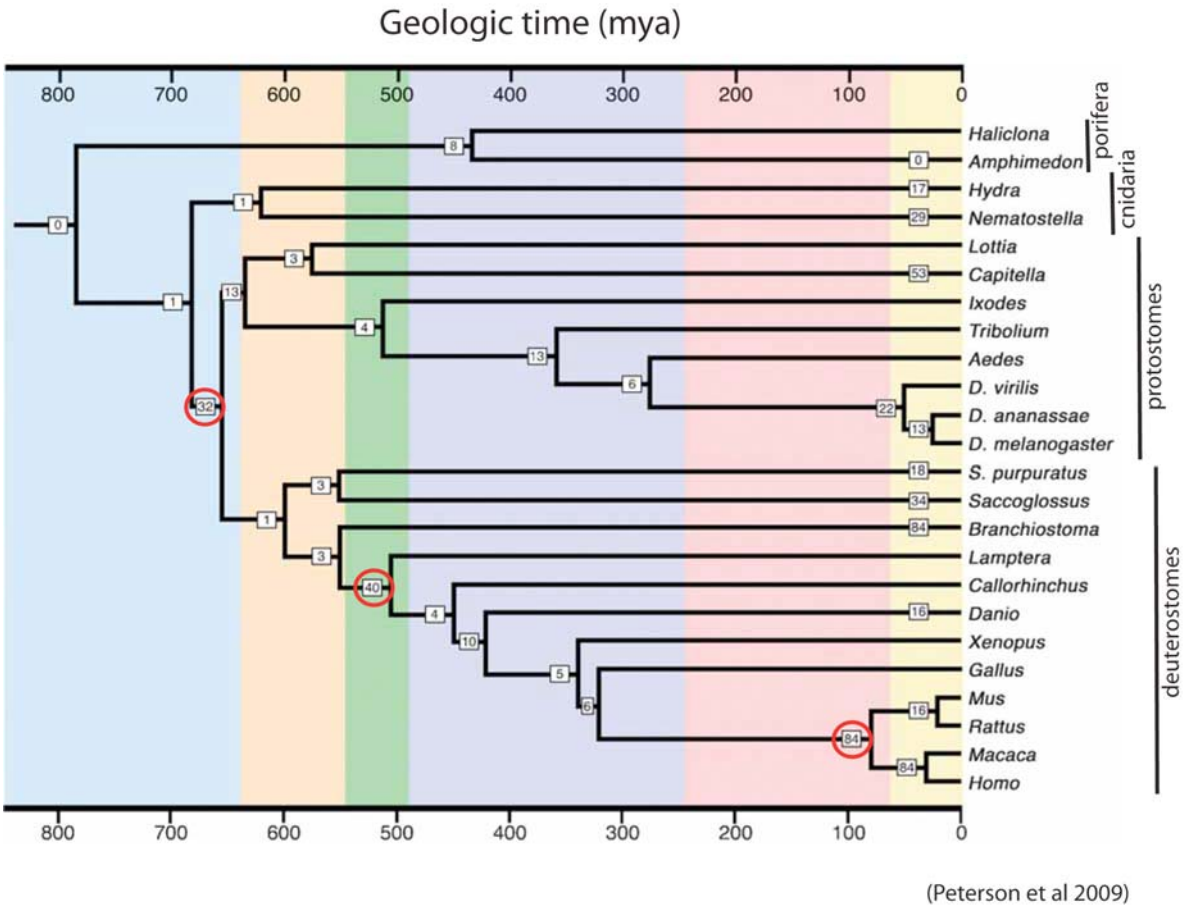


Figure 1.6: **MicroRNA acquisition over geologic time for 24 metazoan taxa.** Each evolutionary node is labelled by the number of miRNAs acquired at the time of its emergence (losses not considered). All eumetazoan lineages have acquired at least one novel microRNA but there are three instances of very high rate of microRNA acquisition: at the base of the protostomes and deuterostomes, at the base of the vertebrates, and at the base of primates (circled in red). Figure adapted after Peterson et al. (2009).

1.7 A comparative study to reveal the ancient site of microRNA activity

Comparative approaches, together with the fossil record, have served a great deal in reconstructing the historical course of animal evolution. By identifying homologous structures across metazoa, comparative studies can decipher common origin from a precursor structure which emerged once and persisted, to different extent, in distinct evolutionary lineages (Arendt, 2005).

1.7.1 Comparing cell types molecular fingerprints in search for homology

The level of homology which is chosen for a comparative study, can greatly affect the end result in terms of how exhaustive it allows the search to be. For example, classical neuroanatomists have deciphered homology for major brain subdivisions such as prosencephalon, mesencephalon and rhombencephalon within vertebrates but the complexity of these brain parts makes any comparison to invertebrate brains hardly possible (Arendt and Nubler-Jung, 1999; Nielsen, 2001; Lacalli, 2003; Reichert and Simeone, 2001; Velasco et al., 2004). Limitations of this kind are overcome with the growing use of cellular characteristics and their molecular profiles as a reference of homology in comparative studies (Arendt, 2008).

A cell's unique combination of genes contributing to its differentiation as well as the regulatory genes which turn on the latter, make up a cell type specific signature (composed of the differentiation and regulatory signatures) equivalent to a molecular fingerprint (Arendt, 2005, 2008). By comparing cell types using their unique molecular fingerprints, homology can be identified even when comparing lineages across long evolutionary distances (Arendt et al., 2004; Tessmar-Raible et al., 2007; Denes et al., 2007).

Just like genes, microRNAs also form part of cell types molecular fingerprints and more specifically they are part of the regulatory signature of cell types (Arendt, 2008). Therefore, specific localisation of a microRNA in a cell type shared between protostomes and deuterostomes should likely reflect the ancient specificity of that microRNA in their last common ancestor (Arendt, 2005). For example, shared restricted expression of miR-1 and miR- 124 in musculature and central nervous system, respectively, in fly and vertebrate reflects conserved roles in the differentiation of these tissues (Aboobaker et al., 2005; Bushati and Cohen, 2007; Kapsimali et al., 2007).

1.7.2 MicroRNA expression studies in bilateria

From the early days of microRNA discovery, upon the realization that some microRNA families are conserved across bilateria, the detection of their site of expression during animal development became top priority. Resolution on microRNAs temporal and spatial localisation gradually increased through the contribution of multiple groups, with expression profiles being originally based on cloning and northern blotting (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001; Aravin et al., 2003; Sempere et al., 2004; Landgraf et al., 2007), microRNA array profiling (Miska et al., 2004; Farh, 2005; Sood et al., 2006; Tang et al., 2007b; Bak et al., 2008), and eventually with the revolutionary application of locked nucleic acid (LNA) oligo probes (Valoczi et al., 2004) whole mount in situ hybridization (Aboobaker et al., 2005; Stark et al., 2005; Wienholds et al., 2005; Ason et al., 2006; Gonzalez-Estevez et al., 2009) as well as *in situ* hybridization on tissue sections (Deo et al., 2006; Kapsimali et al., 2007; Sacheli et al., 2009; Pena et al., 2009).

A vast and very detailed dataset is by now available based on high throughput profiling, revealing the localisation of microRNAs in many vertebrate tissues and cell types (Sempere et al., 2004; Berezikov et al., 2006; Landgraf et al., 2007). However, the degree of detail in vertebrate tissue annotation is often too elaborate to directly compare these data to expression profiles generated from invertebrates, where annotation is based on simpler invertebrate anatomy. This "loss in translation" renders the wealth of profiling data comparatively incompetent to decipher the ancient site of microRNA expression.

In addition, expression comparison between protostomes and deuterostomes has so far been hampered by the fact that in the fast-evolving protostome species *Drosophila* and *Caenorhabditis* microRNAs have acquired distinct localisation patterns concomitant with functional diversification (Liu et al., 2008). This makes it difficult or even impossible to infer ancient sites of microRNA activity from the comparison of expression patterns for most of the conserved bilaterian microRNAs. A similar tendency is apparent in the vertebrates where timing and location of microRNA expression may differ between fish, chicken and mouse (Ason et al., 2006).

What was missing, was a cornerstone dataset which could bridge the information available from protostomes and deuterostomes. For this comparative study, I investigated microRNA expression in the developing marine annelid *Platynereis dumerilii*. In contrast to *Drosophila*, *Caenorhabditis*, or the recently investigated planarian *Schmidtea* (Gonzalez-Estevez et al., 2009), *Platynereis* represents a slow-evolving protostome (Raible et al., 2005) of special value for comparative studies (Denes et al., 2007; Tessmar-Raible et al., 2007), more likely to retain ancient microRNA localisation patterns. In the following

chapter, I shall explain in more detail the advantages of using *Platynereis* as a model organism.

1.7.3 Using *Platynereis* as a model system

1.7.3.1 Phylogenetic position

The polychaete worm *Platynereis dumerilii* is a marine annelid. Its phylogenetic position in the tree of bilateria lies in the tree's third biggest branch, that of lophotrochozoa (Figure 1.7). Except from annelids (Tessmar-Raible and Arendt, 2003), this branch comprises many other marine invertebrate animals such molluscs, and planaria which have only recently been introduced as model systems in molecular studies. Together with the branch of ecdysozoa which comprises all insects, crustaceans and nematodes (Aguinaldo et al., 1997), they comprise the phylum of protostomes (Philippe et al., 2005). In the phylum of deuterostomes, both invertebrates such as echinoderms or tunicates and vertebrates such as fish, reptiles, birds or mammals can be found.

Compared to other bilateria, polychaetes have not changed their habitat since cambrian explosion and therefore did not have to adapt to new ecological niches and change (Weistheide and Rieger, 1996). Almost all characters shared within polychaetes are not polychaete-specific and can be found in other phyla too. Thanks to their "shallow" level of specialization, they still exhibit many ancestral characteristics. *Platynereis* for instance, exhibits developmental characteristics, such as amphistomous gastrulation and primary ciliated larva, which are considered ancestral (Nielsen, 2001). It additionally possesses ancestral morphological features (Prud'homme et al., 2003; Tessmar-Raible and Arendt, 2003), an ancestral type gene inventory (Raible et al., 2005) and ancestral cell types (Arendt et al., 2004; Tessmar-Raible et al., 2007; Denes et al., 2007; Arendt, 2008). For these reasons, *Platynereis* was ideally suited to complement previous microRNA expression data from protostomes which have been limited by technical challenges and derive from studies on fast-evolving species (Aboobaker et al., 2005; Gonzalez-Estevez et al., 2009).

1.7.3.2 Life cycle

Figure 1.8 illustrates the life cycle of *Platynereis dumerilii*. Following fertilization of the egg, a cortical reaction releases a jelly coat which cushiones zygotes from potential mechanical damage and helps them float. A series of spiral cleavages generate small micromeres apically and larger macromeres posteriorly. The developing embryo undergoes gastrulation and starts to differentiate into a trochophore larva which hatches after

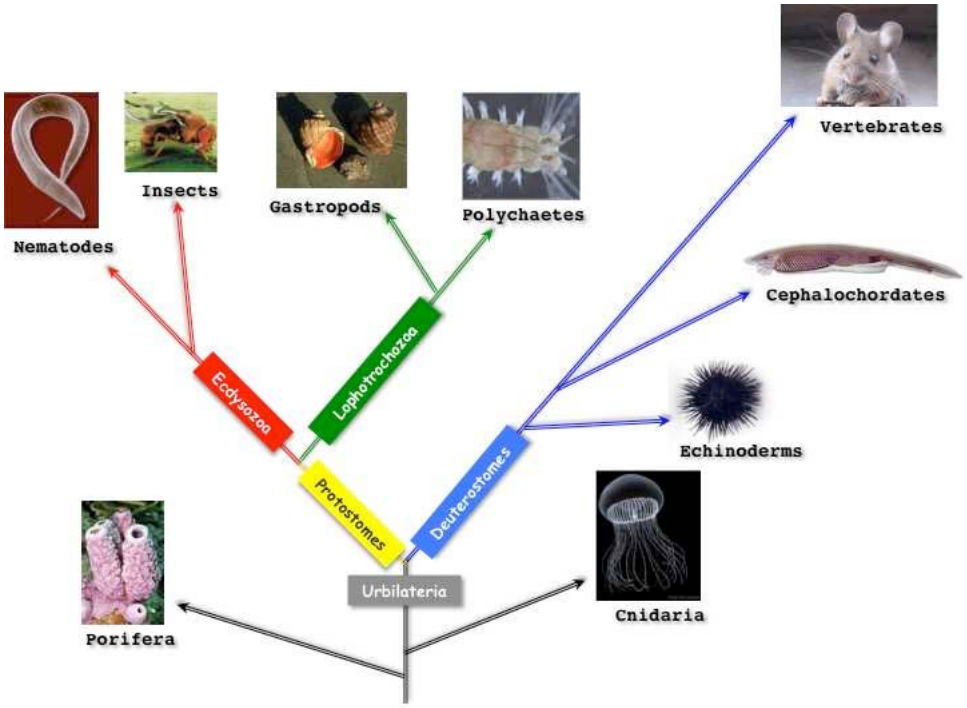


Figure 1.7: **A simplified phylogenetic tree of metazoa.** Sponges are basally positioned as they have no nervous system and very few cell types. Cnidaria, which have a difuse nervous system and only two germ layers pose the closest sister outgroup to bilateria (animals with bilateral symmetry). Bilaterian animals which in the vast majority have a central nervous system and three germ layers are divided in three major phyla; protostomes (yellow) which further split into ecdysozoa (red) and Lophotrochozoa (green) and deuterostomes . The last common ancestor of all bilaterian animals lived approximately 600mya and is refered to as "Urbilateria".

~18 hours post fertilization (hpf). At 48hpf, the swimming larva has an apical tuft, a prototroch ciliary belt (used for phototactic swimming) at the level of the stomodaeum and a telotroch at the posterior end (Fischer and Dorresteijn, 2004). The 48hpf brain of *Platynereis* consists of approximately 2000 cells which include a considerable number of differentiated neurons (Tessmar-Raible et al., 2007). As the larva further develops, segmentation begins with metameric chaetal sacs appearing progressively at 52hpf and from where chaeta and future parapodial appendages will project. By 72hpf, the larva is elongated into a neochaeta which continues to swim and only settles to the benthic floor after 5dpf when the foregut and gut have fully developed and the young worm is ready to start feeding as a scavenger. Upon settlement, new segments bud off the posterior growth zone of the worm until sexual maturation is reached 3-4 months post fertilization. The sexually mature *Platynereis* worms synchronously seek out for partners (following lunar periodicity) by abandoning their benthic tube and swimming in the open sea where they meet and spawn by releasing sperm and eggs which get externally fertilized. Life cycle summary from Fischer and Dorresteijn (2004).

The resulting large batches of fertilized, transparent eggs, combined with a very synchronous and stereotypical development render *Platynereis* a very good model for developmental studies. By now, a number of established molecular techniques including whole mount *in situ* hybridization, immunohistochemistry, micro-injections facilitate experimentation on *Platynereis*. EST and BAC libraries as well as very recent transcriptome data are available resources. Finally, culturing Platynereis in the lab is easy under an artifical lunar cycle and this is how cultures have been grown in the past 60 years.

1.7.3.3 Other model organisms

For selected microRNAs I also investigated microRNA expression in another annelid, *Capitella species I* (*spI*), in a slow-evolving deuterostome, the sea urchin *Strongylocentrotus purpuratus*, and in the sea anemone *Nematostella vectensis*.

Capitella spI is a polychaete marine annelid (Figure 1.7) whose development, like *Platynereis*, goes through swimming planktonic larval stages to mature into benthic adult worms (Hill and Boyer, 2003). Still, *Capitella*'s is phylogenetically distanct to *Platynereis*, showing many differences in development. This made it an excellent reference model to validate conservation of microRNA expression patterns across annelids.

Strongylocentrotus was chosen because it represents a slow evolving basal deuterostome species in the clade of echinoderms (Figure 1.7; Raible et al., 2005) which also goes through swimming planktonic larval development and shares comparable structures with *Platynereis* larvae (such as ciliary bands; Burke, 1978).

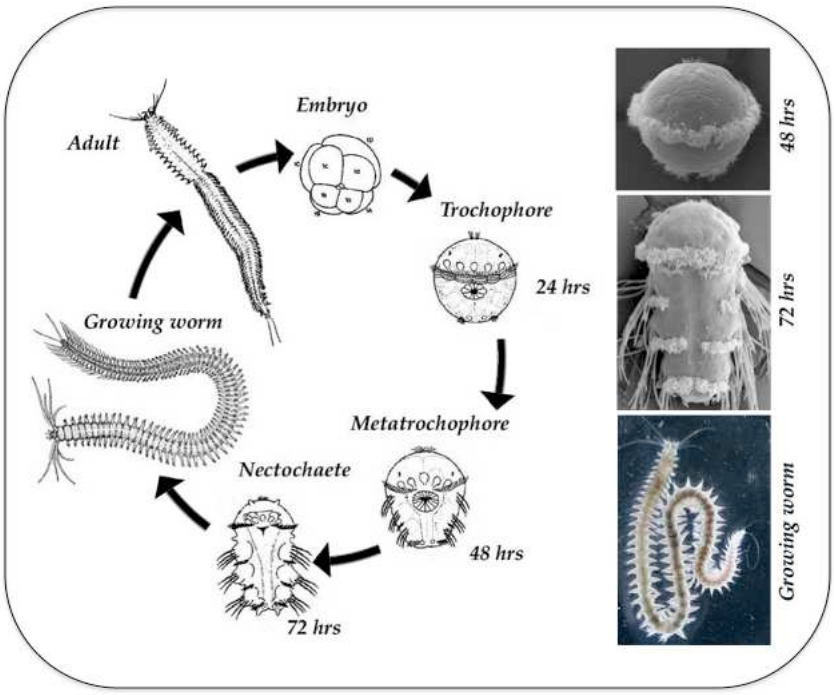


Figure 1.8: **The *Platynereis dumerilii* life cycle.** Upon fertilization, the zygote undergoes spiral cleavages to form a trochophora larva which swims as part of the plankton (in a phototactic manner) for few days. During development, the larva elongates to form a three segmented nectochaete which continues to swim until it is able to start feeding when it takes the decision to settle down in the benthic environment. From then on, the worms grow by adding more segments and in 3-4 months they sexually mature and seek out for partners (following lunar periodicity). The panels on the right give scanning electron microscopy (SEM) pictures of trochophore and nectochate stages. A white light picture depicts a growing worm prior to its sexual maturation. Life cycle figure by G. Balavoine, SEMs by H. Hausen and N. Dray and worm picture by K. Tessmar-Raible.

Finally, the cnidarian *Nematostella* represents an outgroup to the bilaterians (Figure 1.7) which is of special value because without an outgroup we cannot distinguish between loss, conservation or innovation between different lineages of protostomes and deuterostomes (Sullivan et al., 2006).

1.8 Aim of the thesis

Mounting evidence supports an emerging link between the evolution of microRNAs and the evolution of the complex bilaterian body plans (Peterson et al., 2009). Independent studies agree in that a set of microRNAs have been available to regulate gene expression since early in evolution (Hertel et al., 2006; Prochnik et al., 2007; Wheeler et al., 2009). However, to date, few data have been available to indicate what was the role and site of action of these microRNAs when they first evolved?

Large part of this thesis addressed this question through a comparative approach which aimed to unravel the ancestral sites of microRNA expression. I explored the microRNA expression of the slow evolving marine annelid worm studied in our laboratory, *Platynereis dumerilii*, which has been shown to be of special value in comparative studies (Denes et al., 2007; Tessmar-Raible et al., 2007). To make the comparative data more credible, I studied microRNA expression in other slow evolving protostome and deuterostome model organisms such as the marine annelid *Capitella* and the sea urchin *Strongylocentrotus* respectively. The comparative expression data show that in these animals, other than in the fast-evolving fly *Drosophila* and nematode *Caenorhabditis*, microRNAs largely retain their ancient expression sites.

A second question I set out to address was how did the step-wise acquisition of microRNA families contribute to bilaterian complexity. For this, I again used *Platynereis* as a model to localize the expression of some microRNAs specific to the protostome, the lophotrochozoan, the annelid and the *Platynereis* lineage. This gave a first insight into the site of action of lineage specific microRNAs which appear to reinforce the regulatory role of ancient microRNAs by joining them in the same expression sites.

Finally, the ultimate aim of this thesis was to bridge "evolutionary developmental biology" (evo–devo) and the microRNA field which had until recently very little overlap.

2 Results

2.1 The *Platynereis* microRNA repertoire

Mature microRNAs present in *Platynereis* were identified by high-throughput sequencing of complementary DNA libraries generated from 19–24- nucleotide RNAs, isolated from various developmental stages. 1.3 million reads matched annotated microRNA sequences in miRBase (Table 2.1; Griffiths-Jones et al., 2007), distributing to 66 microRNA families of which 36 exist in protostomes and deuterostomes (Table 2.2) and 30 in protostomes only, in accordance with recent studies (Sempere et al. 2006; Wheeler et al. 2009).

Deep sequencing revealed sequence variants for most *Platynereis* microRNAs. Typically, for each microRNA one sequence was highly abundant and therefore considered representative/canonical (analysis done by F. Raible). Remaining sequence variants occurred in lower numbers and exhibited point mutations and differential lengths (Figure 2.1).

2.1.1 microRNA phylogenetic distribution

2.1.1.1 Ancient Bilaterian microRNAs

36 *Platynereis* microRNAs (miR-1, -7, -8, -8*, -9, -9*, -10, -22, -29, -31, -33, -34, -71, -92, -100, -124, -125, -133, -137, -153, -183, -184, -190, -210, -216, -219, -242, -252, -263, -278, -281, -283, -315, -375, -2001 and let-7) were shared between protostomes and deuterostomes (which represent the two major superphyla of the bilaterians including insects, nematodes, annelids and sea urchins, vertebrates, respectively, among others) and should have therefore been present in the last common ancestor of bilateria. One of them, miR-100, is conserved in the sister group of cnidaria (Grimson et al., 2008) and therefore must have been present in the last common ancestor of bilaterians and cnidarians.

miR FAMILY	VARS	OVERALL	REPRESENTATIVE SEQUENCE	READS	READS PER SMALL RNA LIBRARY			deuterostomes	ecdysozoans	lophotrochozoans	cnidaria	CONSERVED AMONG
miR-1	186	62651	UGGAAUGUAAAGAAGUAUGUAG	57587	zygotic|32	larval|20242	worm|37313	+	+	+	-	[mmu][cel][lqi]
miR-7	96	7596	UGGAAGACUAGUGAUUUUGUUGUU	5881	zygotic|7	larval|1335	worm|4539	+	+	+	-	[hsa][dme][lqi]
let-7	24	230	UGAGGUAGUAGGUUGUAUAGU	71	zygotic|24	larval|42	worm|5	+	+	+	-	[hsa][cel][lqi]
miR-8/200/141	199	127338	UAAUACUGUCAGGUAAAGAUGUU	111203	zygotic|163	larval|49446	worm|61594	+	+	+	-	[mmu][dme][lqi]
miR-8*	26	1372	CAUCUUACUGGGCAGCAUUAGA	1291	zygotic|0	larval|758	worm|533	+	+	+	-	[mmu][lmi][pdu]
miR-9	52	1395	UCUUUGGUUAUCUAGCUGUAUGA	943	zygotic|1	larval|86	worm|856	+	+	+	-	[mmu][dme][lqi]
miR-9*/79	31	816	AUAAAGCUAGGUUACCAAAGCU	366	zygotic|27	larval|48	worm|291	+	+	+	-	[spu][lmi][lqi]
miR-10	68	3745	UACCCUGUAGAUCCGAAUUUGU	2891	zygotic|10	larval|1556	worm|1325	+	+	+	-	[hsa][dme][lqi]
miR-22/745	69	33067	AGCUGCCUGGUGAAGAGCUGUC	26280	zygotic|111	larval|13781	worm|12388	+	+	+	-	[dre][dme][lqi]
miR-29	199	16172	UAGCACCAUUUGAAAUCAGUUU	6429	zygotic|354	larval|4264	worm|1811	+	+	+	-	[has][dme][lqi]
miR-31	145	22674	AGGCAAGAUGUUGGCAUAGCUGA	10177	zygotic|4	larval|1290	worm|8883	+	+	+	-	[dre][dme][lqi]
miR-33	2	3	GUGCAUUGUAGUUGCAUUGCA	3	zygotic|0	larval|0	worm|2	+	+	+	-	[hsa][dme][lqi]
miR-34	68	3023	UGGCAGUGUGGUUAGCUGGUUGU	2316	zygotic|0	larval|1624	worm|692	+	+	+	-	[dre][dme][lqi]
miR-71	169	9502	UGAAAGACAUGGGUAGUGAGAUG	4251	zygotic|1	larval|2818	worm|1432	+	+	+	-	[spu][cel][lqi]
miR-92	48	2089	AAUUGCACUUGUCCCGGCCUGC	1605	zygotic|3	larval|1039	worm|563	+	+	+	-	[dre][dme][lqi]
miR-100	4	33	AACCCGUACAACCGAACUUGUG	24	zygotic|2	larval|13	worm|9	+	+	+	+	[hsa][dme][lqi][nve]
miR-124	93	8405	UAAGGCACGCGGUGAAUGCCA	7217	zygotic|5	larval|487	worm|6725	+	+	+	-	[dre][dme][lqi]
miR-125	21	520	UCCCUGAGACCCUAACUUGUGA	384	zygotic|36	larval|274	worm|74	+	+	+	-	[hsa][dme][csp]
miR-133	117	19765	UUGGUCCCCUUCAACCAGCUGU	17186	zygotic|13	larval|3077	worm|14096	+	+	+	-	[hsa][dme][lqi]
miR-137	36	828	UAUUGCUUGAGAAUACACGUAG	336	zygotic|0	larval|57	worm|279	+	+	+	-	[hsa][dme][lqi]
miR-153	92	7811	UUGCAUAGUCACAAAAGUGAUC	6628	zygotic|27	larval|1300	worm|5301	+	+	+	-	[hsa][dme][lqi]
miR-183	99	21428	AAUGGCACUGGUAGAAUUCACGG	16045	zygotic|3	larval|15708	worm|334	+	+	+	-	[dre][dme][lqi]
miR-184	131	26743	UGGACGGAGAACUGAUAAGGGC	23950	zygotic|16	larval|12857	worm|11077	+	+	+	-	[dre][dme][lqi]
miR-190	2	7	AGAUAUGUUUGAUAUAUUUGGU	5	zygotic|0	larval|0	worm|5	+	+	+	-	[hsa][dme][lqi]
miR-210	14	187	CUUGUGCGUGUGACAGUGACAAU	160	zygotic|0	larval|66	worm|94	+	+	+	-	[dre][dme][lqi]
miR-216	1	1	UAAUCUCAGCUGGCAAAAGUGAG	1	zygotic|0	larval|0	worm|1	+	+	+	-	[dre][dme][lqi]
miR-219	4	17	UGAUUGUCCAAACGCAUUUCUUG	10	zygotic|0	larval|0	worm|10	+	+	+	-	[dre][dme][csp]
miR-252	54	7139	CUAAGUACUAGCGCCGCAGGA	4995	zygotic|3	larval|3376	worm|1616	+	+	+	-	[bfl][dme][csp]
miR-263/182	61	3339	CUUGGCACUGGUAGAAUUCACUGA	2851	zygotic|0	larval|576	worm|2275	+	+	+	-	[dre][dme][lqi]
miR-278	19	3195	UCGGUGGGACUUUCGUUCGUUC	2685	zygotic|0	larval|549	worm|2136	+	+	+	-	[spu][dme][lqi]
miR-281	29	2934	UGUCAUGGAGUUGCUCUCUUUA	2550	zygotic|2	larval|380	worm|2168	+	+	+	-	[bfl][dme]lqi]
miR-283	14	371	AAAUAUCAGCUGGUAAUUCUGAG	313	zygotic|0	larval|160	worm|153	+	+	+	-	[dre][dme][lqi]
miR-315	101	17744	UUUUGAUUGUUGCUCAGAAAGCC	15996	zygotic|18	larval|9647	worm|6331	+	+	+	-	[pem][dme][lqi]
miR-375	9	304	UUUGUUCGUCCGGCUCGCGUUA	279	zygotic|0	larval|0	worm|279	+	+	+	-	[dre][dme][lqi]
miR-2001	1	1818	UUGUGACCGUUACAAUGGGCA	1818	zygotic|2	larval|1442	worm|374	+	-	+	-	[spu][lqi]
miR-242	7	3012	UUGCGUAGGCGUUGUGCACAGA	2757	zygotic|2	larval|1791	worm|964	+	-	+	-	[bfl][csp]
bantam	64	9469	UGAGAUCAUUGUGAAAACUGAUU	8261	zygotic|50	larval|1915	worm|6296	-	+	+	-	[dme][csp]
miR-2	397	90976	UAUCACAGCCAGCUUUGAUGAGC	20736	zygotic|215	larval|9452	worm|11069	-	+	+	-	[dme][lqi]
miR-12	12	225	UGAGUAUUACAUCAGGUACUGA	193	zygotic|0	larval|21	worm|172	-	+	+	-	[dme][lqi]
miR-67	6	1035	UCACAACCUGCAUGAAUGAGGUA	454	zygotic|0	larval|323	worm|131	-	+	+	-	[dme][lqi]
miR-87	184	55442	GUGAGCAAAGUUUCAGGUGUGU	50218	zygotic|109	larval|7069	worm|43040	-	+	+	-	[dme][lqi]
miR-277	30	1196	UAAAUGCAUUAUCUGGUAUGUA	920	zygotic|9	larval|753	worm|158	-	+	+	-	[dme][csp]
miR-317	93	3985	UGAACACAGCUGGUGGUAUCUCUUU	1376	zygotic|1	larval|1051	worm|324	-	+	+	-	[dme][lqi]
miR-279	148	44057	UGACUAGAUCCACACUCAUCC	38150	zygotic|1817	larval|21297	worm|15036	-	+	+	-	[dme][lqi]
miR-981	5	856	UUCGUUGUCGUCGAAACCUGCCU	743	zygotic|0	larval|162	worm|581	-	+	+	-	[dme][lqi]
miR-993/10*	3	74	AGAAGCUCGGUUCUACAGGUAU	69	zygotic|0	larval|38	worm|31	-	+	+	-	[dpu][csp]
miR-305	4	10	AUUGUACUUCAUCAGGUGCUCUGG	4	zygotic|0	larval|4	worm|0	-	+	-	-	[dme]
miR-318	4	33	UCACUGGGCUUUGUUUAUCUCA	30	zygotic|22	larval|6	worm|2	-	+	-	-	[dme]
miR-989	3	9	UGUGAUGUGACGUAGUGGAACA	7	zygotic|3	larval|4	worm|0	-	+	-	-	[dme]
miR-996	3	8	UGACUAGAUUUCAUGCUCGUCUA	6	zygotic|5	larval|1	worm|0	-	+	-	-	[dme]
miR-36	10	7787	UCACCGGGUAUACAUUCAUCCG	6118	zygotic|19	larval|3751	worm|2348	-	-	+	-	[sme]
miR-96	7	2370	CUUGGCACUGGCGGAAUAAUCAC	1335	zygotic|1	larval|600	worm|734	-	-	+	-	[lqi]
miR-133*	3	641	AGCUGGUUGAAAUAGGGCCAAAU	384	zygotic|0	larval|156	worm|228	-	-	+	-	[lqi]
miR-750	3	33	UCAGAUCUAACUCUUCCAGCUCA	17	zygotic|0	larval|2	worm|15	-	-	+	-	[sme]
miR-1175	3	414	UGAGAUUCAACUCCUCCAACUGC	370	zygotic|0	larval|71	worm|299	-	-	+	-	[lqi]
miR-1175*	5	723	AGUGGAGAGAGUUCUAUCUCAUC	561	zygotic|0	larval|156	worm|405	-	-	+	-	[lqi]
miR-1986	1	21	UGGAUUUCCCAUGAUCCGUAAC	21	zygotic|0	larval|21	worm|0	-	-	+	-	[lqi]
miR-1989	18	9222	UCAGCUGUCGCGAUGCCUUCUU	3230	zygotic|0	larval|248	worm|140	-	-	+	-	[lqi]
miR-1992	26	17584	UCAGCAGUUGUACCACUGAUGUG	16048	zygotic|12	larval|2536	worm|13500	-	-	+	-	[lqi]
miR-1993	1	80	UAUUAUGCUGUUAUUCACGAGA	80	zygotic|0	larval|15	worm|65	-	-	+	-	[lqi]
miR-1994	10	7830	UGAGACAGUGUGUCCUCCCUCG	7174	zygotic|5	larval|2292	worm|4877	-	-	+	-	[lqi]
miR-1996	27	22111	AUCAAGUGAGGUCAGAUCUUGG	19127	zygotic|24	larval|1850	worm|17253	-	-	+	-	[csp]
miR-1997	6	2867	UCUGCAGGUUCACAUCAGCCCCA	2743	zygotic|10	larval|1152	worm|1581	-	-	+	-	[csp]
miR-1998	21	13688	UUGAACGCAGAGAUGUACAUCA	13126	zygotic|10	larval|6902	worm|6214	-	-	+	-	[csp]
candidate L49	1	22	UAACUCAGUCAGAUCACAGGC	22	zygotic|2	larval|13	worm|7	-	-	+	-	[csp]
candidate K	1	16674	UGACUAAAAGUAUUGAAGGCUU	16674	zygotic|20	larval|10917	worm|5737	-	-	-	-	
candidate C	1	6894	UUUGGCACCAGUGGAAUAGUCAC	6894	zygotic|4	larval|386	worm|6504	-	-	-	-	

Table 2.1: **The *Platynereis* microRNA repertoire**. (vars, sequence variants for a given microRNA; overall, total number of related solexa reads; representative sequence, solexa read with highest occurrence for a group or related sequences; reads, occurrence of the representative sequence; reads per library, occurrence of representative sequence in each solexa library; source, initials of species with orthologos microRNA sequence- see Appendix).

Foregut	Locomotor cilia	Neurosecretory brain tissue	Sensory brain tissue	General musculature	General CNS	Sensory organs	Gut	Other
miR-100	miR-29	miR-7	miR-9	miR-22#	miR-71	miR-8	miR-216	miR-315
let-7^	miR-92	miR-137	miR-9*	miR-1	miR-124	miR-183	miR-283	miR-281
miR-125^	miR-34	miR-153		miR-133	miR-184	miR-263		miR-210^
miR-375^					miR-190	miR-252^		miR-33
miR-10					miR-219	miR-2001		
miR-278								
miR-31								

miRNAs are classified according to their site of expression in Platynereis. Grouped miRNAs are genomically clustered or processed from one common polycistronic transcript in at least one bilaterian animal (Suppl. Table 3).
^Also expressed in gut
#Also expressed in locomotor cilia

Table 2.2: **Ancient bilaterian microRNAs**

Sequence	counts	frequency
UCUUUGGUUAU-CUAGCUGUAUGA	943	71%
UCUUUGGUUAU-CUAGCUGUAUG	215	16%
UCUUUGGUUAU-CUAGCUGUAU	75	6%
CUUUGGUUAU-CUAGCUGUAUGA	64	5%
UCUUUGGUUAU-CUAGCUGUAUGAA	18	1%
UCUUUGGUGAUUUUAG	9	1%
UCUUUGGUGAUUUUAGCUGUAUG	6	0%
UCUUUGGUUAU-CUAGCUGUAC	5	0%

(a) **miR-9 variants identified in the dataset.** Variants include 3'-variants (colored in blue), insertions (yellow) and a 5' variant predicted to shift the seed region of the miRNA. Only variants with at least 5 representatives in the dataset are shown. Frequencies reflect the relative abundance of each variant among the displayed set.

Sequence	counts	frequency
AGCUGCCUGGUGAAGAGCUGUC	26280	53%
AGCUGCCUGGUGAAGAGCUGUCU	2691	5%
AGCUGCCUGGUGAAGAGCUGU	2226	4%
AGCUGCCUGGUGAAGAGCUGUCA	1274	3%
AGCUGCCUGGUGAAGAGCUG	362	1%
AGCUGCCUGGUGAAGAGCUGC	305	1%
AGCUGCCUGGUGAAGAGCUGUCC	210	0%
AGCUGCCUGGUGAAGAGCUGGC	182	0%
AGCUGCCUGGUGAAGAGCUGCC	128	0%
AGCUGCCUGGUGAAGAGCUGUU	55	0%
AGCUGCCUGGCGAAGAGCUGUC	50	0%
AGCUGCCUGGUAAAGAGCUGUC	Cap-miR-745a	
GAGCUGCCAAGUGAAGGGCUGU	9905	20%
GAGCUGCCAGGUGAAGGGCUGU	4193	8%
GAGCUGCCAAGUGAAGGGCUGC	1053	2%
GAGCUGCCAGGUGAAGGGCUGC	412	1%
GAGCUGCCAAGUGAAGGGCUG	100	0%
GAGCUGCCAGGUGAAGGGCUG	62	0%
GAGCUGCCAAGUGAAGGGCUGUU	51	0%
GAGCUGCCAAGUGAAGGGCUGU	Cap-miR-745b	

(b) **miR-22/-745 variants identified in the dataset.** Variants fall into two categories of 5'-variants (colored in blue and yellow, respectively) that correlate with the two aligned miRNA paralogs identified in Capitella. Only variants with at least 50 representatives in the dataset are shown.

Figure 2.1: **Examples of 3'/5' variability within our analyzed sequence dataset**

2.1.1.2 Protostome specific microRNAs

Deep sequencing for small RNAs has only been materialized for few lophotrochozoan species (Friedländer et al. 2009; Lu et al. 2009; Wheeler et al. 2009). As a result the lophotrochozoan microRNA list is still incomplete in miRBase and matching reads only represent a fraction of protostome microRNAs:

30 *Platynereis* microRNAs were found only in the protostome lineage. Of these, 14 were shared between ecdysozoa and lophotrochozoa (bantam, miR-2, -12, -67, -87, -277, -279, -305, -317, -318, -981, -989, -993/10*, -996) while the remaining 16 appeared to be lophotrochozoan specific microRNAs (miR-36, -96, -133*, -746, -750, -1175, -1175*, -1986, -1989, -1992, -1993, -1994, -1996, -1997, -1998, -L49) most of which have also been cloned and sequenced by (Wheeler et al., 2009).

Reads that did not match any annotated microRNA sequence of miRBase were mapped to sequenced lophotrochozoan genomes by F. Raible in search for candidate microRNA loci which fulfilled structural stability criteria (Lu et al. 2008b; Hofacker et al. 1994) and could give rise to precursor microRNA characteristic folds (hairpin structures). A list of 27 candidate lophotrochozoan microRNAs was assembled upon identification of a candidate microRNA locus in any of the following lophotrochozoan genomes: *Capitella spI, Helobdella robusta, Lottia gigantea* (see table in appendix).

To systematically confirm lophotrochozoan microRNA candidates, I adapted an in vivo method used to validate *Drosophila* predicted microRNA candidates (Sandmann et al., 2007). By cloning genomic loci which host candidate microRNAs of interest in the closely related annelid *Capitella spI* and over-expressing them in S2 cell lines, I could verify by Northern blot analysis if these candidate primary transcripts get processed by the cells' microRNA biogenesis machinery into mature 21- to 23-mers. Using this technique, I could rule out the validity of 8 out of the 27 lophotrochozoan specific microRNA candidates, 3 of which had been cloned (but not validated in Wheeler et al., 2009) and deposited on miRbase as: miR-1987, -1995 and -2000. There are still 19 candidates remaining to be validated (see appendix) aside from the now annotated (Wheeler et al., 2009) but earlier just validated by me: miR-242, -1986, -1996, -1998, and L49 (See appendix for northern blots).

2.1.1.3 *Platynereis* specific microRNAs

As we lack information about the complete *Platynereis* genome, all remaining sequences which could not map in other genomes were treated as candidate *Platynereis* specific microRNAs (solexa reads with hits in rRNA or tRNA databases had been removed already).

miRNA clusters	Pdu	Lgi	Dpu	Dme	Cel	Bfl	Dre	Cin	Mmu	Hsa
100-let7-125	100-let7	100-let7-125	125-let7	100-let7-125	100-let7-125	let7-100-125	100-let7-125	125-let7	100-let7-125	100-let7-125
2-71		2-71	2-71		2-71					
12-216-283	12-216	12-216	12-283	12-283-304†						
317-277-34			317-34	317-277-34						
1-133		1-133				1-133	1-133		1-133	1-133
183-263	183-263	183-263			183-263 (64-65-66)	183-183	183-182		183-182	183-182
mir9-79	9-79	9-79	9-79	9-79 ‡				9-79		

Table 2.3: **Conservation of miRNA clusters across bilateria.** List of conserved microRNA clusters commonly identified in different bilaterian genomes. ***Pdu***- *Platynereis dumerilii*; ***Lgi***- *Lottia gigantea*; ***Dpu***- *Daphnia pulex*; ***Dme***- *Drosophila melanogaster*; ***Cel***- *Caenorhabditis elegans*; ***Bfl***- *Branchiostoma floridae*; ***Dre***- *Danio rerio*; ***Cin***- *Ciona intestinalis*; ***Mmu***- *Mus musculus*; ***Hsa***- *Homo sapiens*.Highlighted clusters have been shown to give rise to polycistronic transcripts in ***Pdu*** (verified by cDNA amplification-see materials and methods for genbank accession numbers), in ***Dme*** (verified by Sokol et al. 2008), in ***Dre*** (predicted and verified by Thatcher et al. 2008) and in ***Mmu*** (verified by Xu et al. 2007). Table assembly based on a number of different sources: all cluster predictions for ***Cel***, ***Dme***, ***Dre***, ***Mmu*** and ***Hsa*** were retrieved from miRGen database (Megraw et al., 2007). ***Cin*** clusters reported by (Legendre et al., 2005). ***Lgi*** clusters discussed in (Sokol et al. 2008; Prochnik et al. 2007). ***Bfl*** microRNA clusters predicted by Luo and Zhang 2009. ***Dpu*** clusters retrieved from Daphnia Genomics Consortium. † miR-304 is the **Dme** ortholog of miR-216. ‡ In ***Pdu***, ***Lgi***, ***Cin***, miR-79 is actually 9* (belongs to pre-miR-9). In ***Dme*** miR-79 is a distinct miRNA gene found in the same cluster as miR-9 but further downstream.

These sum to at least 32 microRNA candidates and of them 2 have been experimentally validated by WMISH were named miR-C and miR-K (Table 2.1, Figure 2.24 on page 96 and 2.22 on page 94).

2.1.2 Conservation of microRNA genomic clusters across bilateria

Many microRNAs appear in clusters on a single polycistronic transcript (Lee et al., 2002; Mourelatos, 2002; Lai et al., 2003). Certain microRNA clusters are commonly identified in different bilaterian genomes. They exhibit conservation in microRNA content, synteny and for some it has been experimentally verified that they get processed from one single polycistronic transcript (Table 2.3).

In *Platynereis* the presence of three conserved microRNA clusters (miR-100/let7; miR-263/183 and miR-12/216 every single one of which was detected as one polycistronic transcript) was confirmed experimentally through RT-PCR. The identified *Platynereis* cluster members do not share any sequence similarity except from the miR-263/183 cluster where one of the two microRNAs arose after gene duplication of its paralog followed by mutational accumulation.

Consistent with their organization in the genome, *Platynereis* cluster members showed a similar expression pattern during development (Table 2.2, Figure 2.26 and Figure 2.27).

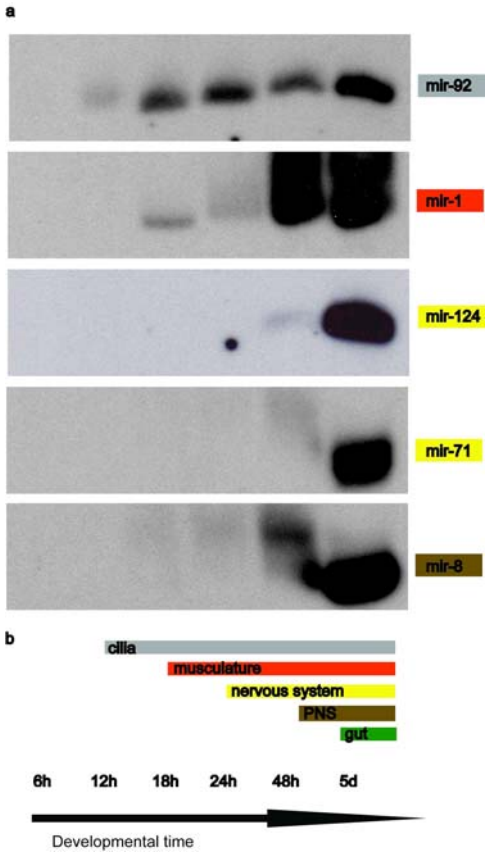


Figure 2.2: **miRNA expression onset is coupled to differentiation.** Northern blots were used to detect the expression onset of selected miRNAs. **a,** miRNAs whose expression demarcates specific tissues are detectable only when **b,** the corresponding tissues start to differentiate.

2.2 Tissue specific activity of *Platynereis* microRNAs

To investigate the temporal and spatial localisation profile of conserved bilaterian microRNAs, whole-mount in situ hybridizations (WMISH) were conducted using Locked Nucleic Acids as probes (Valoczi et al., 2004). In all cases, localisation of mature *Platynereis* microRNAs was spatially restricted and, as revealed by comparison to marker gene analysis, almost exclusively occurred in differentiating tissues. Table 2.2 summarizes the tissue affinity of all conserved bilaterian microRNAs. WMISH comparison with differentiation markers and Northern Blots for selected microRNAs at six consecutive developmental stages revealed that regardless of tissue identity, microRNA expression started with the onset of differentiation (Fig 2.2) and became more widespread concomitant with the increase in number of differentiated cells.

Specific sets of microRNAs showed very similar, spatially restricted expression (Table

2.2). In many cases, co-expressed microRNAs originate from the same hairpin (miR-9/-9*), from the same transcript (pri-miR-183-263), or from duplicated genes; in other cases, however, they proved entirely unrelated (Table 2.1).

Whenever ancient bilaterian microRNAs had been reported to genomically cluster in any of the sequenced bilaterian genomes, or originated from the same transcript in *Platynereis* or in any other bilaterian model (Table 2.3), these microRNAs proved to be specifically co-expressed in *Platynereis* (black brackets in Table 2.2).

Two distinct evolutionary scenarios can explain the similar, highly restricted expression of seemingly unrelated, ancient bilaterian microRNAs in *Platynereis*. Either, these microRNAs were initially expressed elsewhere and have converged secondarily on the very same tissue. Or, they have been expressed in this tissue ever since they evolved. This latter scenario would imply that the tissues were in place before the respective microRNAs evolved, i.e. existed already in the protostome-deuterostome ancestor. To distinguish between these possibilities, I compared microRNA expression in *Platynereis* with that in other slow-evolving metazoans.

2.2.1 The localisation of ancient bilaterian microRNAs

2.2.1.1 miR-100, -125, let-7 and the ancient mouth

The most ancient, conserved microRNA, miR-100, shared by cnidaria and bilateria (Grimson et al. 2008; Wheeler et al. 2009), highly specifically localised to two small groups of cells in the larval foregut (Figure 2.3a, f). Let-7, which plays a conserved role in developmental timing in fly and nematode (Pasquinelli et al. 2000Reinhart et al. 2000; Caygill and Johnston 2008; Sokol et al. 2008), was also detected in these cells, from 5 days post fertilization (5dpf) onwards (Figure 2.3b,g). As in fly (Sokol et al. 2008Feinbaum and Ambros, 1999), *Platynereis* miR-100 and let-7 are processed from a single polycistronic transcript; yet, the similar highly restricted localisation of the mature microRNAs must be due to post-transcriptional co-regulation (Heo et al., 2008; Lehrbach et al., 2009) given that the precursor transcript was detected more broadly in the larval foregut and in part of the brain (Figure 2.3d). miR-125, another conserved microRNA derived from the same transcript and similarly exerting a conserved role in the control of developmental timing (Sokol et al. 2008; Olsen and Ambros 1999) was also detected in the *Platynereis* foregut but in a less restricted pattern (Figure 2.3c). These cells also expressed miR-375 (Fig 2.3e), known to be highly specifically expressed in foregut-related, neurosecretory/endocrine cell populations of the vertebrate pituitary and pancreas (Poy et al., 2004).

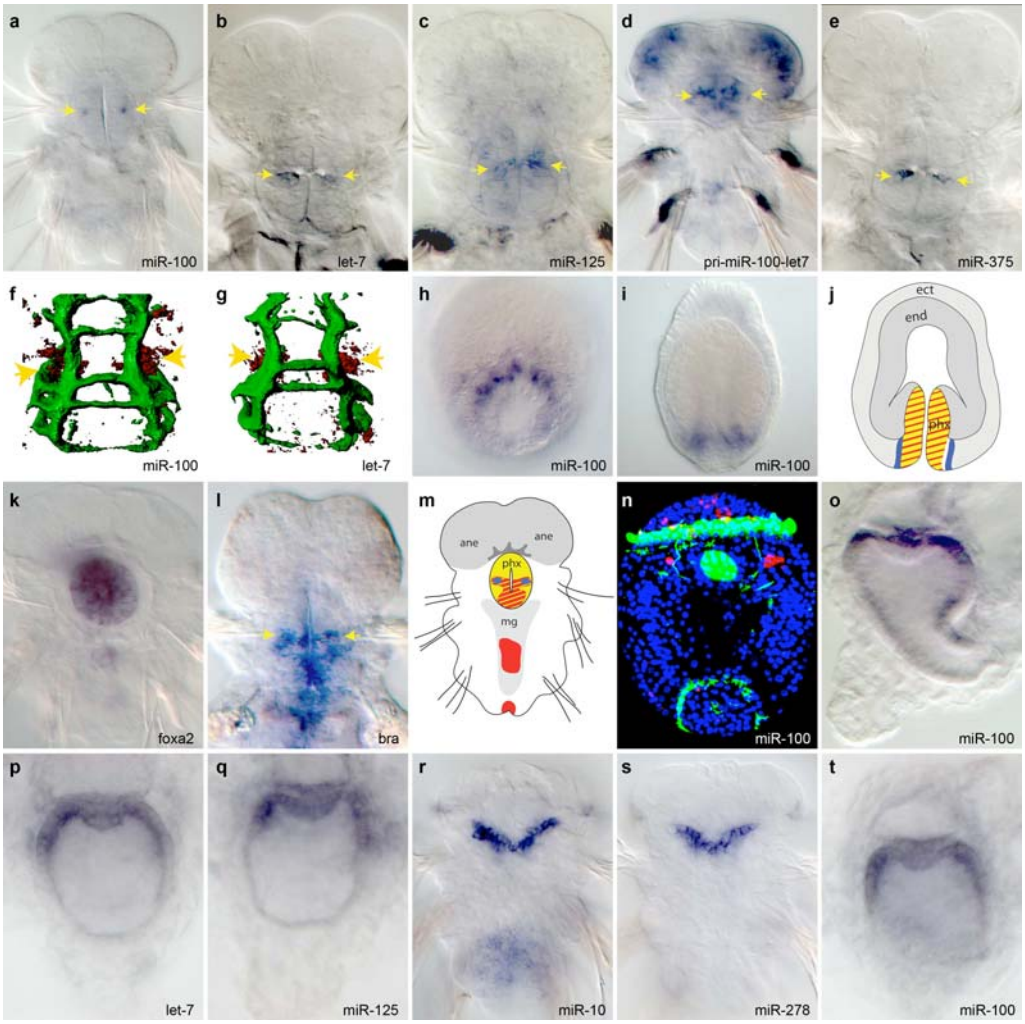


Figure 2.3: **Foregut-related expression of conserved microRNAs. a,** ventral view of 72 hours post fertilization (hpf) *Platynereis* with miR-100 expression restricted to a small group of cells in the foregut. **b,** dorsal view of 5dpf *Platynereis* with foregut let-7 expression. **c,** ventral view of 5dpf *Platynereis* showing miR-125 expression in the foregut. **d,** ventral view of 72h *Platynereis* larva with strong expression of pri-miR-100-let-7 in the brain, foregut and gut. **e,** dorsal view of 5dpf *Platynereis* with foregut miR-375 expression. **f,** 3D reconstruction of the 5 days post fertilization (dpf) *Platynereis* foregut (acTub green, miR-100 expression red). **g,** 3D reconstruction of the 5 dpf foregut (acTub green, let-7 expression red). **h,** oral view of the *Nematostella* planula larva expressing miR-100 in cells around the pharynx. **i,** lateral view of the *Nematostella* planula larva expressing miR-100 **j,** scheme of *Nematostella* planula with pharynx demarcated by *brachyury* (Scholz and Technau, 2003) expression in red, *foxa* (Martindale, 2004) expression in yellow and miR-100 positive cells in blue (phx, pharynx; end, endoderm; ect, ectoderm). **k,** ventral view of 72hpf *Platynereis* larva expressing *foxa2* (*forkhead*). **l,** dorsal view of 5dpf *Platynereis* with foregut and midgut *brachyury* expression. **m,** scheme of 72hpf *Platynereis* with foregut expressing *foxa* (yellow), *brachyury*(Arendt et al., 2001) (red) and miR-100 (blue). **n,** ventral view of *Capitella* stage 4 larva expressing miR-100 in few cells on each side of the foregut and in the brain. **o,** lateral view of *Strongylocentrotus* with miR-100 expression in the sphincter region interconnecting esophagus and stomach. **p-q,** dorsal view of 4 week old sea urchin pluteus larvae expressing let-7 and miR-125 in the sphincter region interconnecting esophagus and stomach. **r,** ventral view of 72hpf *Platynereis* expressing miR-10 in cells along the mouth and at the posterior ventral nervous system. **s,** ventral view of 72hpf larva with restricted expression of miR-278 in tissue along the mouth. **t,** dorsal view of 4 week old sea urchin pluteus larvae expressing miR-100 in the sphincter region interconnecting esophagus and stomach.

Since miR-100, -125 and let-7 are expressed much more broadly in fly (Sokol et al., 2008), and vertebrate (Table 2.4 on page 89), I investigated the expression of miR-100 in a cnidarian, the sea anemone *Nematostella* to test if the highly specific *Platynereis* pattern is conserved. miR-100 was exclusively detected in single cells located around the pharynx anlage (Figure 2.3h-i), demarcated by *brachyury*(Scholz and Technau, 2003) and *foxa* (Martindale, 2004) expression (blue, red and yellow in Figure 2.3j scheme). Intriguingly, miR-100, *brachyury*(Arendt et al., 2001) and *foxa2* also specifically co-localise in the *Platynereis* foregut (Figure 2.3k-m scheme). Together, these findings indicate that early in animal evolution miR-100 was active in a small population of cells located around a digestive opening.

To further test evolutionary conservation of the foregut as ancestral expression site of miR-100, -125 and let-7, I investigated their expression in the annelid *Capitella* (Figure 2.3n) and in the sea urchin *Strongylocentrotus* (Figure 2.3o-q, t). In both species localisation was similarly restricted to foregut tissue.

Expression analysis of another conserved bilaterian microRNA, miR-10, lent further support to an ancient activity of miR-100, -125 and let-7 in the foregut. The *miR-10* gene is a miR-100 duplicate that jumped into the Hox cluster (Aboobaker et al., 2005), thus acquiring a hox4- like expression pattern (Kulakova et al., 2007) also in *Platynereis* (posterior staining in Figure 2.3r).

miR-10 was additionally detected in a prominent V-shaped domain enclosing the mouth opening (Figure 2.3r), adjacent to but not overlapping the miR- 100/ let-7+ cells. This, combined with the shared proximity to the mouth, indicates a common evolutionary origin of the miR-10 and miR-100 expression domains from an ancient mouth-related pattern. The V-shaped domain is further linked to miR-100+ / let-7+ cells through a shared marker between the two cell types' molecular fingerprints; miR-278. The latter microRNA is known to play a role in hormonal control of energy homeostasis in the fly (Teleman et al., 2006) and was localised in both tissues (Figure 2.3s) which are distinctly expressing miR-10 and miR-100/let-7 (Table 2.5 on page 108).

Insights into *Platynereis* pharynx anatomy and characterization of miR-100/ let-7+ cells To characterize the nature of the miR-100+, let-7+ foregut cells in *Platynereis* it was necessary to investigate the anatomy of the entire foregut to better understand the arrangement of these cells in relation to the stomatograstric nervous system and all other cell types that comprise the foregut. Since very little was known about the *Platynereis* foregut, both structural (through white light and confocal microscopy) and molecular (using gene markers and antibody markers already available in the lab) characterization

of the organ was undertaken.

At 5dpf, the foregut is a well visible oval shaped structure under white light microscopy. Its length occupies large part of the young worm's body as it spans from the mouth opening (mid-head region) down to the end of the first body segment where the midgut begins (Figure 2.4a). In width, the foregut almost occupies the entire body volume (from ventral to dorsal) as it is positioned right behind the ventral nerve cord and extends to the dorsal epidermis.

Three foregut structures (pointed with black arrows in Figure 2.4a) are easily recognizable under white light microscopy: 1) the digestive tract which runs vertically from anterior to posterior and divides the foregut in two halves 2) a pair of small oval shaped compact structures, positioned right in the middle of the foregut, adjacent to the digestive tract and 3) a sphincter-like structure at the very posterior end of the foregut which connects it to the midgut.

A more detailed study on foregut cells' nuclei arrangement (stained by DAPI), revealed a number of distinct cellular populations which vary in position, density and arrangement/orientation. The foregut is divided into four quadrants, characteristic for their very dense cell population. The two anterior quadrants are smaller than the two posterior ones and get divided by another group of cells, which appear less dense than their neighbours (circled in Figure 2.4b), positioned right in the middle of the foregut. These "sandwiched" cells' position is reminiscent of the two oval shaped structures which are visible under white light (Figure 2.4a). An elaborate arrangement of longitudinal and dorso-ventral axons "wraps" all the aforementioned cell clusters to enervate them (Figure 2.4c-d). A distinct line of cells (note vertical arrangement of nuclei pointed by red arrow Figure 2.4b) runs along the lateral most part of the foregut encapsulating both the stomatograstric nervous system and all the cells that it enervates in foregut.

Investigation of the foregut at the molecular level by marker gene analysis allowed the annotation of all these distinct cell clusters and structures in greater detail. The highly specific localisation of miR-100 and the foregut expression of let-7 and miR-375 proved to coincide in position with the small oval structures in the middle of the foregut (Figure 2.4b, Figure 2.4e). Under white light, the latter are best visible from a dorsal view. Similarly, miR-375 detection was recorded predominantly in the dorsal most stacks of the confocal scan and after reconstruction of the structure and rotation of the stomatograstric nervous system to a lateral view, it became apparent that the cells' coordinates are dorsomedial (Fig 2.4j).

Co-expression of *synaptotagmin*, a neuronal differentiation marker (Denes et al., 2007) and of *prohormone convertase 2* (Tessmar-Raible et al., 2007) (Figure 2.4i), responsible

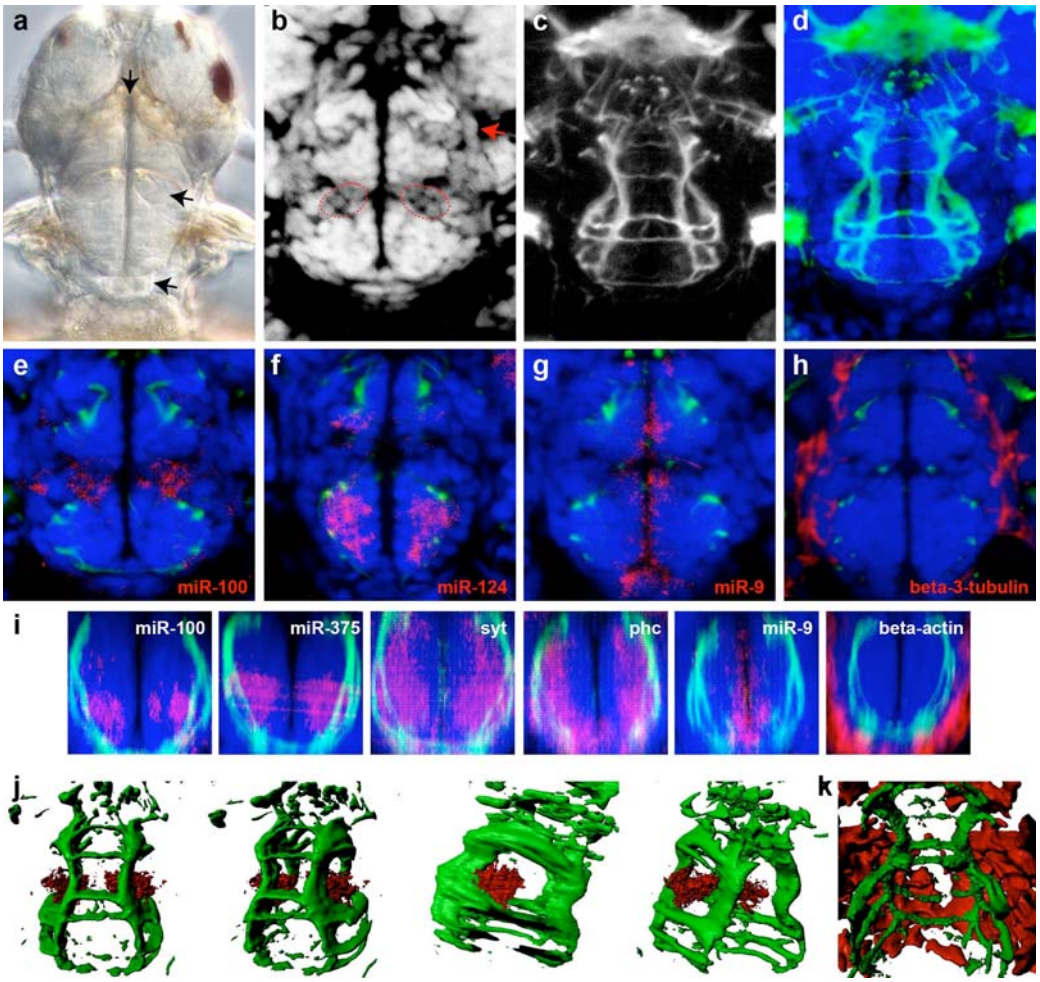


Figure 2.4: **Characterization of foregut cell types.** **a,** dorsal view of living 5 days post fertilization (dpf) *Platynereis* worm head region plus first segment. **b,** 2 micron thick virtual cross section of *Platynereis* 5dpf foregut stained with DAPI to label the cells' nuclei (in red circles are the nuclei of miR-100+ cells). **c,** z-projection of *Platynereis* 5dpf foregut with immunostaining against acetylated tubulin to visualize the axonal scaffold. **d,** z-projection of *Platynereis* 5dpf foregut with combined DAPI (blue) and acTub (green) stainings **e,** 2 micron thick virtual cross section of *Platynereis* 5dpf foregut with medial stomodeal miR-100 staining (acTub green, DAPI blue, miR-100 expression red) **f,** 2 micron thick virtual cross section of *Platynereis* 5dpf foregut with miR-124 staining in foregut's anterior and posterior quadrants (miR-124 expression red) **g,** 2 micron thick virtual cross section of *Platynereis* 5dpf foregut with digestive tract epidermal miR-9 staining (red) **h,** 2 micron thick virtual cross section of *Platynereis* 5dpf foregut with beta-3-tubulin immunostaining in differentiated muscle cells (beta-3-tubulin red) **i,** horizontal virtual sections of different foreguts showing the expression of miR-100, miR-375, *synaptotagmin* (*syt*)(Denes et al., 2007), *prohormone convertase 2* (*phc2*) (Tessmar-Raible et al., 2007), miR-9 and beta-3-tubulin (Leiss et al., 1988) (ventral is up and dorsal is down). **j,** clockwise rotation of the 3D reconstruction of 5dpf *Platynereis* foregut (acTub green, miR-375 red, in lateral view left is dorsal and right is ventral). **k,** 3D reconstruction of 5dpf *Platynereis* foregut (beta-3-tubulin red)

for the maturation of neuropeptides in the vertebrate endocrine system (Seidah and Chrétien, 1999) indicated that these cells are differentiated secretory neurons. *Synaptotagmin* and *prohormone convertase 2* showed a more extensive expression which additionally included all four quadrants of the foregut. However, microRNA marker analysis offered a higher resolution in the molecular dissection of the foregut as microRNAs either showed specific expression in the four quadrants (miR-124 Figure 2.4f) or in the miR-100+, let-7+ cells (never in both cell types). Taking into account the conserved roles of let-7 and miR-125 in developmental timing control in nematode and fly(Pasquinelli et al., 2000; Caygill and Johnston, 2008; Sokol et al., 2008), it is tempting to speculate that the annelid let-7+/miR-125+ cells are neurosecretory cells involved in the control of annelid metamorphosis (see discussion).

Two additional cell types of the foregut were characterized after relating the marker gene expression analysis back to morphological data. The expression pattern of a sensory-associative specific miR-9 (see results on page 78) demarcated very superficial cells right at the digestive tract lining (Figure 2.4g). These epidermal cells are likely to have a sensory identity given their direct contact with ingested food. Immuno-staining for a muscle specific marker, beta-3-tubulin, uncovered the identity of the outer-most vertically positioned cells which seem to encapsulate the foregut (Figure 2.4h). Cross section view (Figure 2.4i) as well as reconstruction (Figure 2.4k) of the foregut counter stained with beta-3-tubulin (Leiss et al., 1988) demonstrate how the muscle layer encapsulates all fore mentioned cell types together and probably facilitates contractions of the foregut structure.

2.2.1.2 MicroRNAs expressed in locomotor ciliated cells

Locomotor ciliated cells represent a defining character of primary larva in various marine invertebrate groups in both protostomes and deuterostomes (Nielsen, 2001). Used for larval swimming and/or larval feeding (Nielsen, 2001), they are typically arranged uniformly in the epidermis or in bands which line key larval body parts in order to create the appropriate water current that will allow locomotion or food uptake. At post-larval/post-metamorphic stages locomotor ciliated cells disappear (Mcdougall et al., 2006).

A subset of conserved bilaterian microRNAs, miR-92, -34 and -29, demarcated the ciliary bands of the *Platynereis* trochophora larva, composed of locomotor ciliated cells, from early larval stages onward (Figure 2.5a-c). The large ciliated prototroch cells are the first cells to differentiate during *Platynereis* development, already at 12hpf to form a ciliary belt around the larva (A. Fischer unpublished data, Figure 2.5d). Expression of miR-92 and miR-29 was likewise detected through northern blot analysis as early as

12hpf (Figure 2.2).

In addition to their ciliary band expression, miR-92, -34 and -29 were also detected in motile ciliated cells of the apical organ (Fig 2.5e), projecting from the crescent cell (Figure 2.5f and position of the apical organ at 48hpf larva in Figure 2.5g). While expression was similar for all 'ciliary' microRNAs during early larval stages, partly complementary patterns were observed after 48hpf and in the developing young worm (Figure 2.5h-j). In the trunk, all three microRNAs marked trunk ciliated cells (arrows in Figure h-k) just like "ciliary" marker *alpha-tubulin*. However, in the head region, aside from demarcating the crescent cell, miR-92 and -29, exhibited a superficial and lateral expression pattern which is typical for sensory brain region-markers (Figure 2.5l, n and see Figure 2.28 on page 102). Only miR-34 showed a closer resemblance in expression pattern to *alpha-tubulin* with overlap in ciliary band and crescent cell expression (Figure 2.5m, o).

Aside to ciliary bands and the crescent cell, two of the ancient bilaterian "ciliary" microRNAs uniquely stained at least one additional structure during the course of development. miR-92 transiently localised in the midline from 48h-72h (Figure2.5h and Figure 2.6a) and at 5dpf it specifically localised in the miR-100+ cells of the foregut and in few cells of the midgut and hindgut (arrows in Figure 2.6e) while its "ciliary" affinity persisted by marking the prototroch (Figure 2.6f). miR-34 retained a very "clean" localisation specific to ciliary bands (with notably strong expression in the prototroch-not observed for miR-92 and -29) throughout development (Figure 2.6 b, h). It was therefore the only of the ciliary microRNAs which showed exclusive expression in ciliary bands throughout development. At 72h, miR-29 showed fainter (yet real) expression in the midline cells compared to miR-92 which on the contrary faintly marked the thinner ciliary bands abutting the stomodeal opening (Figure 2.6c) in which miR-29 was strongly expressed. After 5dpf, miR-29 was localised in the miR-100+ cells of the foregut with stronger expression than miR-92 but restricted to the foregut as no gut expression was detectable (arrows in Figure 2.6g). Ciliated cells in the pharynx have also been reported in the pelagosphera larva of the lophotrochozoan *Phascolosoma agassizii* (Tzetlin and Purschke, 2006).

A similar localisation to ciliary bands (Hill and Boyer, 2003) was observed in one more annelid, *Capitella spI,* for all three "ciliary" conserved bilaterian microRNAs (Figure 2.7a-c). In line with the *Platynereis* results, miR-92 and -29 were also localised in the *Capitella* stomodaeum (arrows in Figure 2.7b, c). I also investigated the same microRNAs in the sea urchin pluteus larva and indeed found miR-92 expressed in the ciliary bands lining the 2 week old pluteus arms and foregut (Figure 2.7d, e; Burke, 1978; Nakajima, 1986) . miR-34 was also localised in ciliated cells lining the 3 week old foregut and stomach (Figure

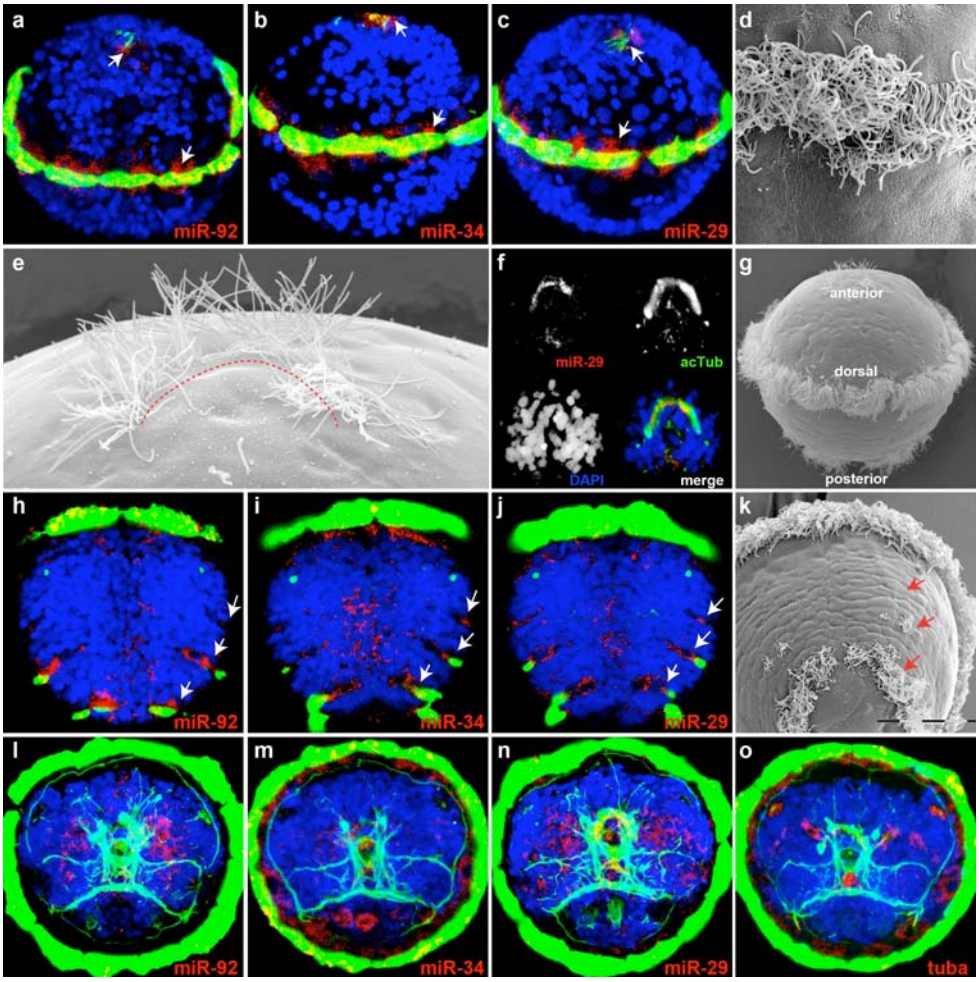


Figure 2.5: **microRNAs expressed in ciliated cells.** **a-c,** ventral-anterior views of 24hpf *Platynereis* trochophora larvae with microRNA expression in prototroch cells and apical organ (arrowheads) (blue: DAPI; green: acTub; red: microRNA). **d,** scanning electron micrograph of the *Platynereis* prototroch cilia at 48hpf. **e,** scanning electron micrograph of the 48hpf *Platynereis* apical organ. **f,** 3micron slice of the apical organ at 48hpf with expression of miR-29 in the crescent cell (red: microRNA; green: acTub; blue: DAPI). **g,** scanning electron micrograph of a 48hpf *Platynereis* trochophora larva (dorsal view). **h,** ventral view of 48hpf Platynereis with miR-92 expression in locomotor ciliated cells of the trunk (arrows) and midline (blue: DAPI; green: acTub; red: microRNA). **i,** ventral views of 48hpf *Platynereis* with miR-34 expression in locomotor ciliated cells (including prototroch) and midline (blue: DAPI; green: acTub; red: microRNA). **j,** ventral view of 48hpf Platynereis with miR-92 expression in locomotor ciliated cells of the trunk (arrows) and midline (blue: DAPI; green: acTub; red: microRNA). **k,** scanning electron micrograph of a 48hpf *Platynereis* depicting the ciliary bands' arrangement on the ventral side of the larva. **l,** apical view of 48hpf Platynereis larval brain with miR-92 staining in crescent cell and lateral brain but also a small group of dorsally positioned cells. **m,** apical view of 48hpf Platynereis larval brain with miR-34 staining in crescent cell and prototroch cells. **n,** apical view of 48hpf Platynereis larval brain with miR-29 staining in crescent cell and lateral brain. **o,** apical view of 48hpf Platynereis larval brain with *alpha-tubulin* staining in all differentiated ciliated cells of the head. Scanning electron micrograph pictures were taken by H. Hausen (e, g) and N. Dray (d, k) .

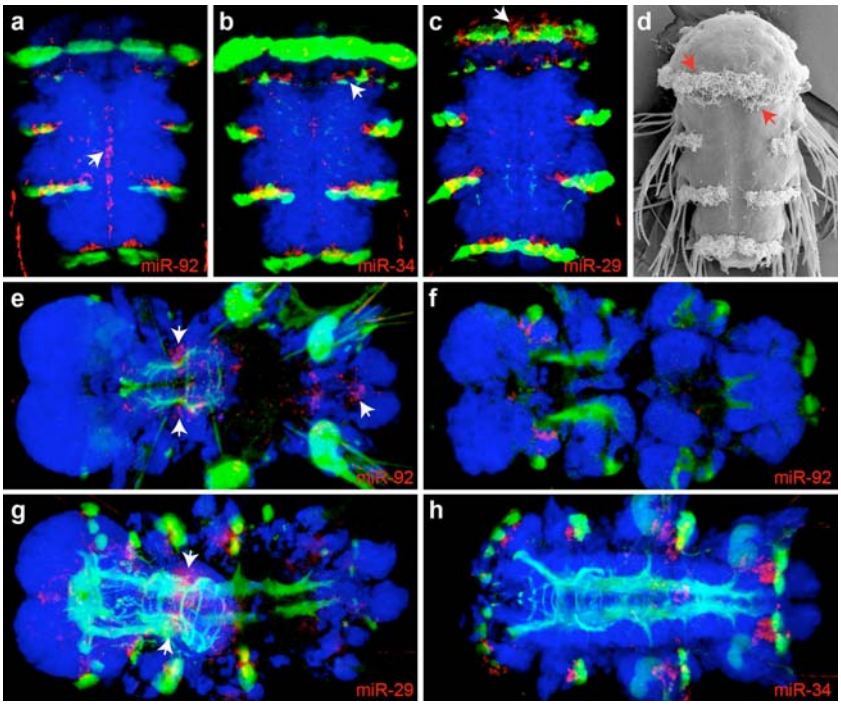


Figure 2.6: **Ciliary microRNA expression later in *Platynereis* development. a,** ventral view of 72hpf *Platynereis* with miR-92 expression in locomotor ciliated cells of the trunk and midline (arrow) (blue: DAPI; green: acTub; red: miR-92). **b,** ventral view of 72hpf *Platynereis* showing miR-34 expression in locomotor ciliated cells of the trunk and in the prototroch (arrow). **c,** ventral view of 72hpf *Platynereis* with expression of miR-29 in locomotor ciliated cells of the trunk. **d,** Transmission electromicrograph of the 72hpf *Platynereis* depicting the arrangement of ciliary bands along the head and trunk. **e,** z-projection of a 24micron slice on the dorsal side of the 5dpf *Platynereis* worm with white arrows pointing at miR-92 expression in the miR-100 positive cells of the foregut and cells in the gut. **f,** 2micron slice on the ventral side of the 5dpf *Platynereis* worm with miR-92 expression in prototroch cells. **g,** dorsal view of 5dpf *Platynereis* with miR-29 expression in ciliated cells of the head, trunk and in the miR-100 positive cells of the foregut. **h,** ventral view of 5dpf *Platynereis* with miR-34 expression in ciliated cells of the head and trunk. The scanning electron micrograph pictures was taken by N.Dray.

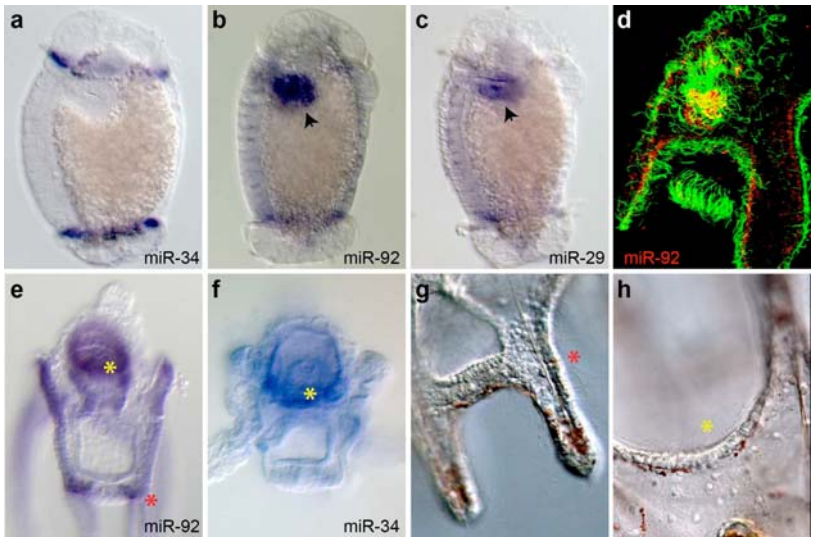


Figure 2.7: **Ciliary microRNA expression in *Capitella spI.* and *Strongylocentrotus purpuratus*. a,** lateral view of stage 6 larva with miR-34 expression in ciliary bands. **b-c,** lateral view of stage 6 larva with miR-92 and miR-29 expression in ciliary bands and stomodaeum (black arrows). **d,** ventral-lateral view of *Strongylocentrotus* pluteus larva with miR-92 expression in ciliated cells lining the foregut and the pluteus arms. **e,** dorsal view of 2 week old *Strongylocentrotus* pluteus larva expressing miR-92 in ciliated cells along the lining of the arms and in mouth and gut. **f,** dorsal view of 3 week old *Strongylocentrotus* pluteus larva expressing miR-34 in ciliated cells of the gut and the arms. **g,** close up of the ciliated pre-oral arms in a living 1 week old *Strongylocentrotus* pluteus larva (by E. Arboleda). **h,** close up of the 1 week old *Strongylocentrotus* ciliated gut (by E. Arboleda).

2.7f). Expression of miR-29 and of control, muscle specific miR-1, was not detected at any of these stages (data not shown).

To ensure that the identity of the sea urchin ciliated cells is indeed locomotory, living sea urchin plutei larvae were observed and documented under white light microscopy with the help of E. Arboleda (Figure 2.7g, h, coloured asterisks indicate the equivalent structures in fixed material of panels e, f). The 2 weeks old pluteus larva epidermis is covered in single motile cilia. However, there is a clear distinction between these and the condensed patches of multiciliated cells arranged in locomotor bands which can steer swimming (Wada et al., 1997) and control feeding behaviour (Strathmann, 1971). All microRNA-expressing ciliated cells belonged to such condensed ciliary patches. The similar specific affiliation of these microRNAs to locomotor ciliated cells in both protostome and deuterostome larvae strongly suggests that this is an ancient bilaterian trait that existed already in the protostome/deuterostome ancestor.

2.2.1.3 Demarcating ancient brain centres

Two sets of conserved bilaterian microRNAs showed localised expression in defined parts of the brain; the neurosecretory and the sensory-associative brain centres.

The neurosecretory brain centre Previous work done by K. Tessmar-Raible, who thoroughly studied the neurosecretory brain part of *Platynereis*, linked the annelid's medial brain to the vertebrate hypothalamus (Tessmar-Raible et al., 2007). The overall arrangement of the developing neurosecretory centres of zebrafish and *Platynereis* share similarities with respect to their "molecular topography" and their early axonal scaffolds. As already reported for several vertebrates (Corbin et al., 2003; Murakami et al., 2001), subdivision of the forebrain by transcription factors *nk2.1* and *pax6* into medial and lateral respectively is also observed in *Platynereis* (Fig 2.8 scheme a). In vertebrates the *nk2.1+* region gives rise to the preoptic region and the hypothalamus (Varga et al., 1999)and K. Tessmar-Raible identified the fraction of *nk2.1+* cells with neurosecretory identity. In both zebrafish and *Platynereis*, these cells additionally express transcription factors; *rx*, a marker of the developing zebrafish forebrain (Chuang et al., 1999), and *otp*, necessary for the terminal differentiation of a subset of neuropeptidergic hypothalamic neurons in the mouse brain (Acampora et al., 1999).

The *nk2.1/rx/otp* + medial region of zebrafish and *Platynereis* brains (Figure 2.8b) hosts, among other differentiating neurosecretory cells, light sensitive vasotocinergic neurons. In *Platynereis*, at 48hpf, vasotocinergic cells are positioned medially, deep in the brain (Figure 2.8c red outlines), adjacent to the ciliary photoreceptors (extraocular photoreceptors of the larva; Arendt et al. 2004). More anteriorly, in a *nk2.1/otp+* but *rx-* region, chemosensory FMRFamidergic cells are situated medially but more superficially in the brain. FMRFamidergic cells bear dendrites with sensory cilia that reach the surface of the subcuticular extracellular space (Fig 2.8d EM reconstruction).

A stronger molecular link between vasotocinergic and FMRFamidergic cells was established after I investigated the localisation of conserved bilaterian miR-7, already known for its highly specific expression in the medial forebrain of zebrafish (Kapsimali et al., 2007; Wienholds et al., 2005) and medaka (Ason et al., 2006). *Platynereis* miR-7 expression was restricted to differentiating neurosecretory tissue (Figure 2.8e and Figure 2.9 on page 76b), demarcated by marker gene *phc-2* (Figure 2.9 on page 76a), and comprised both vasotocinergic (yellow arrows pointing the ciliary photoreceptors adjacent to vasotocinergic cells in Figure 2.8f, h) and FMRFamidergic cells (white arrows in Figure 2.8g, i). I localized the expression of miR-7 in zebrafish and K. Tessmar-Raible spatially correlated their restricted population to the axonal scaffold of the 37hpf forebrain con-

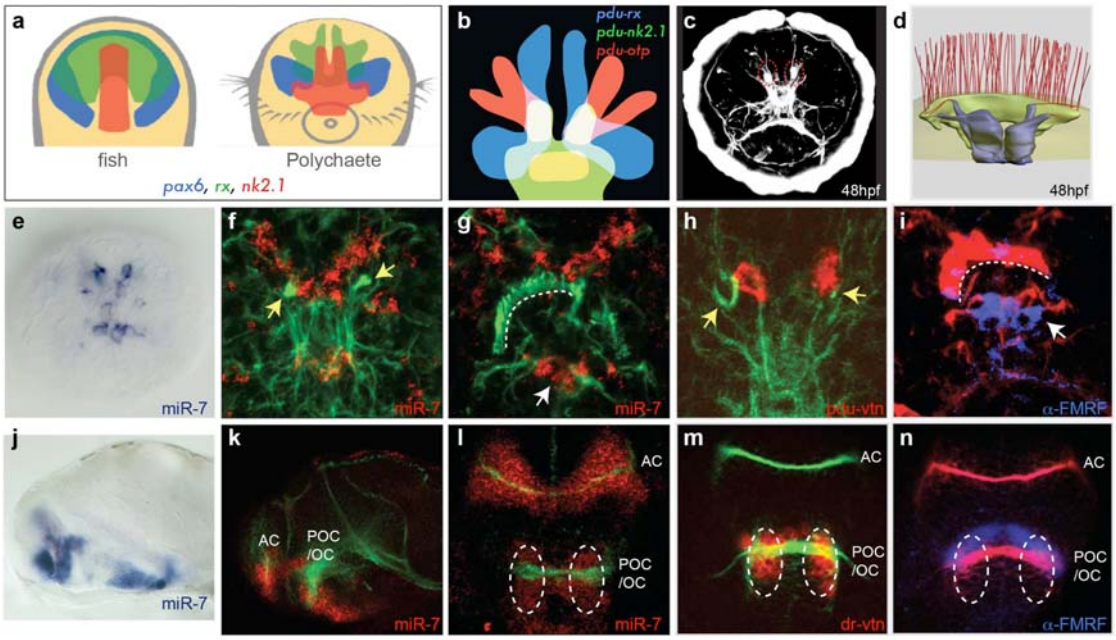


Figure 2.8: **Conserved neurosecretory cell types in *Platynereis* and zebrafish forebrain.** **a,** schematic view of *pax6* (blue), *rx* (green) and *nk2.1* (red) expression patterns in the zebrafish and *Platynereis* forebrain. **b,** schematic representation of the intersection (white) between the expression patterns of *pdu-rx* (blue), *pdu-nk2.1* (green) and *pdu-otp* (red) where the medial *vtn+* cells are found in the *Platynereis* 48hpf brain (outlined in red in panel c). **c,** apical view of 48hpf *Platynereis* larva immunostained for acetylated tubulin with dashed red circles highlighting the location of the studied vasotocinergic cells in comparison to the axonal scaffold. **d,** reconstruction of the 48hpf FMRFamidergic cellular arrangement from transmission electromicrographs depicting their superficial position related to the apical organ and their apical dendrites' cilia which protrude in the subcuticular extracellular space. **e,** apical view of 48hpf *Platynereis* larva with miR-7 expression in dorso-medial brain. **f,** apical view of 48hpf larva with miR-7 expression (red) in relation to the axonal scaffold (green: acTub) in the deep brain region where ciliary photoreceptors reside (yellow arrows). **g,** apical view of the 48hpf *Platynereis* larva with miR-7 expression (red, FMRFamidergic cells pointed by white arrow) in relation to the axonal scaffold and cilia (green: acTub) of the apical organ at the anterior most level of the larval brain (cilia of crescent cell underlined by white dashed line). **h,** apical view of the 48hpf *Platynereis* larval brain with the axonal scaffold (green: acTub) of the neurosecretory region around ciliary photoreceptors (yellow arrows) and *pdu-vasotocin* (*pdu-vtn*) expression in red. **i,** apical view of 48hpf *Platynereis* larva immunostained for FMRF amide (blue: FMRFamidergic cells, pointed by white arrow) and acetylated tubulin (acTub: red) depicting the position of FMRFamidergic cells in relation to the axonal scaffold and cilia (underlined by white dashed line) of the apical organ's crescent cell. **j,** lateral view of 3dpf zebrafish head with miR-7 expression in the forebrain. **k,** lateral view of the 37hpf zebrafish brain with miR-7 expression (red) in relation to the axonal scaffold (green: acTub). **l-m** ventral view of 37hpf zebrafish brain with miR-7 ,and *dr-vasotocin* (*vtn*) expression respectively (red) in relation to the axonal scaffold (green: acTub). **n,** ventral view of the 37hpf zebrafish brain immunostained for FMRFamide (blue) and acetylated tubulin (red). AC, anterior commissure; POC/OC, post-optic commissure/ optic chiasm. All miR-7 WMISH were done by me and all pictures except *Platynereis* miR-7 were taken by K. Tessmar-Raible.

firming our findings in *Platynereis* (all fish confocal scans taken by K. Tessmar-Raible in Figure 2.8k-l). Zebrafish vasotocinergic neurons co-expressed miR-7 (compare dashed outlines between Figure 2.8l and m) and the miR-7 forebrain expression also comprised FMRFamidergic neurons (compare dashed outlines between Figure 2.8l and n). Through this comparative analysis of the miR-7 expression, the molecular fingerprint of these two types of neurosecretory cells was enriched by an additional marker and the evidence pointing towards a conserved molecular anatomy in the region which gives rise to important neurosecretory forebrain centres was strengthened. In the history of comparative studies, it was in the publication of Tessmar-Raible et al. (2007) where for the first time a microRNA was used as a marker to uncover conserved cell types between distantly related species such as annelid and fish.

Aside from miR-7, the *Pdu-phc-2+* neurosecretory tissue (Figure 2.9a) co-expressed another two additional microRNAs; miR-137 and -153. The latter, which are not related by sequence, showed very similar restricted expression in the dorso-medial 48hpf *Platynereis* brain just like miR-7 (Figure 2.9b-d). All three microRNAs were expressed in differentiated vasotocinergic and FMRFamidergic (Figure 2.8f-i and Tessmar-Raible et al., 2007) as well as serotonergic neurons (Figure 2.9e, i). As opposed to other paneural microRNA markers (see miR-71 in Figure 2.13 on page 80), the localisation of miR-137, -153 and -7 remained restricted to a subset of the brain even at later developmental stages. At 5dpf, neurosecretory cells of the dorsal brain, the nuchal organ and the adult eyes co-expressed all three microRNAs (Figure 2.9f-h). Mushroom bodies and palpae additionally expressed miR-7 which showed a more expanded pattern in the 5dpf brain (white arrows in Figure 2.9h). miR-137 was also localised in serotonergic neurons of the trunk nervous system (Figure 2.9i) and few cells in the miR-124+ region of the foregut (white arrows in Figure 2.9j and see Figure 2.4 on page 67f). miR-7 and -153 were exclusively expressed in the brain (Figure 2.9k-l).

miR-7, -137 and -153 also showed brain-restricted expression in *Capitella* (Figure 2.10a-c) and have already been reported to show spatially localised expression in the zebrafish brain, including neurosecretory brain parts of the hypothalamus (Kapsimali et al., 2007 and Table 2.4 on page 89). These comparative data thus indicate that all three microRNAs co-evolved in neurosecretory tissue in an ancient bilaterian brain.

The sensory-associative brain centre A conserved pair of complementary microRNAs, miR-9 and miR-9*/miR-131 (Deo et al., 2006), also showed highly restricted expression in the annelid brain, in two bilaterally symmetrical, ventro-lateral sets of differentiated neurons (Figure 2.11a-c). In mouse, miR-9 and miR-9*/miR-131 are detected broadly

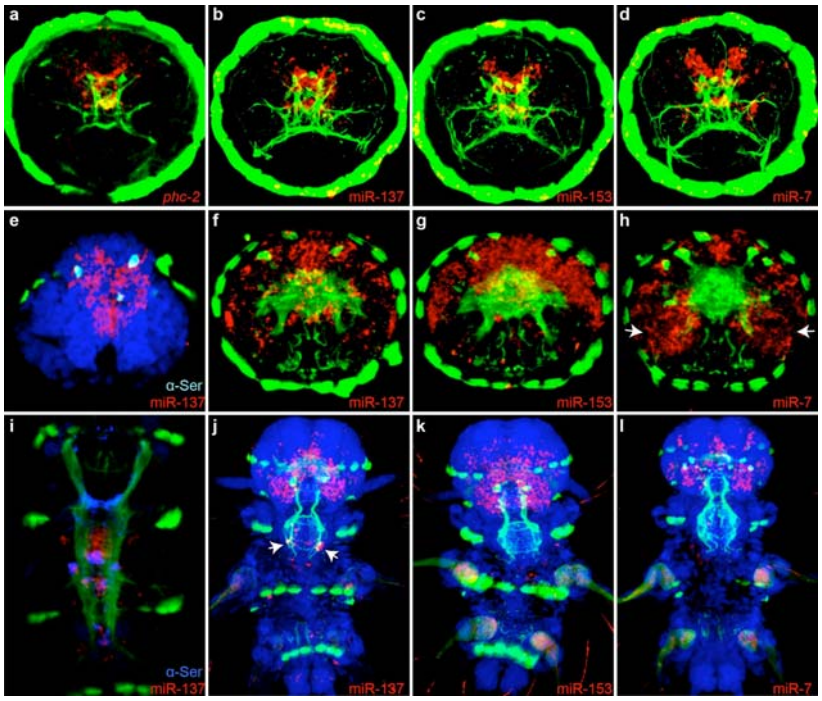


Figure 2.9: **Expression of neurosecretory tissue specific microRNAs. a-d,** apical view of 48hpf *Platynereis* larval brain with *prohormone convertase* (*phc2*), miR-137, miR-153 and miR-7 expression in dorso-medial brain (green; acTub, red;gene or microRNA). **e,** apical view of 72hpf *Platynereis* brain section showing co- localisation of miR-137 with serotonin (blue: DAPI, green: acTub, red: miR-137, cyan: serotonin). **f-g,** apical view of 5dpf *Platynereis* worm brain with miR-137 and miR-153 expression in neurosecretory tissue, adult eyes and nuchal organ. **h,** miR-7 expression in the 5dpf worm brain comprising neurosecretory tissue, adult eyes, nuchal organ, mushroom bodies and palpae. **i,** ventral view of the 5dpf *Platynereis* trunk axonal scaffold depicting the co- localisation of miR-137 with serotonin (blue: serotonin, green: acTub, red: miR-137). **j,** dorsal view of 5dpf *Platynereis* worm with miR-137 expression in dorsal brain and few cells in the foregut. **k-l,** dorsal view of 5dpf *Platynereis* worm with miR-153 and miR-7 expression in dorsal brain.

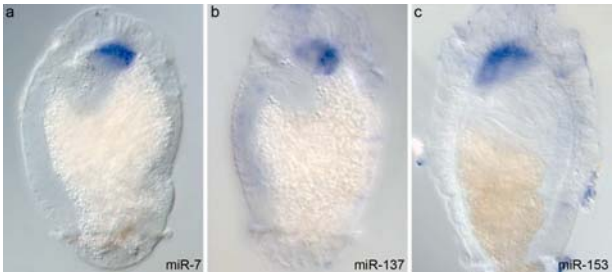


Figure 2.10: **Expression of neurosecretory tissue specific microRNAs in *Capitella spI*. a-c,** lateral view of stage 6 *Capitella* swimming larva with miR-7, miR-137 and miR-153 expression restricted in the brain.

in neuronal precursors but among differentiated neurons they are expressed only in the telencephalon (Deo et al., 2006), the olfactory brain centre.

In *Platynereis*, the most apical cells expressing mature miR-9 and miR-9* were located at the base of the antennae (Figure 2.11d), a pair of head appendages considered chemosensory sense organs (p.58 in Purschke, 2005). Expression was not only localised superficially in the brain but extended deeper to reach few cells of the developing mushroom bodies of the 5dpf brain (shaded red structures in Figure 2.11 on the next pagec identified and characterized by R. Tomer). Addional head expression was detected in the nuchal organ (arrows in Figure 2.11e-g), a dorsally positioned ciliated epithelial thickening with bipolar sensory cells (Purschke et al., 1997), which shows structural analogies to arthropod olfactory sensory organs (Schlotzer-Schrehardt, 1986). Lastly, miR-9 and -9* showed expression in sensory cells (p.58 inPurschke, 2005) of the pharyngeal epithelium (Figure 2.11e-f and Figure2.4g). In closely related *Capitella*, miR-9* was detected in a subset of the brain and in single sensory cells of the epidermis (Figure 2.11h).

The above findings indicate that miR-9 and -9* may play a conserved role in defining neurons involved in some sort of olfactory/chemosensory information processing. As in vertebrates, both miR-9 and miR-9* are detected in relatively high abundance in *Platynereis* (see Table 2.1) and thus represent a pair of complementary microRNAs conserved in bilaterians (Wheeler et al., 2009).

2.2.1.4 Defining four core organ systems

A large fraction of the conserved bilaterian core microRNAs was expressed more broadly in one of four major organ systems, representing the central nervous system (Figure 2.12a-d), peripheral sensory tissue (Figure 2.12e-h), musculature (Figure 2.12i-m) or gut (Figure 2.12n-q). For each of these four groups, expression of individual microRNAs was largely overlapping, but at the same time complementary to those belonging to other groups.

Central Nervous System miR-124 (Figure 4a), known to maintain neuronal identity in the vertebrates (Farh, 2005; Kapsimali et al., 2007) and expressed in developing CNS in fly (Aboobaker et al., 2005) and planarian (Gonzalez-Estevez et al., 2009), was found in differentiating neurons in the brain and ventral nerve cord in *Platynereis*, as evidenced by *synaptotagmin* co-expression (Figure 2.12c). miR-71, lost in vertebrates and insects (Wheeler et al., 2009) and restricted to CNS and parenchyma in planarians (Gonzalez-Estevez et al., 2009), also proved nervous system-specific in *Platynereis* (Figure 2.12b; Fig, 2.13b). While both microRNAs were expressed in differentiating neurons, a clear

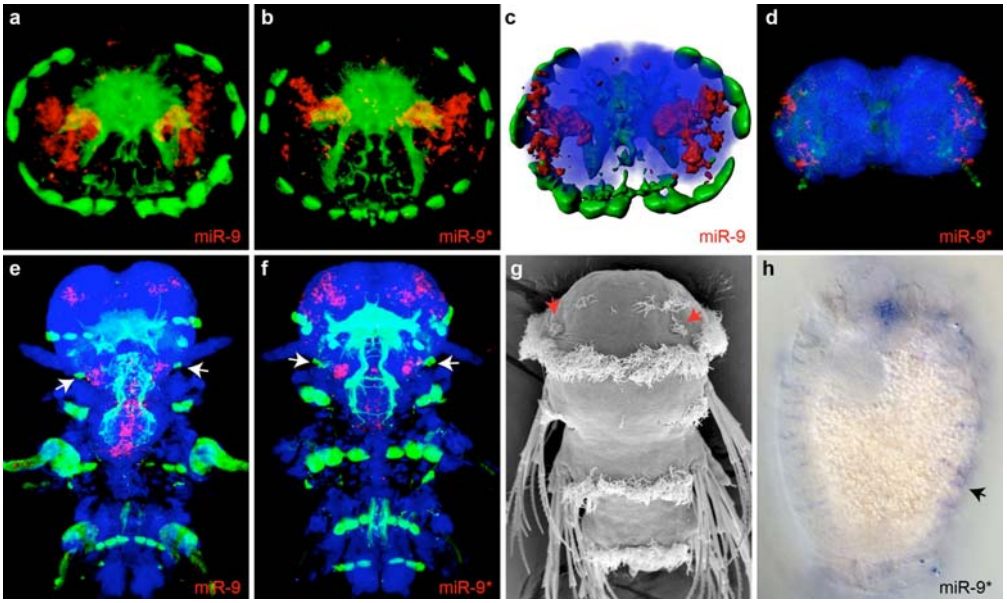


Figure 2.11: **Expression of microRNAs in sensory-associative brain centres. a-b,** apical views of 5dpf *Platynereis* brain showing the expression of miR-9 and miR-9* (green: acTub, blue: DAPI, red: miRNA). **c,** 3D reconstruction of 5dpf apical brain with superficial miR-9 expression at the base of the antennae and deeper brain expression in mushroom bodies. **d,** apical view of 5dpf *Platynereis* anterior brain section (6 micron thick) depicting the very superficial miR-9*+ cells and their antennal projections. **e-f,** dorsal views of 5dpf *Platynereis* expressing miR-9 and -9* in brain, nuchal organs (white arrows) and pharyngeal epidermis in foregut. **g,** Transmission electromicrograph of the 72hpf *Platynereis* dorsal view with red arrows pointing at nuchal organs (by N.Dray). **h,** lateral view of stage 6 *Capitella* swimming larva with miR-9* expression in the brain and single cells of epidermis (black arrow).

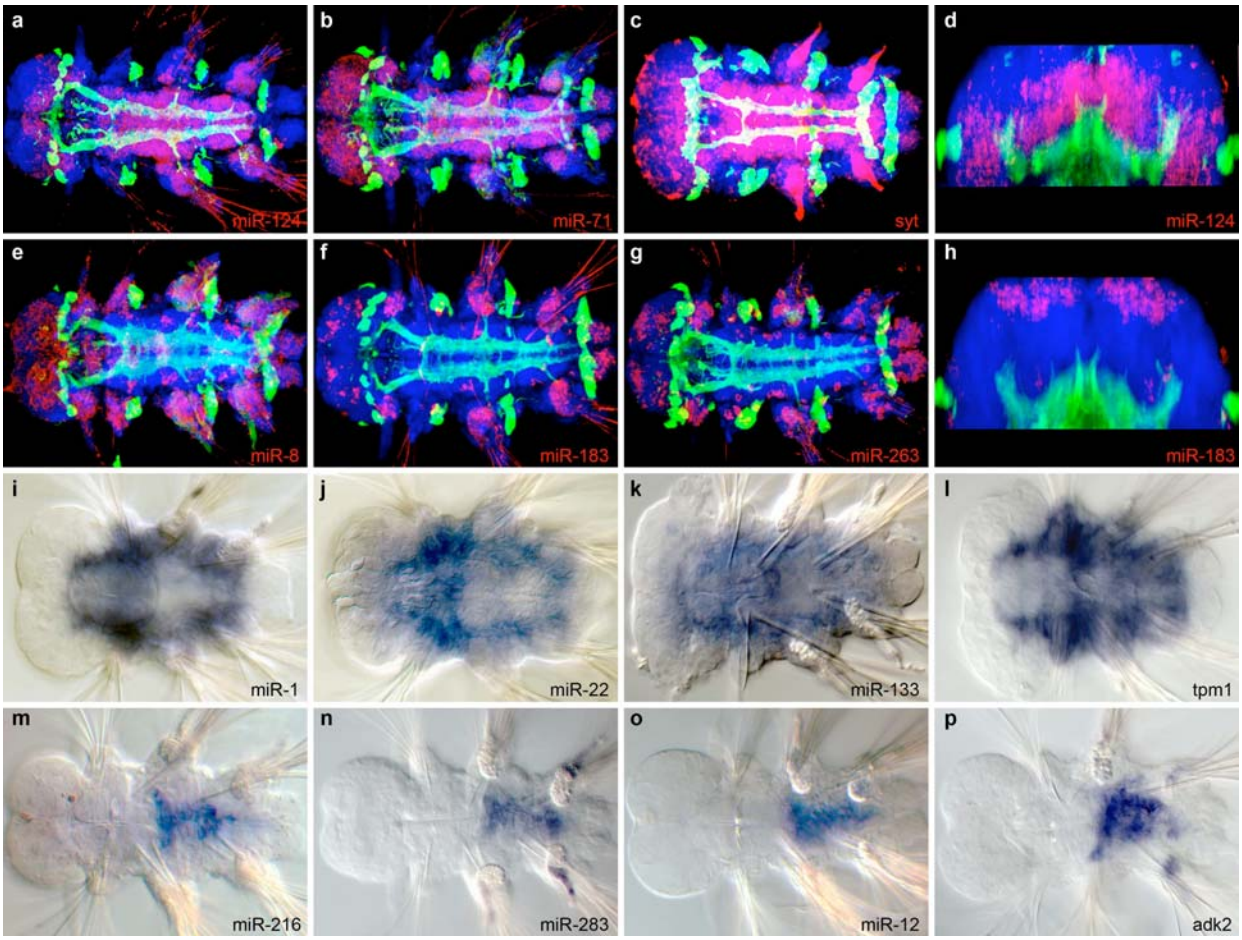


Figure 2.12: **microRNAs demarcating organ systems. a-b,** ventral views of 5dpf worms with restricted expression of miR-124 and -71 in the CNS (blue: DAPI, green: acTub, red: microRNA). **c,** *synaptotagmin* (*syt*) expression at 5dpf ventral view. **d,** virtual section of 5dpf *Platynereis* brain exemplifying the basal expression of CNS-specific microRNAs. **e-g,** ventral views of 5dpf worms with restricted expression of miR-8, -183 and -71 in the PNS (blue: DAPI, green: acTub, red: microRNA). **h,** section of 5dpf *Platynereis* brain exemplifying the apical expression of PNS-specific microRNAs. **i-m,** ventral views of 72hpf *Platynereis* expressing miR-1, -22, -133 and *tropomyosin 1* (*tpm1*) in differentiated musculature. **n-q,** dorsal views of 5dpf *Platynereis* expressing miR-216, -283, -12 and a gut marker, *adenosine kinase 2* (*adk2*) in the differentiating midgut.

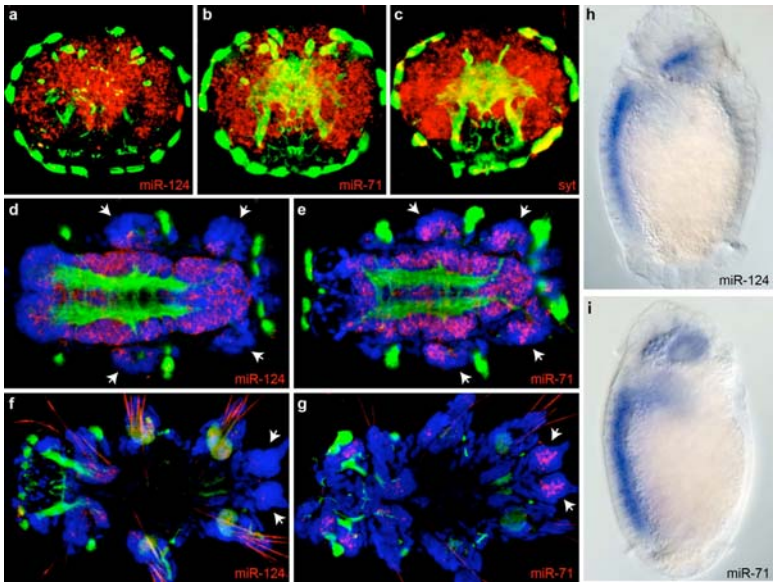


Figure 2.13: **Paneural microRNAs. a-c,** apical view of 5 day old brains with miR-124 restricted expression versus miR-71 and *synaptotagmin* (*syt*) whole brain expression (green; acTub, red; miRNA). **d-e,** ventral view of 5dpf *Platynereis* worm with miR-124 expression restricted to the trunk nervous system and miR-71 additional expression in cells of the parapodia pointed by white arrows. **f-g,** ventral view of 5dpf worm with differential expression of miR-124 and miR-71 in the pygidial lobe. **h-i,** lateral view of stage 6 *Capitella* swimming larva with miR-124 and -71 expression in the brain and ventral nerve cord.

distinction could be made between miR-124 and miR-71 in that the latter was expressed more broadly in all types of neurons (including those of the peripheral nervous system) just like *synaptotagmin*. In the head, miR-71 was additionally expressed in the nuchal organs, a dorsally positioned group of sensory cells and in palpae therefore attributing a broader pattern (Figure 2.13b). In the trunk, miR-71 was expressed in sensory cells of the parapodial epidermis and the pygidial lobe again resulting in a broader pattern to that of miR-124 which was solely expressed in neurons of the CNS (Figure 2.13f-g). In *Capitella* both microRNAs demarcated the brain and the ventral nerve cord albeit miR-71 showed additional expression in the stomodaeum (Figure 2.13h-i).

Spatially restricted localisation in the *Platynereis* CNS (in a subset of differentiating neurons of the brain and the trunk) was observed for miR-184 which in chicken and mouse is expressed in the brain and in corneal epidermis (Darnell et al., 2007; Nomura et al., 2008; Ryan et al., 2006). miR-190, also found to be expressed in the zebrafish eye (Kloosterman, 2006) resembled a lot the miR-184 expression pattern (Figure 2.14b-c). In *Platynereis* both microRNAs' expression largely coincided with glutamatergic neurons

as revealed by *glutamate transporter 1 (glt1)* expression (Figure 2.14a-c and e-g). In the 5dpf brain, glutamatergic neurons are positioned very deeply at the anterior end of the brain axonal scaffold (Figure 2.14i). miR-184 brain expression included some additional neurons of the ventral brain (Figure 2.14f) aside from the glutamatergic neurons (Figure 2.14j). On the contrary, miR-190 was detected in additional neurons of the dorsal brain (Figure 2.14g) aside from the glutamatergic cells.

miR-219, a brain specific microRNA in both zebrafish and mouse (Kapsimali et al., 2007; Wienholds et al., 2005; Cheng et al., 2007), showed in *Platynereis* the most restricted localisation compared to any other CNS "subset" marking microRNA. It proved specific for two unknown neurons at the end of the 5dpf nerve cord and another two neurons positioned in the ventral brain (Figure 2.14d, h). The identity of miR-219+ cells was revealed through FVRI-amide immunostainings which among other cells also highlighted the restricted miR-219 expression pattern both in the trunk (Figure 2.14k) and the brain (Figure 2.14l).

Sensory organs At 5dpf, differentiating sensory organs of the young *Platynereis* worm include the eyes (red pigmented cells in Figure 2.15a), the cirri and pygidial lobe (white and red arrows respectively in Figure 2.15b), the antennae and palpae (red and black arrows respectively in Figure 2.15c and stained structures in d-e), sensory superficially positioned cells of the ventral nerve cord and parapodia (Figure 2.15f) and dorsal superficial sensory cells of the head and the trunk (Figure 2.15g). Most of the aforementioned sensory structures express the gene marker *distaless* (*pdu-dlx*) (Figure 2.15h). In *Platynereis* the distaless expression pattern harbors single cells positive for sensory markers *atonal* (*pdu-ath*) and TRP-family cation channel (*trpv*) (Denes et al., 2007) like already reported in fly and vertebrates (Boekhoff-Falk, 2005).

miR-8 proved to be an excellent marker for sensory organs as it demarcated the antennae, palpae, cirri, sensory cells of trunk and parapodia, the pygidial lobe, covering both neural and non-neural tissue (Figure 2.12e and Figure 2.15d-e). Aside from marking peripheral sensory organs, miR-8 showed additional strong expression in the nuchal organ and the foregut where its expression overlapped completely with miR-100, miR-9 and only partially with miR-124. LNA probes also detected miR-8*, albeit more weakly, in a similar head pattern (Figure 2.15i). A sensory tissue affiliation is likely to be evolutionarily ancient for these microRNAs since the related miR-200a, miR-200b and miR-141 likewise show restricted expression in sensory organs such as nose and lateral line in the vertebrates (Table 2.4 on page 89 and Wienholds et al., 2005). The sensory tissue affinity of miR-8 was confirmed in another lophotrochozoan, *Capitella SpI*, where expression was

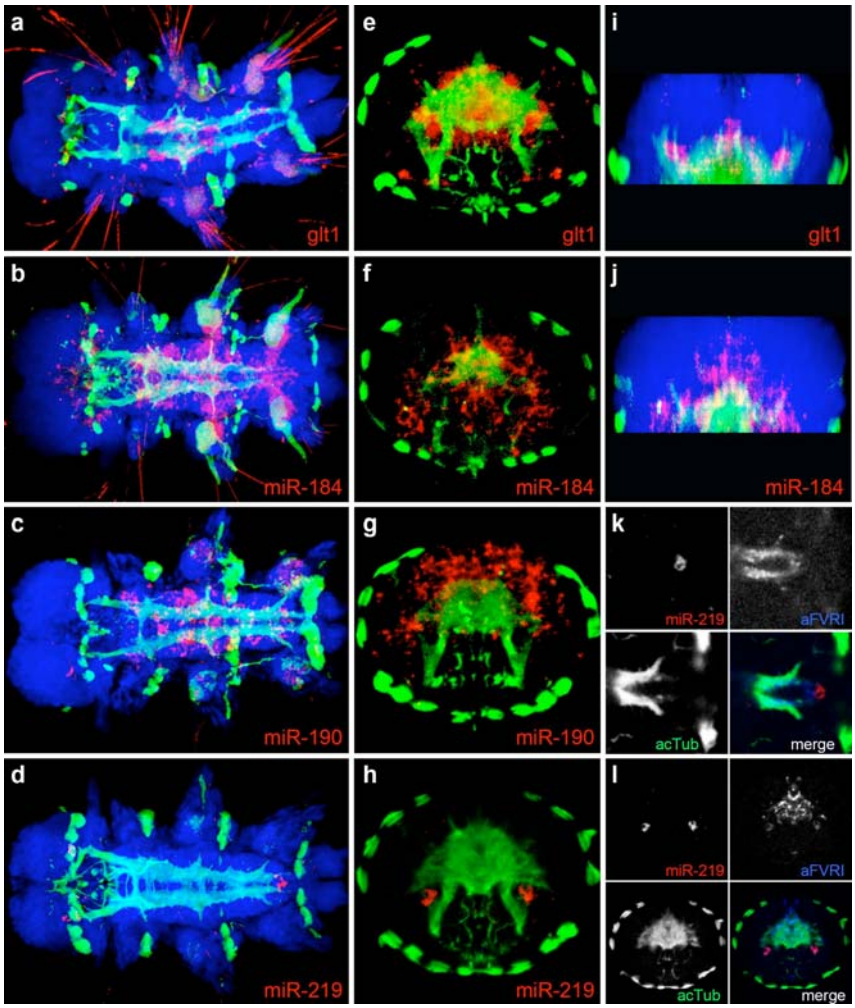


Figure 2.14: **microRNAs expressed in subset of CNS. a-c,** miR-184, miR-190 and *glutamate transporter 1* (*glt1*) show shared expression sites in subsets of the ventral nerve cord including glutamatergic neurons (blue; DAPI, green; acTub, red; miRNA). **d,** ventral view of 5dpf *Platynereis* with highly specific miR-219 expression in few cells of the CNS. **e-g,** apical view of 5dpf brains with miR-184, miR-190 and glt1 sharing parts of their expression domains. **h,** apical view of 5dpf brain with exclusive expression in very few cells of the ventral brain. **i-j,** virtual vertical sections of 5dpf *Platynereis* brains exemplifying the basal expression of *glutamate transporter 1* (*glt1*) and the largely overlapping miR-184. **k,** 2micron vertical slice of the 5dpf posterior segment with expression of miR-219 in two FVRI-amidergic cells of the ventral nerve cord (red: microRNA; blue: αFVRI; green: acTub). **l,** 2micron horizontal slice of the 5dpf brain with expression of miR-219 in few FVRIamidergic cells of the ventral brain (red: microRNA; blue: αFVRI; green: acTub).

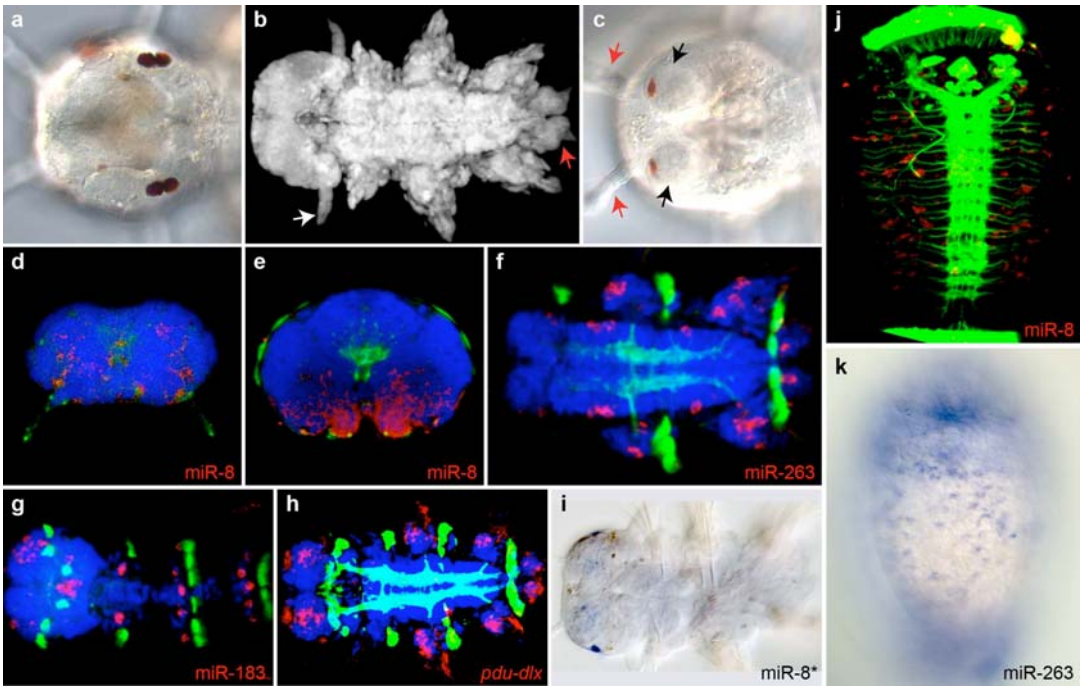


Figure 2.15: **microRNAs expressed in sensory organs of *Platynereis*.** **a,** dorsal view of 5dpf head with focal plane on red pigmented eyes. **b,** ventral view of 5dpf worm stained with DAPI to label the cells' nuclei (white arrows: cirri; red arrows: pygidial lobe). **c,** ventral view of 5dpf head with antennae at the anterior (red arrows) and palpae projecting structures ventrally (black arrows). **d,** apical view of 5dpf *Platynereis* anterior brain section (2 micron thick) depicting the very superficial miR-8+ cells and their antennal projections (green: acTub, blue: DAPI, red: miR-8). **e,** apical view of 5dpf deep head cross section at the level of palpae (demarcated in red by miR-8 expression; green: acTub, blue: DAPI). **f,** ventral view of 5dpf trunk superficial slice (10micron thick) depicting single sensory cells of the nerve plate an parapodia (green: acTub, blue: DAPI, red: miR-263). **g,** dorsal view of 5d worm with miR-183+ sensory cells of the head (dorsal sensory organ) and trunk (green: acTub, blue: DAPI, red: miR-183). **h,** ventral view of 5dpf worm with *distaless* expression in the PNS (green: acTub, blue: DAPI, red: *pdu-dlx*). **i,** ventral view of 5dpf worm with miR-8* expression in palpae and antennae. **j-k,** ventral view of stage 6 *Capitella* larvae with miR-8 and miR-263 in single cells of epidermis (green: acTub, red: miRNA).

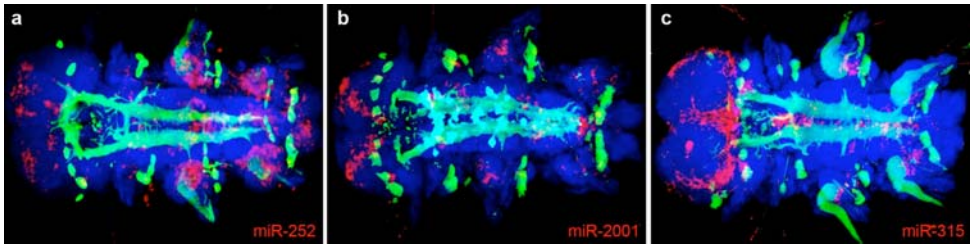


Figure 2.16: **Other ancient bilaterian microRNAs with similar sensory expression. a,** ventral view of 5dpf *Platynereis* with miR-252 expression in antennae, palpae, sensory cells of parapodia, pygidial lobe and gut. **b,** ventral view of 5dpf worm with miR-2001 expression in antennae, palpae, sensory cells of parapodia, few neurons on the ventral nerve cord and gut. **c,** ventral view of 5dpf worm with miR-315 expressed in antennae, palpae, head epidermis and midline of the ventral nerve cord (green: acTub, blue: DAPI, red: microRNA).

localised in very superficially positioned single cells of the epidermis (Figure 2.15).

miR-183 and miR-263, clustered in the *Lottia* genome and processed from the same transcript in *Platynereis* (Table 2.3) also showed a conserved affiliation with sensory organs differentiation (Figure 2.12f-h and Figure 2.15f-g). This has been as previously reported for various other bilaterians (Kapsimali et al., 2007; Aboobaker et al., 2005). Compared to miR-8, the two sequence related microRNAs exhibited a more restricted expression solely in superficial/peripheral sensory cells and organs (i.e. no expression in foregut- compare Figure 2.12e to f-g). Notably, the expression of the sensory tissue-specific miR-183 and of the CNS-specific miR-124 was mutually exclusive (compare Figure 2.12d and h).

Very similar "sensory" patterns were detected for two additional ancient microRNAs which were lost in the vertebrate lineage (Wheeler et al., 2009). miR-252 and -2001 also localized in the antennae, palpae and sensory cells of the parapodia (Figure 2.16a-b). Besides their affinity for sensory organs, both microRNAs were additionally expressed in the gut. Unlike miR-252, miR-2001 showed no expression in the pygidial lobe but was additionally expressed in very few cells of the foregut and several superficially positioned cells of the ventral plate. Up to date, no comparative expression data are available for neither miR-2001 nor miR-252. Only one study indirectly suggests a "neural" affinity for miR-252 which was computationally identified in genetic loci required for synaptic structure and function in *Caenorhabditis* (Sun et al., 2006).

miR-315, also lost in the vertebrate lineage (Heimberg et al., 2008; Wheeler et al., 2009), showed a very unique pattern in *Platynereis* making it hard to group it together with other microRNAs. Like all the above mentioned sensory organ microRNA markers, miR-315 was expressed in the antennae, the dorsal sensory organ of the head, the palpae

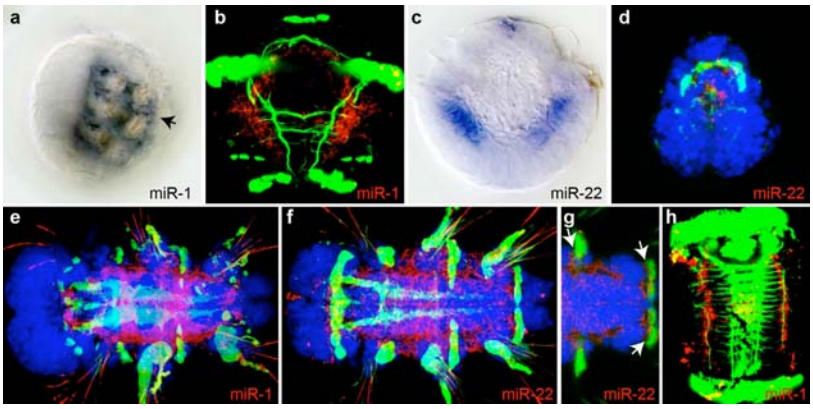


Figure 2.17: **microRNAs expressed in musculature. a,** lateral view of 48hpf *Platynereis* larva with miR-1 expression in longitudinal (vertical) and dorso-ventral (horizontal, black arrow) muscles. **b,** ventral view of 48hpf larva with miR-1 expression in differentiating muscles and their arrangement compared to the axonal scaffold (green: acTub, blue: DAPI, red: miR-1). **c,** ventral view of 48hpf larva with miR-22 expression in musculature and ciliated cells of the head. d, apical cross section of the 48hpf brain with miR-22 expression in ciliated cells of the apical organ. **e,** ventral view of 5dpf worms with miR-1 expression in longitudinal and dorsoventral muscles of the trunk. **f,** ventral view of 5dpf worm with miR-22 expression in musculature and ciliated cells of the trunk and head. **g,** 5micron slice of 5dpf ventral view with miR-22 expression in ciliated cells of the trunk. **h,** ventral view of stage 6 *Capitella* larva with miR-1 expression in differentiated muscle.

but it was not expressed in other sensory structures such as the cirri, the pygidial lobe or sensory cells of the trunk and parapodia. In the head, miR-315 was additionally expressed in nuchal organ, the adult eyes and in head epidermis. In the trunk, expression was exclusive for midline neurons (Figure 2.16c). In *Drosophila* miR-315 is similarly expressed in the brain and a subset of the ventral nerve cord (Aboobaker et al., 2005).

Musculature miR-1 and miR-133, genomically clustered in vertebrates and in mollusks yet not sequence related (Table 2.3) showed almost identical expression in the differentiating musculature (Figure 2.12i, l). The expression pattern of miR-1, one of the most abundant microRNAs in *Platynereis* (Table2.1), was detectable in longitudinal as well as dorsoventral muscles from the larval stage of 48hpf onwards (Figure 2.17a-b, e). miR-1 showed a similar restricted expression in larval musculature of closely related annelid *Capitella spI* (Figure 2.17h). Less abundant miR-133 (Table 2.1), was not detectable in larval stages but was likewise expressed in the differentiated longitudinal and dorsoventral muscles of the 72hpf worm (Figure 2.12l). In vertebrates, these microRNAs are robustly induced upon myotube differentiation concomitant with reduced expression of their target messengers (Farh, 2005; Rao et al., 2006; Sood et al., 2006).

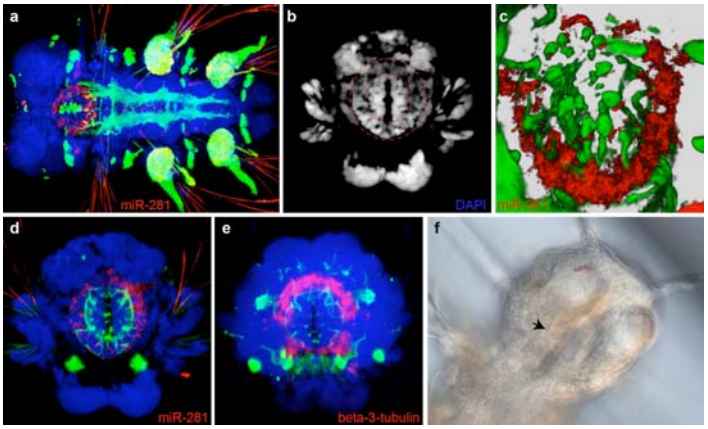


Figure 2.18: **miR-281 is expressed in a subset of musculature. a,** ventral view of 5dpf *Platynereis* worm with miR-281 expression in ring structure below the mouth and in nuchal organ (green: acTub, blue: DAPI, red: miR-281). **b,** 1micron thick apical view of the 5dpf *Platynereis* worm stained with DAPI to label the cells' nuclei (miR-281+ cells highlighted by red dashed lined ring). **c,** 3D reconstruction of 5dpf stomatogastric nervous system (green: acTub) with apically stemed mouth axons innervating vertically the circularly arranged miR-281+cells (red). **d,** apical view of 5dpf worm with miR-281 expression in innervated cells which form a ring between the mouth opening and the foregut (green: acTub, blue: DAPI, red: miR-281). **e,** apical view of 5dpf worm with beta-3-tubulin expression in innervated cells which form a ring between the mouth opening and the foregut (green: acTub, blue: DAPI, red: miR-281). **f,** ventro-lateral view of 5dpf *Platynereis* worm highlighting the mouth ring muscle which is visible under white light.

Similar musculature-specific expression was observed for miR-22 (Figure 2.12k) that together with miR-1 and -133 has been reported to have myoD and myogenin upstream binding sites in the vertebrates (Rao et al., 2006). miR-22 showed an equally early expression as miR-1 and in like manner demarcated both dorsoventral and longitudinal larval muscles (Figure 2.17c). However, together with differentiating musculature, miR-22 was also detected in a small group of cells in the anterior brain (Figure 2.17c). Careful examination revealed the identity of these cells which bared motile cilia (Figure 2.17d). This affinity for two tissues was further confirmed in later developmental stages where miR-22 was expressed in musculature and ciliated cells of the trunk and the head (Figure 2.17f-g).

miR-281, another ancient bilaterian microRNA which was lost in the vertebrate lineage and for which no expression data were available, showed a very restricted ring-shaped expression in the mouth region and the nuchal organ (Figure 2.18a). Expression of the microRNA was only detectable after 72hpf in conjunction with a similar observation in *Drosophila* (Xiong et al., 2009). Detailed characterisation of these uknown to date

circularly arranged miR-281+ cells (Figure 2.18b) was necessary in order to classify this microRNA as a muscle marker. The position of this ring was right below the mouth opening and above the foregut, setting a clear border between the two organs. Axons stemming from the mouth opening innervated the miR-281+ cells (Figure 2.18c). Marker gene analysis (done by A. Fischer) identified a number of different muscle markers, such as beta-3-tubulin, expressed in the same cells as miR-281 among other differentiated muscle cells (Figure 2.18d-e). Their identity was therefore resolved and by re-examining the expression patterns of all above muscle specific microRNAs at 5dpf, I could identify these cells within their general muscle demarcation. The mouth muscle ring was in fact identifiable even under white light microscopy (arrow in Figure 2.18g) but never observed before so it was thanks to the miR-281 expression pattern that a subset of musculature with a possible role in feeding was discovered in *Platynereis*.

Gut Finally, miR-12, -216 and -283 showed identical expression in the differentiating midgut at 5dpf (Figure 2.12n-p). miR-12 clusters with miR-216 in *Platynereis* (Table *2.3*) and in *Lottia* (Prochnik et al., 2007) and with miR-283 in *Drosophila* (Aboobaker et al., 2005) (Table 2.3) indicating that these three microRNAs evolved from the same precursor gene. In the vertebrates, expression of miR-216 is characteristic of pancreatic tissue (Szafranska et al., 2007) and targets of miR-216 are expressed at lower levels in pancreatic than in other tissue (Sood et al., 2006), indicating that the ancient site of activity of the miR-12/-216/- 283 precursor had indeed been the gut.

2.2.2 Comparative analysis of microRNA expression between annelid and vertebrates

Integrating all information available in the literature on expression of microRNAs in fish and mouse was of special value for the comparative approach I followed. In many cases, expression comparison revealed homologous tissues/cell types conserved between annelid and vertebrates (highlighted in bold for fish in Table 2.4 on page 89a and colour coded with dots for mouse in pie chart 2.4 on page 89b and Table 2.4 on page 89a).

Of particular interest were the microRNAs for which no expression data were available or had exhibited a ubiquitous/broad expression pattern in vertebrates, in contrast to a highly restricted tissue specific pattern documented in *Platynereis* (ie. let-7, miR-100, -125, -10, -31, -29, -190, -22, -210). Earlier studies of the same microRNAs in other invertebrates had yielded very limited information about the localisation of the mature microRNA form and the *Platynereis* dataset came to fill in the gaps and reveal their tissue affinity.

Lastly, the microRNA expression analysis I performed in *Platynereis* generated a wealth of information for a number of ancient bilaterian microRNAs which have been lost in the vertebrate lineage and have been poorly or not at all studied in other invertebrates (ie. miR-278, -71, -252, -2001, -315, -281).

2.3 The localisation of microRNAs specific to the protostome lineage

2.3.1 Conserved protostome specific microRNAs

A whole mount in situ hybridization screen, using LNAs, was carried out for the most abundant protostome specific microRNAs according to Table 2.1. The resulting dataset, comprised of specific patterns, could be classified into groups of microRNAs sharing affinity for certain tissues/cell types just like what was done for the ancient bilaterian microRNAs.

2.3.1.1 Protostome specific microRNAs expressed in locomotor ciliated cells

miR-277 and -317, two protostome-specific microRNAs (Wheeler et al., 2009) that cluster with miR-34 in *Anopheles* and in *Drosophila* and thus likely to emerge from a polycistronic transcript (Table 2.3; Winter et al., 2007) were likewise detected in the ciliary bands of *Platynereis* swimming larva (Figure 2.19, Figure 2.5). miR-317 showed an identical expression pattern to miR-34 as it stained all locomotor cilia of the head and the trunk, plus the nuchal organ at 48hpf up until 5dpf (Figure 2.19a-c, h, white arrows point at nuchal organ). miR-277 showed higher specificity to a subset of locomotor ciliated cells of the trunk, as it was not detectable in the prototroch at neither 48hpf nor 72hpf (Figure 2.19d-g). At 72hpf the trunk expression of miR-277 was exclusive to the metatroch (arrow in Figure 2.19g) and at 5dpf the microRNA was no longer detectable concominant with a drop in solexa reads in young worm libraries (Table 2.1 6th column).

2.3.1.2 Protostome specific microRNAs expressed in the CNS

Three protostome specific microRNAs were found localised in the CNS of the developing *Platynereis* worm; miR-2, miR-87 and bantam.

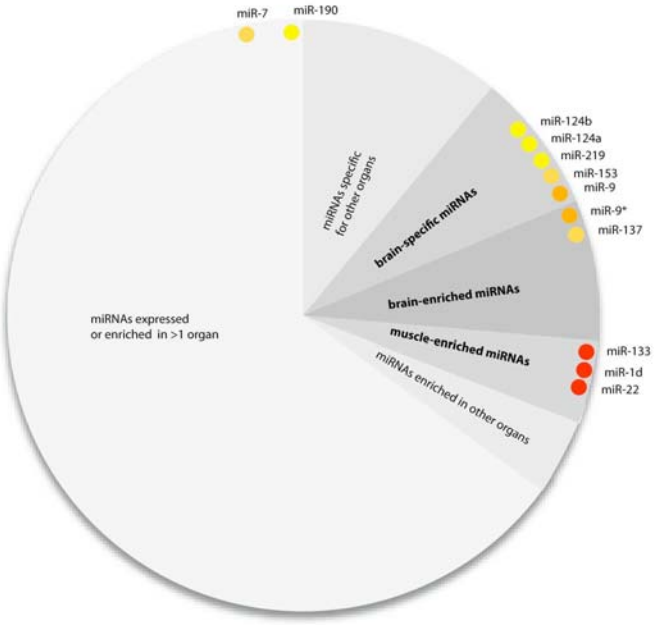
miR-2, the expression of which in Planaria and *Caenorhabditis* was localised in head and body neurons (Gonzalez-Estevez et al., 2009; Martinez et al., 2008), was likewise expressed in the ventral nerve cord, the foregut and the brain of *Platynereis* throughout

miRNA	Zebrafish expression patterns as described by (Wienholds 2005 and Kapsimali 2007)	miRNA	*Platynereis* expression patterns
miR-100	CNS/ brain (hindbrain, diencephalon); spinal cord	miR-100	secretory foregut neurons
let-7a	brain, spinal cord	let-7	secretory foregut neurons and gut
miR-125b	CNS/ brain; spinal cord; cranial ganglia	miR-125	secretory foregut neurons, brain, gut
miR-375	**pituitary, pancreatic islet, few hypothalamic cells**	miR-375	secretory foregut neurons, gut
miR-10	posterior trunk; spinal cord	miR-10	secretory neurons along mouth (gland tissue)
miR-278	lost miRNA in vertebrates	miR-278	secretory neurons along mouth (gland tissue)
miR-31	ubiquitous (early), no expression (late)	miR-31	differentiated neurons in foregut
miR-29	no expression pattern	miR-29	ciliated cells, foregut
miR-92b	**periventricular and adjacent cells** of the ventral and dorsal subpallium in brain	miR-92	ciliated cells, foregut
miR-34a	nucleus of the **paraventricular organ**, cells in forebrain, midbrain, hindbrain, spinal cord	miR-34	ciliated cells
miR-7	pancreatic islet, Brain (neurons in forebrain; diencephalon/**hypothalamus**)	miR-7	neurosecretory brain tissue
miR-137	caudal telencephalon, diencephalon, dorsal midbrain and **hypothalamus**, cranial nerves/ganglia	miR-137	neurosecretory brain tissue, serotonergic neurons of ventral nerve cord
miR-153	brain (fore- mid- and hindbrain, diencephalon/**hypothalamus**)	miR-153	neurosecretory brain tissue
miR-9	periventricular/proliferating/differentiating cells of the **telencephalon**, diencephalon, optic tectum, spinal cord	miR-9	sensory associative brain tissue
miR-9*	proliferating cells of brain, spinal cord and eyes | in mouse miR-9 expressed throughout **telencephalon**	miR-9*	sensory associative brain tissue
miR-71	lost miRNA in vertebrates	miR-71	differentiated neurons (all)
miR-124	**differentiated neurons** in the brain, eyes, spinal cord and cranial ganglia	miR-124	differentiated neurons in eyes, brain and ventral nerve cord
miR-219	thalamus, periventricular pretectum, hindbrain, spinal cord	miR-219	4 FVRI amidergic cells in brain and ventral nerve cord
miR-184	lens, epidermis	miR-184	CNS subset (glutamatergic neurons)
miR-190	no expression pattern	miR-190	CNS subset (glutamatergic neurons)
miR-200b	**sensory cells** (olfactory epithelium, taste buds, ear, neuromasts)	miR-8	sensory organs
miR-182	**sensory cells** (photoreceptors, neuromasts, olfactory epithelium and ear)	miR-263	sensory organs
miR-183	**sensory cells** (photoreceptors, neuromasts, olfactory epithelium and ear) cranial ganglia	miR-183	sensory organs
miR-252	lost miRNA in vertebrates	miR-252	sensory organs
miR-2001	lost miRNA in vertebrates	miR-2001	sensory organs
miR-22	ubiquitous	miR-22	general musculature
miR-1	body, head and fin **muscles**	miR-1	general musculature
miR-133a	body, head and fin **muscles**	miR-133	general musculature
miR-216	brain, spinal cord, proliferative cells of eyes, **pancreas**, body muscles	miR-216	gut
miR-315	lost miRNA in vertebrates	miR-315	midline and head epidermis
miR-281	lost miRNA in vertebrates	miR-281	subset of muscle (muscle ring of mouth opening)
miR-210	ubiquitous	miR-210	CNS and gut
miR-33	no expression pattern	miR-33	no expression pattern

miRNas are grouped according to their tissue/organ affinities observed in *Platynereis*. In bold, we highlight tissue/cell type homology which becomes apparent after comparing miRNAs expression sites. Possible conservation of ancient expression sites in zebrafish is highlighted in bold.

MiRNAs are grouped according to their tissue/organ affinities in *Platynereis*. In bold, we highlight homologous tissue/cell types in fish hat express the same microRNAs. Closed circles demarcate microRNAs that are brain and muscle-specific in *Platynereis* and have been included in the comparative expression study depicted in (b). Open circles mark miRNA not inlcuded in this study

(a) **Comparison of microRNA expression between zebrafish and *Platynereis***



The pie chart summarizes the tissue-specificity of 91 miRNAs investigated in mouse and human. Note that the population of brain-specific and -enriched miRNAs covers all except two of the miRNAs that are brain-specific in *Platynereis*.(These two miRNAs still cluster with the brain-specific miRNAs in human and mouse.)

(b) **Comparison of microRNA expression between mouse and *Platynereis* (based on Sempere et al 2004)**

Table 2.4: **Comparison of microRNA expression in *Platynereis*, zebrafish and mouse.**

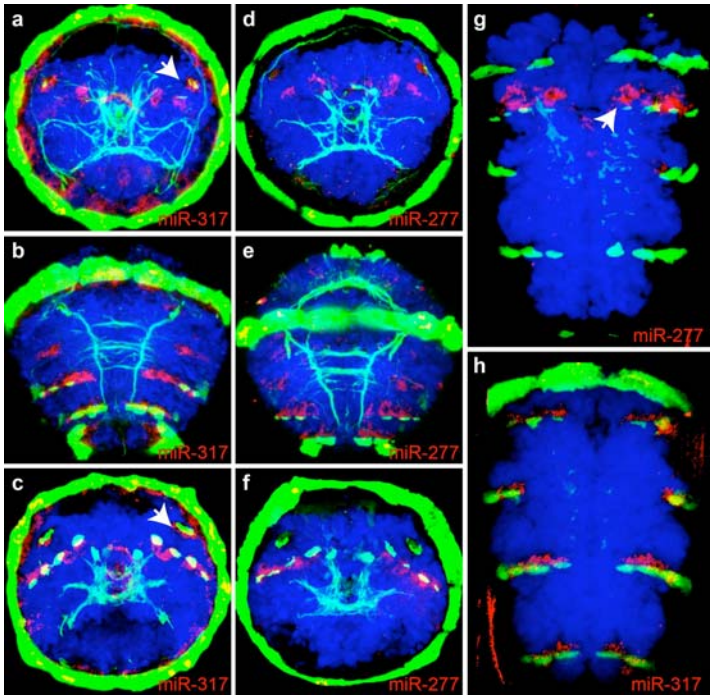


Figure 2.19: **Protostome specific microRNAs expressed in locomotor ciliated cells. a,** apical view of 48hpf *Platynereis* larva with miR-317 expression in ciliated cells of the head, the nuchal organ (white arrow) and prototroch (blue: DAPI; green: acTub; red: miR-317). **b,** ventral view of the 48hpf larva with miR-317 expression the prototroch and ciliated cells of the trunk. **c,** apical view of 72hpf larva with miR-317 expression in head ciliated patches, the nuchal organ (white arrow) and prototroch. **d,** apical view of 48hpf larva with miR-277 expression in ciliated cells of the head, and the nuchal organ. **e,** ventral view of 48hpf larva with miR-277 expression in ciliated cells of the trunk. **f,** apical view of 72hpf larva with miR-277 expression in head ciliated patches. **g,** ventral view of 72hpf larva expressing miR-277 exclusively in metatroch (white arrow). **h,** ventral view of 72hpf embryo expressing miR-317 in prototroch, metatroch, paratroch and telotroch.

development (Figure 2.20a). Its paneural pattern was very similar to that of miR-71 and *synaptotagmin* (Figure 2.12b-c). In closely related annelid *Capitella*, miR-2 was expressed identically to *Platynereis*; in the ventral nerve cord, the stomodaeum and the brain of the swimming larva (Figure 2.20b) confirming the paneural affinity the microRNA exhibited in *Platynereis*. Another protostome specific microRNA with paneural expression was miR-87 (Figure 2.20c), as it was localised in differentiating neurons of the trunk and the brain. Affinity for neural tissue has already been reported for miR-87 in Planaria (Gonzalez-Estevez et al., 2009).

bantam, which in *Drosophila* stimulates cell proliferation during wing-disc and eye development by inhibiting the translation of pro-apoptotic gene hid (Brennecke et al., 2003) and plays a role in circadian rythhmicity (Kadener et al., 2009a), has been detected by in situ hybridisation in brain neurons of both Planarian and *Caenorhabditis* (Gonzalez-Estevez et al., 2009; Martinez et al., 2008). In *Platynereis*, bantam was localised in a subset of the CNS giving a very unique pattern compared to any other neural microRNA. Throughout development, the trunk expression of bantam was restricted to a handful of glutamatergic neurons (Figure 2.20d-e compare to 2.14a) of the ventral nerve cord. At 72hpf brain bantam expression was restricted to few glutamatergic neurons and the adult eyes (Figure 2.20f, white arrows pointing at adult eyes). Later in development, brain expression included the antennae, palpae and cells of the mouth opening (white arrow in Figure 2.20g) with glutamatergic neurons of the midbrain exhibiting a notably stronger expression.

2.3.1.3 miR-279 demarcates the epidermis and PNS

The affinity of miR-279 for sensory organs has already been reported for both early developing as well as metamorphosing stages of *Drosophila* (Stark et al., 2005; Cayirlioglu et al., 2008). In *Platynereis*, miR-279 covered the entire epidermis of both head and trunk (Figure 2.21a) from 48hpf onwards. Later in development, aside from epidermis, the microRNA demarcated differentiating sensory organs of the PNS such as the antennae (Figure 2.21b), palpae (Figure 2.21c white arrows), mouth opening (white arrow in Figure 2.21d), pygidial lobe and sensory cells of the parapodia.

2.3.2 Lophotrochozoan specific microRNAs

Despite the early identification of certain lophotrochozoan specific microRNAs (Palakodeti, 2006) which was recently enriched by (Wheeler et al., 2009), no expression information was generated until recently in adult Planaria (Gonzalez-Estevez et al., 2009). Still, the

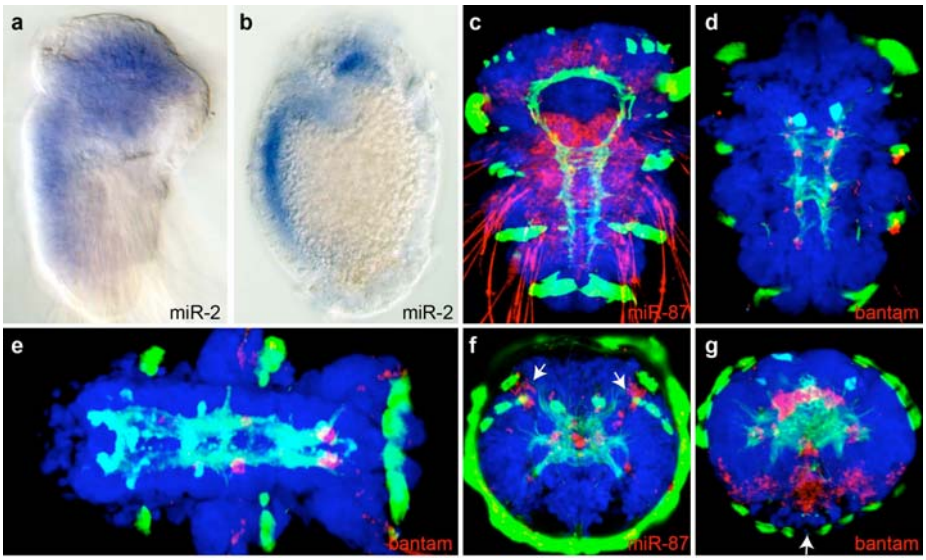


Figure 2.20: **Protostome specific microRNAs of the CNS. a,** lateral view of 72hpf *Platynereis* larva with miR-2 expression in the brain, stomodaeum and ventral nerve cord. **b,** lateral view of stage 6 *Capitella* swimming larva with miR-2 expression in the brain, stomodaeum and ventral nerve cord. **c,** ventral view of 72hpf *Platynereis* with miR-87 expression in the brain and ventral nerve cord (blue: DAPI; green: acTub; red: miR-87). **d,** 5micron thick ventral view of 72hpf *Platynereis* larva with bantam expression in glutamatergic neurons of the ventral nerve cord. **e,** 10micron thick ventral view of 5dpf *Platynereis* worm with bantam expression in glutamatergic neurons of the ventral nerve cord. **f,** apical view of 72hpf brain with bantam expression in adult eyes (white arrows) and glutamatergic neurons of the midbrain. **g,** apical view of 5dpf brain with bantam expression in glutamatergic neurons of the midbrain, in palpae and mouth opening (white arrow).

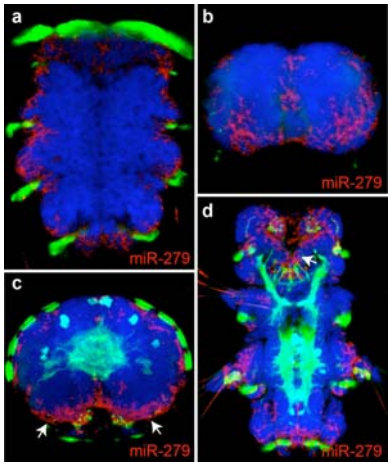


Figure 2.21: **Protostome specific miR-279 expression in PNS and epidermis. a,** 2micron thick section of 72hpf *Platynereis* ventral view with miR-279 expression in trunk epidermis (blue: DAPI; green: acTub; red: miR-279). **b,** 5micron thick slice of 5dpf brain apical view with miR-279 expression in antennae and head epidermis. **c,** apical view of 5dpf deeper slice of the brain with peripheral miR-279 expression in head epidermis and demarcation of palpae (white arrows) and mouth opening. **d,** 13micron thick slice of 5dpf worm ventral view with miR-279 expression in palpae, mouth opening, sensory cells of the parapodia and the pygidial lobe.

site of action and inferred role for lineage specific microRNAs during lophotrochozoan development remains uknown. The *in situ* hybridization screen I performed for a small set of abundantly expressed lophotrochozoan specific microRNAs in *Platynereis* (Table 2.1), revealed their sites of action to be in locomotor cilia, the CNS, PNS and gut.

2.3.2.1 Lophotrochozoan microRNAs expressed in locomotor ciliated cells

miR-1998 showed strong expression in locomotor ciliated cells from as early as 24hpf. Expression in the 48hpf swimming larva was restricted to the developing nuchal organ, locomotor ciliated cells of the head, the prototroch ring and ciliated cells of the trunk (Figure 2.22a). The locomotor ciliated cell-specific pattern persisted throughout development with head ciliary patches, nuchal organ, prototroch, metatroch, paratroch and telotroch expressing miR-1998 (Figure 2.22b-c).

Another microRNA, cloned from *Platynereis* small RNA libraries but without any orthologous sequence identified so far and therefore a potential *Platynereis* specific microRNA candidate, was also expressed in locomotor ciliated cells. miR-K showed an equally strong expression pattern as miR-1998 and its expression persisted in ciliated cells in all stages investigated (Figure 2.22d-f).

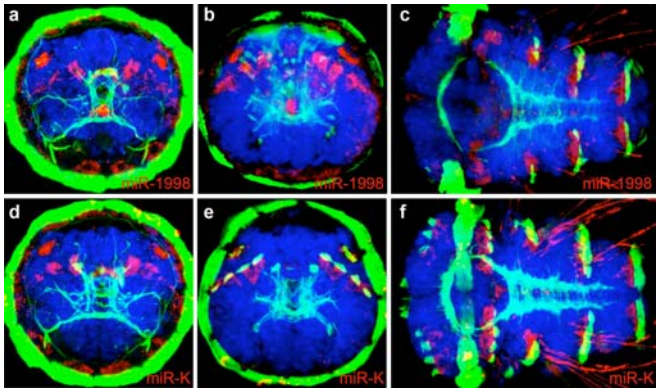


Figure 2.22: **Lophotrochozoan specific microRNAs expressed in motile cilia. a-b,** apical view of 48hpf and 72hpf *Platynereis* brain respectively with miR-1998 expression in the prototroch and motile ciliated cells of the head (blue: DAPI; green: acTub; red: miR-1998). **c,** ventral view of 72hpf larva with miR-1998 expression in ciliary bands of the trunk. **d,** apical view of 48hpf brain with *Platynereis* specific microRNA candidate miR-K expressed in the prototroch and motile ciliated cells of the head. **e,** apical view of 72hpf brain with miR-K expression in ciliary patches of the head and the nuchal organ. **f,** ventral view of 72hpf larva with miR-K expression in ciliary bands of the trunk and ciliated patches of the head.

2.3.2.2 Lophotrochozoan microRNAs expressed in the CNS

Three of the tested lophotrochozoan microRNAs were localised in the CNS; miR-36, -1996 and -1997. In all cases, expression was paneural, demarcating differentiating neurons of the trunk, foregut and of the head. For miR-36, the *Platynereis* paneural expression (Figure 2.23a) was confirmed in annelid *Capitella* (Figure 2.23b). miR-1996 and -1997 were likewise localised in neural tissue in a similar pattern as that of *synaptotagmin* (Figure 2.23c-d).

2.3.2.3 Lophotrochozoan microRNAs expressed in sensory organs

miR-1992, gave a highly restricted pattern in the 48h and 72h brain where it specifically demarcated all photoreceptor cells (Figure 2.24a). Previous expression studies of photoreceptor cells in *Platynereis* have identified two distinct photoreceptor cell types; a) rhabdomeric photoreceptors, which express differentiation marker rhabdomeric opsin (r-opsin) and can be found in the developing larval and adult eyes of the swimming larva (Arendt et al., 2002) and b) ciliary photoreceptors, which express ciliary opsin (c-opsin) and are positioned in the medial brain from where they project axonal connections into the neurosecretory plexus of the larval mid-brain (Arendt et al., 2004; Tessmar-Raible et al., 2007). Both rhabdomeric (adult and larval eyes pointed by white arrows in Figure

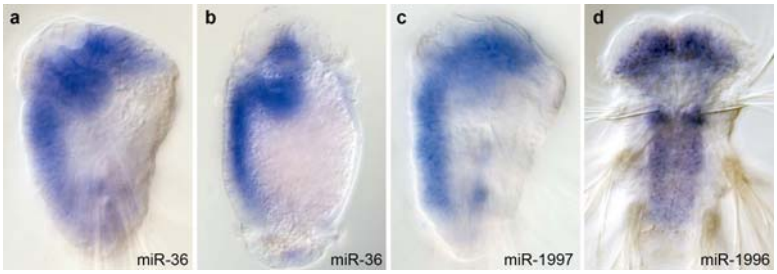


Figure 2.23: **Lophotrochozoan microRNAs expressed in the CNS. a-b,** lateral view of 72hpf *Platynereis* and stage 6 *Capitella* swimming larva respectively with miR-36 expression in brain, stomodaeum and the ventral nerve cord. **c,** lateral view of 72hpf *Platynereis* with miR-1997 expression in the brain, stomodaeum and ventral nerve cord. **d,** ventral view of 72hpd *Platynereis* with miR-1996 expression in the brain and ventral nerve cord.

2.24a, c) and ciliary photoreceptors (yellow arrow in Figure 2.24a and close up for ciliary photoreceptors in Fig 2.24b) strongly expressed miR-1992 throughout larval brain development. Before the discovery of miR-1992, no other marker gene able to highlight the molecular link between two distinct photoreceptor cell types of *Platynereis* was available. The high specificity of miR-1992 made it an excellent marker of photoreceptor cell studies.

At 5dpf, head expression of miR-1992 persisted in photoreceptor cells (adult eyes pointed by white arrows and ciliary photoreceptors by yellow arrow in Figure 2.24d-e) but was additionally detected in the palpae (Figure 2.24d, f). Trunk expression for miR-1992 was only detectable after 72hpf in few cells of the ventral plate (data not shown). However, at 5dpf miR-1992 expanded its trunk expression and demarcated sensory organs such as the cirri, sensory cells of the parapodia and the pygidial lobe (Figure 2.24f).

The second tested *Platynereis* specific microRNA candidate (with no orthologous sequence yet identified in other lophotrochozoan small RNA libraries), named miR-C, was localised in very few superficially positioned cells of the larval brain from 48hpf onwards (Figure 2.24g-j). The identity of these cells could not be deciphered by other gene marker analysis since no marker was ever exclusively detected in a similar pattern as miR-C during early development. At 5dpf, miR-C exhibited a very superficial/epidermal expression in the brain (Figure 2.24k-l) while the trunk expression was identical to any of the previously presented sensory organ specific microRNAs with miR-C+ cells found in palpae, cirri, sensory cells of the trunk and parapodia and in the pygidial lobe (Figure 2.24m). The microRNA therefore exhibited an epidermal/sensory organ affinity throughout development.

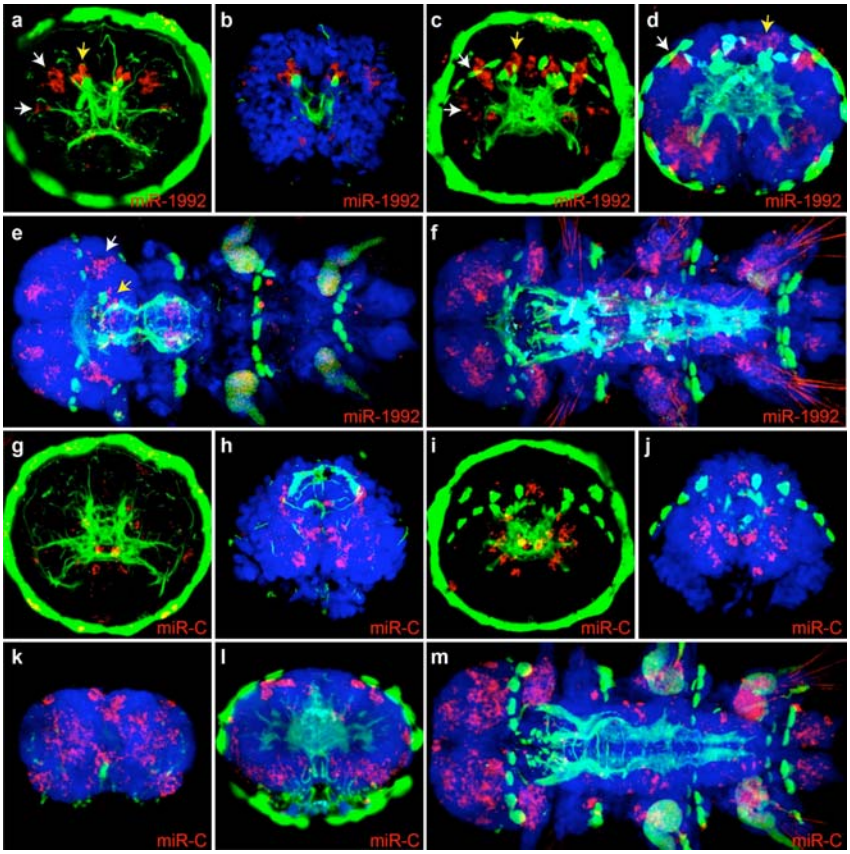


Figure 2.24: **Lophotrochozoan microRNAs expressed in sensory organs.** **a,** apical view of 48hpf *Platynereis* brain with miR-1992 expression in rhabdomeric (white arrows pointing at larval and adult eyes) and ciliary (yellow arrow) photoreceptor cells (green: acTub; red: miR-1992). **b,** 10micron slice of 48hpf brain apical view with miR-1992 expression in ciliary photoreceptors. **c,** apical view of 72hpf brain with miR-1992 expression in rhabdomeric (white arrows) and ciliary (yellow arrow) photoreceptor cells. **d,** apical view of 5dpf brain with miR-1992 expression in adult eyes (white arrow), ciliary photoreceptors (yellow arrow) of the dorsal brain and palpae. **e,** dorsal view of 5dpf *Platynereis* worm depicting the position of adult eyes (white arrow) and ciliary photoreceptors which are positive for miR-1992 along cells of the apical brain and neurons of the foregut. **f,** ventral view the 5dpf worm with miR-1992 expression in palpae, cirri, sensory cells of the parapodia, few neurons of the ventral nerve cord and the pygidial lobe. **g-h,** z-projection and 15micron slice of the 48hpf brain apical view respectively with *Platynereis* specific microRNA candidate miR-C expression in superficially positioned cells of the brain. **i-j,** z-projection and 20micron slice of the 72hpf brain apical view respectively depicting the superficial expression of miR-C. **k-l,** 20micron slice of the 5dpf apical brain and 30micron slice of the deeper brain respectively to highlight the epidermal and peripheral staining of miR-C additionally to expression in antennae and palpae. **m,** ventral view of 5dpf worm with miR-1992 expression in antennae, palpae, cirri, sensory cells of the trunk and parapodia and the pygidial lobe.

2.3.2.4 Lophotrochozoan microRNAs expressed in the gut

Four lophotrochozoan specific microRNAs showed expression restricted to the 5dpf gut; miR-750, -1175, -1175* and -1989.

miR-750 expression was detectable in cells of the differentiating midgut from 72hpf (Figure 2.25a). At 5dpf, expression of miR-750 had spread to demarcate the entire midgut (Figure 2.25b). miR-1175 showed a very unique expression pattern at 72hpf where it labelled few cells of the developing gut (Figure 2.25c), as well as single cells in the periphery of the stomodaeum (arrows in Figure 2.25c). Later in development, unlike other gut marker microRNAs, miR-1175 was not only expressed in the differentiating mid-gut but also in cells lining the exterior side of the pharyngeal muscle sheath (arrows in Figure 2.25d). This additional expression domain was encountered for the first time to point out a an interesting set of cells with uknown identity. Its reverse complement, miR-1175* was predominantly expressed in the stomodaeum at 72hpf with faint expression in the developing mid-gut region (Figure 2.25e). Later in development miR-1175* was localised in the 5dpf differentiating mid-gut (Figure 2.25f). Lastly, miR-1989 showed a very intriguing transition in its expression pattern from differentiating musculature between 48h-72hpf (Figure 2.25g) to restricted expression in the 5dpf mid-gut (Figure 2.25h).

2.4 Post- transcriptional regulation of *Platynereis* primary microRNA transcripts

Differences in expression levels (Thomson et al., 2006) and localisation (Obernosterer, 2006) between primary transcripts (pri-miRNA) of microRNA genes and their confined microRNAs which need to be processed in order to reach maturity/active form, have been reported for a number of microRNAs in mammals (Thomson et al., 2006). The underlying post- transcriptional regulation in biogenesis steps has recently been uncovered for specific microRNAs (Winter et al., 2009). In mammalian cell lines, let-7 biogenesis inhibition by target Lin28 has been proposed to happen at both important processing steps which lead to microRNA maturation; the Drosha (Newman et al., 2008; Viswanathan et al., 2008) and Dicer (Rybak et al., 2008; Heo et al., 2008) step. However, more recent results from an in vivo study carried out in *C. elegans* favor biogenesis regulation to be exerted in the cytoplasm, at the Dicer step (Lehrbach et al., 2009).

In *Platynereis*, sequence information about pri-miRs was originally not available. Without a sequenced genome, it was impossible to map the generated solexa reads to reference

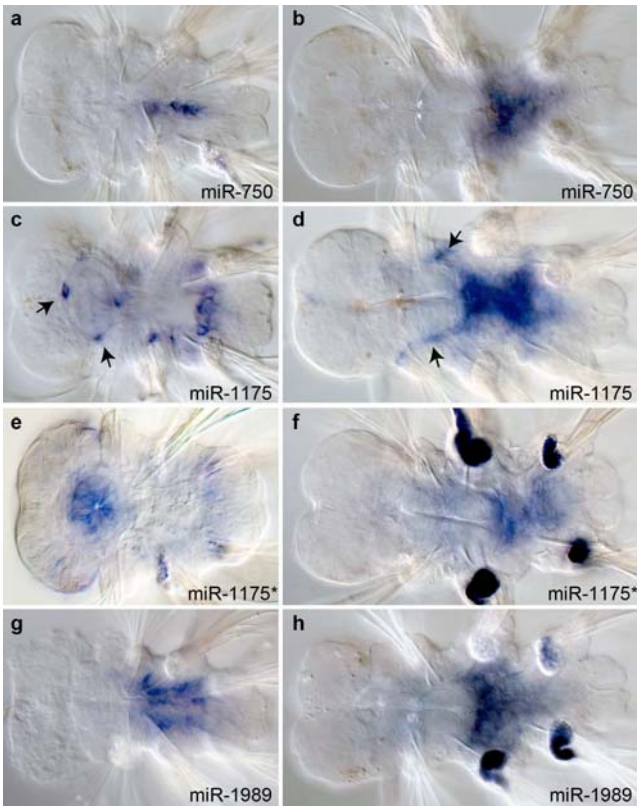


Figure 2.25: **Lophotrochozoan microRNAs expressed in gut.** **a-b,** dorsal view of 72hpf and 5dpf *Platynereis* worm respectively with miR-750 expression in the differentiating gut. **c,** dorsal view of 72hpf larva with miR-1175 expression in single cells of the stomodeal periphery (black arrows) and in the developing gut. **d,** dorsal view of 5dpf worm with miR-1175 expression in the gut. **e,** dorsal view of 72hpf larva with stomodeal expression of miR-1175*. **f,** dorsal view of 5dpf worm with miR-1175* expression in gut. **g,** dorsal view of 72hpf larva with miR-1989 expression restricted to a subset of trunk musculature. **h,** dorsal view of 5dpf with miR-1989 expression in the gut.

sequences and extract information about every cloned mature microRNA's genomic locus and its corresponding pri-miRNA sequence. However, after deducing that some microRNA clusters show conserved synteny across bilateria (Table 2.3), I generated cluster enriched cDNA libraries using the 3' most microRNA of each conserved cluster as a gene specific primer. In this manner, I could successfully clone through RT-PCR three distinct clusters; pri-miR-100-let-7, pri-miR-183-263 and pri-miR-12-216 and make riboprobes for each of them in order to visualise their expression patterns. M. Tosches independently cloned pri-miR-1992 for which she also generated a riboprobe and performed a WMISH. Out of the four tested pri-miRs, three exhibited a broader (and in one case developmentally earlier onset of) expression compared to that of their confined mature microRNAs.

Pri-miR-100-let-7, was detectable from 48hpf with broad expression in the lateral region of the larval brain (like miR-8 expression in Figure 2.28e and Appendix microRNA expression atlas), the stomodaeum and neurons of the ventral nerve cord. At 72hpf, lateral brain and stomodeal expression persisted and additional expression was observed in the developing gut (Figure 2.26a-b). Later in development, brain expression was restricted to the ventral side specifically labelling the antennae and palpae (Figure 2.26c). In the trunk, pri-miR-100-let-7, was additionally detected in cells of the foregut and midgut of the 5dpf worm (Appendix microRNA expression atlas). On the contrary, mature miR-100 was never detected in the brain (Figure 2.26d, f) and was exclusively expressed in few cells of the foregut (Figure 2.26e and appendix for 5dpf). The processing of miR-100 out of its primary transcript therefore seems to be regulated in *Platynereis* too, both temporally and spatially.

The expression of pri-miR-183-263 was likewise broader (Figure 2.27a) compared to the highly restricted localisation of either miR-183 or miR-263 in antennae, palpae, cirri and the pygidial lobe (Figure 2.27b-c). At 5dpf, pri-miR-183-263 showed additional strong expression in the parapodial epidermis and the trunk epidermis of the first and second segments (compare Figure 2.27a to b-c). This result revealed a second example of possible post- transcriptional regulation in yet another *Platynereis* pri-miRNA which exhibited differential localisation to its mature counterparts. A similar finding was reported for pri-miR-1992 by M. Tosches who documented it's broader expression compared to mature miR-1992 (data not shown and Figure 2.24a-d). Lastly, pri-miR-12-216 showed identical expression to its mature microRNAs (Appendix expression atlas and Figure 2.12m, o).

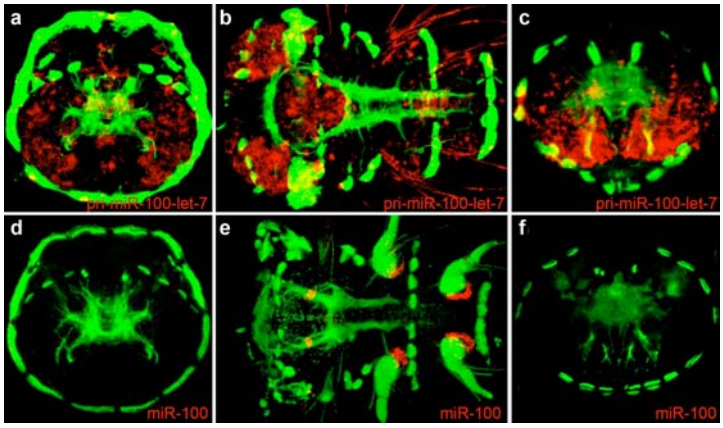


Figure 2.26: **Post- transcriptional regulation of pri-miR-100-let-7. a,** apical view of 72hpf brain with pri-miR-100-let-7 expression in lateral and dorsal brain (green: acTub; red: pri-miR-100-let-7). **b,** ventral view of 72hpf larva with pri-miR-100-let-7 expression in developing palpae (ventral brain), stomodaeum and developing gut. **c,** apical view of 5dpf brain with pri-miR-100-let-7 expression in the antennae and palpae. **d,** apical view of 72hpf brain showing no detectable miR-100 expression (green: acTub). **e,** dorsal view of 72hpf larva with miR-100 expression restricted to few cells of the stomodaeum (chaetal sacs staining is probe trapping). **f,** apical view of 5dpf brain with no expression of miR-100 detectable.

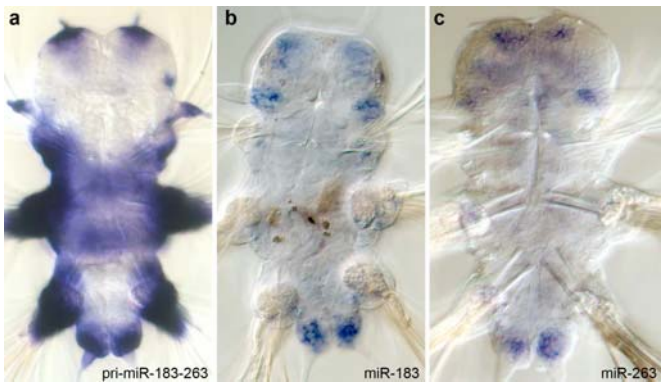


Figure 2.27: **Post- transcriptional regulation of pri-miR-183-263. a,** dorsal view of 5dpf *Platynereis* worm with pri-miR-183-263 expression in antennae, palpae, cirri, epidermis of the parapodia and two of the three trunk segments and pygidial lobe. **b-c,** ventral view of 5dpf Platynereis worms with miR-183 and miR-263 expression respectively in antennae, palpae, cirri and pygidial lobe.

2.5 miRNA localisation in the 48h larval brain

Apace with my microRNA in situ screen, R. Tomer in the laboratory developed a protocol for Whole Mount In Silico Expression Profiling (WMISEP). This protocol utilizes the highly stereotypical features of *Platynereis* larval development to generate a common reference average axonal scaffold image of the 48h brain (derived from 36 distinct larval brain images), upon which other 3D images can be aligned. In this manner, expression information documented for any gene of interest can be brought into the same coordinate system, thus allowing high- throughput gene co-expression comparisons in silico (R. Tomer PhD thesis 2008). To ensure reliability over the expression pattern of the genes brought into the system, R. Tomer computed an averaged expression pattern for every gene of interest by averaging 3-5 distinct confocal brain scans coming from different larvae.

A minimum of 3 brain scans per microRNA were used to produce the average microRNA expression patterns which were then aligned to the common reference brain. Only microRNAs with strong enough (and therefore detectable under the confocal microscope) expression pattern in the 48h brain were incorporated into the WMISEP database resulting in a total set of 19 aligned averaged expression patterns of mature microRNAs (Figure 2.28). Aligning all microRNA patterns to a common reference nervous system, made it especially easy to group those with related spatial localisation together. The four resulting groups demarcated distinct regions of the 48h developing brain such as; neurosecretory tissue (Figure 2.28a-d), differentiating sensory and epidermal tissue (Figure 2.28e-h), the cerebral ganglia (Figure 2.28i-l) and motile ciliated cells (Figure 2.28m-r). A molecular dissection of the 48h brain topology into medial and lateral (blue and pink in Figure 2.28t scheme) became possible after considering the expression pattern of paneural microRNAs (medially) versus that of sensory microRNAs (laterally) (see scheme in Figure 2.28t).

In-silico co-expression comparisons among the set of averaged microRNA expressions provided a unique visualisation of the very strict tissue demarcation microRNAs exert in the 48h brain. While microRNAs which belonged in the same groups showed extensive overlap (depicted in white Figure 2.29a-e), the opposite was observed from different classes of microRNAs whose expressions never "met" (Figure 2.29f-y).

Positive cells for any of the sensory or epidermal microRNAs never co-expressed a microRNA which could also be found in either neurosecretory tissue, in neurons of the cerebral ganglia or in motile ciliated cells (Figure 2.29f-o). Likewise, "ciliary" microRNA markers never "met" in expression with any other tissue's microRNA marker (Figure

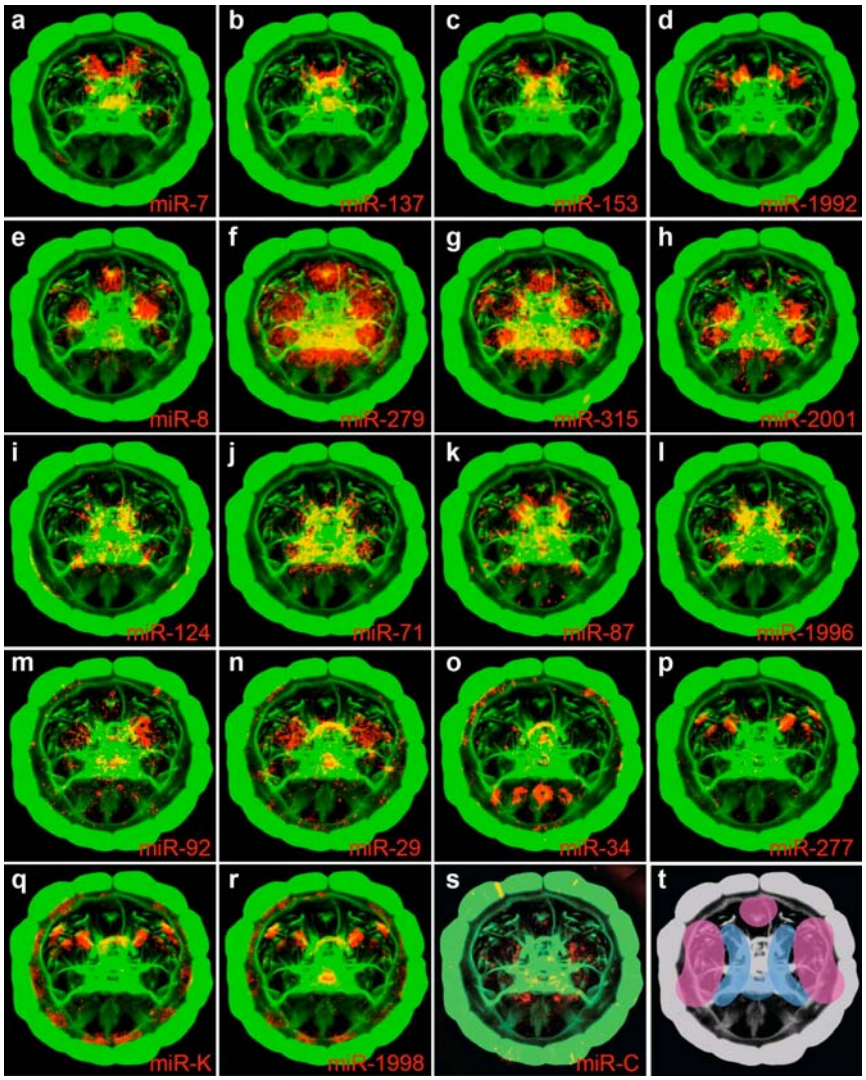


Figure 2.28: **Average expression patterns in the 48h larval brain for all microRNAs aligned using WMISEP.** 48hpf brain expression of microRNAs in: **a-d,** neurosecretory tissue. **e-h,** differentiating sensory and epidermal tissue. **i-l,** the cerebral ganglia. **m-r,** motile ciliated cells. **s,** uknown cells of the the apical brain and epidermis. **t,** scheme of the 48hpf brain depicting the molecular dissection of the 48h brain topology into medial/neural (light blue) and lateral/sensory (pink) domains.

2.29o-x). Finally, neurosecretory and paneural microRNAs did show some overlap in the medio-dorsal brain (depicted in white in Figure 2.29y).

Such mutual exclusion between microRNAs with different tissue affinity has already been reported for *Drosophila* during neuroectoderm development (Stark et al., 2005) and seems to have a role in brain development of *Platynereis* too.

2.6 microRNA target analysis

The strict tissue demarcation by different sets of microRNAs which I observed in *Platynereis* raised a number of questions concerning the biological significance of this phenomenon. Do the targets of all these microRNAs actively avoid them by being expressed in a complementary, non overlapping manner (adjacent tissues) as reported for *Drosophila* (Stark et al., 2005) or can they be found co-expressed in the same cells/tissues as shown in mammals by array experiments (Farh, 2005)?

To relate expression of the conserved bilaterian microRNAs to that of their predicted targets, I performed WMISH for 92 different genes and annotated the expression of their transcripts' differential expression (Figure 2.30, Figure 2.31 and Figure 2.5) with a total of 433 predicted target sites (predictions done by K. Trachana). The large set of riboprobes I used for the WMISH had already been generated by H. Snyman and R. Tomer.

Aside from experimental validation (which was impossible in the limited timed frame I had for this analysis), target prediction quality can be scored in a number of ways (Bartel, 2009). TargetScan algorithm can distinguish between evolutionarily conserved and non-conserved target sites (Lewis et al., 2005), however, this option could not be applied for *Platynereis*. Unfortunately, at the level of 3' UTRs sequence similarity of orthologs between lophotrochozoa or even between closely related annelids like *Capitella* was too small to allow their alignment (which is essential for TargetScan to identify conserved binding sites).

To overcome this problem O. Simakov in the laboratory followed an alternative approach, by taking advantage of *Platynereis* highly polymorphic transcriptome (which he assembled based on solexa reads) and investigating the rate of SNPs found in target sites versus those found in the remaining 3' UTR. Predicted microRNA binding sites exhibited reduced SNP frequencies (1/60 bp for predicted target sites versus 1/45 bp for general 3'UTRs) indicating an overall positive selection on these sites. Still, the overall SNP frequency was too low in order to distinguish between conserved and non-conserved microRNA binding sites for the entire set of predicted targets. Because of the above

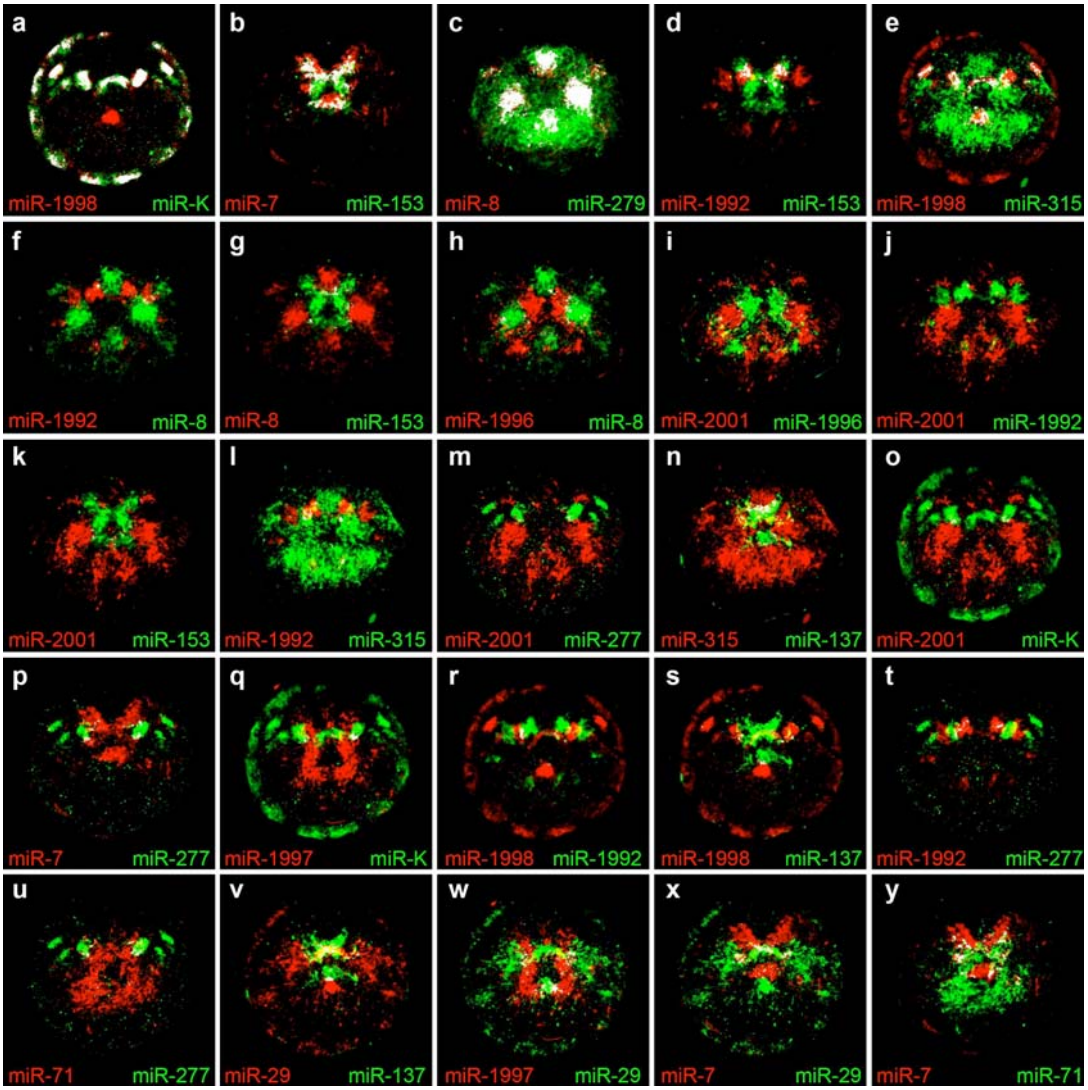


Figure 2.29: **In-silico co-expression comparisons among the set of averaged microRNAs.** White regions depict co-expression between microRNAs.

mentioned limitations, the only means of increasing prediction stringency was by using a second prediction algorithm (PITA) and only consider shared predictions between the two.

Statistical analysis done by O. Simakov who used the detailed annotation atlas I generated after documenting the expression patterns of 92 genes and 34 microRNAs at 5dpf (Table 2.5), revealed a subset of tissues in which microRNAs were less frequently co-expressed with their predicted targets than expected to occur by chance (Table 2.6; co-expression before or after 5dpf was not examined). This may, at least in part, be due to selective avoidance of target sites in co-expressed messengers (Shkumatava et al., 2009). Other tissues showed the opposite trend (Table 2.6), indicating different modes of involvement of the conserved bilaterian microRNAs in regulating the establishment and/or maintenance of tissue identity (Shkumatava et al., 2009). When we increased prediction stringency by combining TargetScan and PITA algorithms, this led to a general increase of bootstrap support values for co-expression of microRNAs and predicted targets (compare Table 2.6 to 2.7), indicating that the (unknown) set of 'real' targets may be further enriched for co-expression events. Statistically, we found that a similar trend is highly improbable for any random subset of similar size.

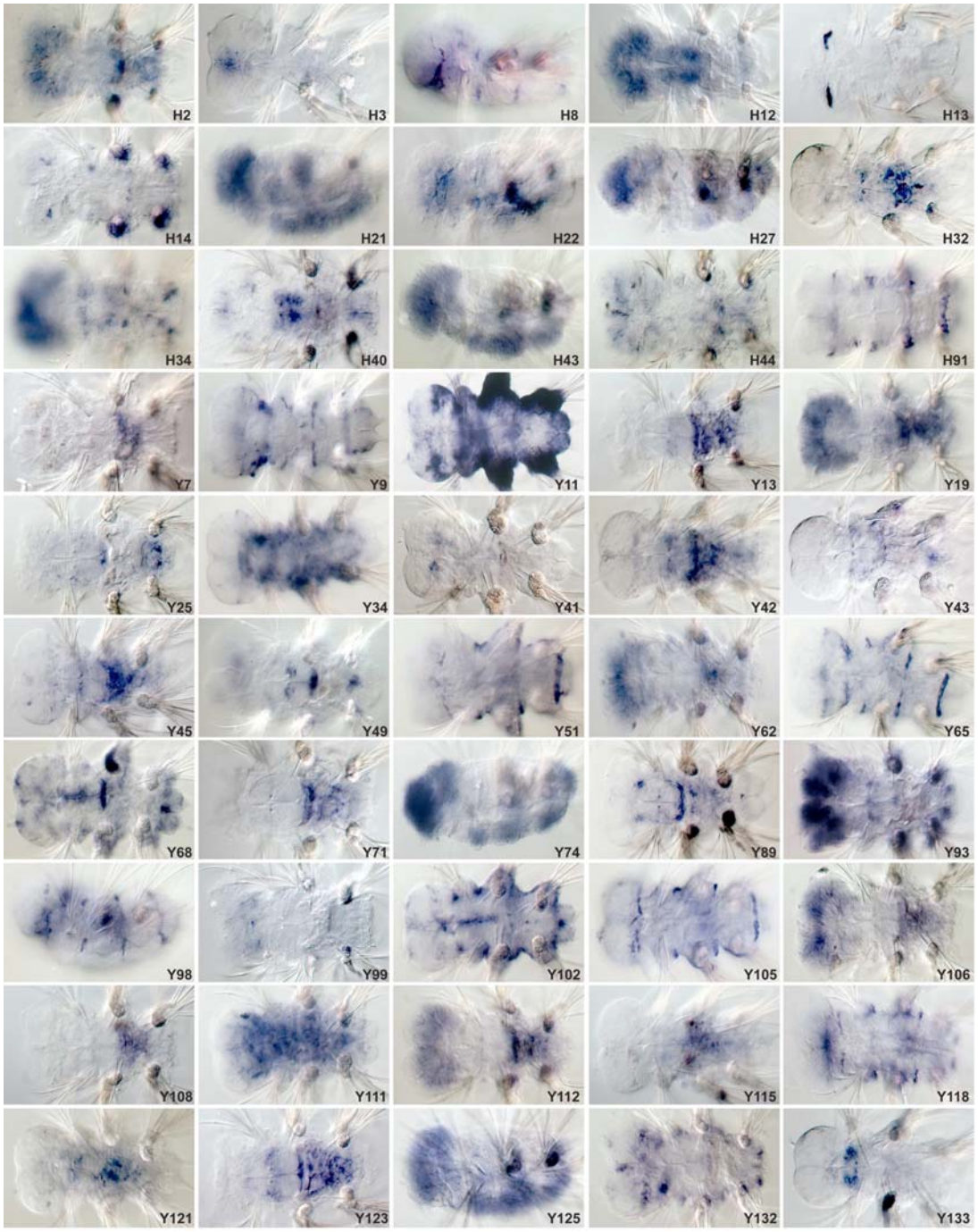


Figure 2.30: **Expression patterns of 50 target messengers in 5dpf worms.** CNS expression is documented from lateral view, peripheral sensory nervous system expression from ventral view, foregut, gut, musculature and brain expression from dorsal view and cilia from either lateral or ventral view. Detailed anatomical annotations for their expression patterns are provided in Table 2.5. Targets are named after the RNA probe IDs used for the in situ screen. Their genebank accession numbers are available in Appendix.

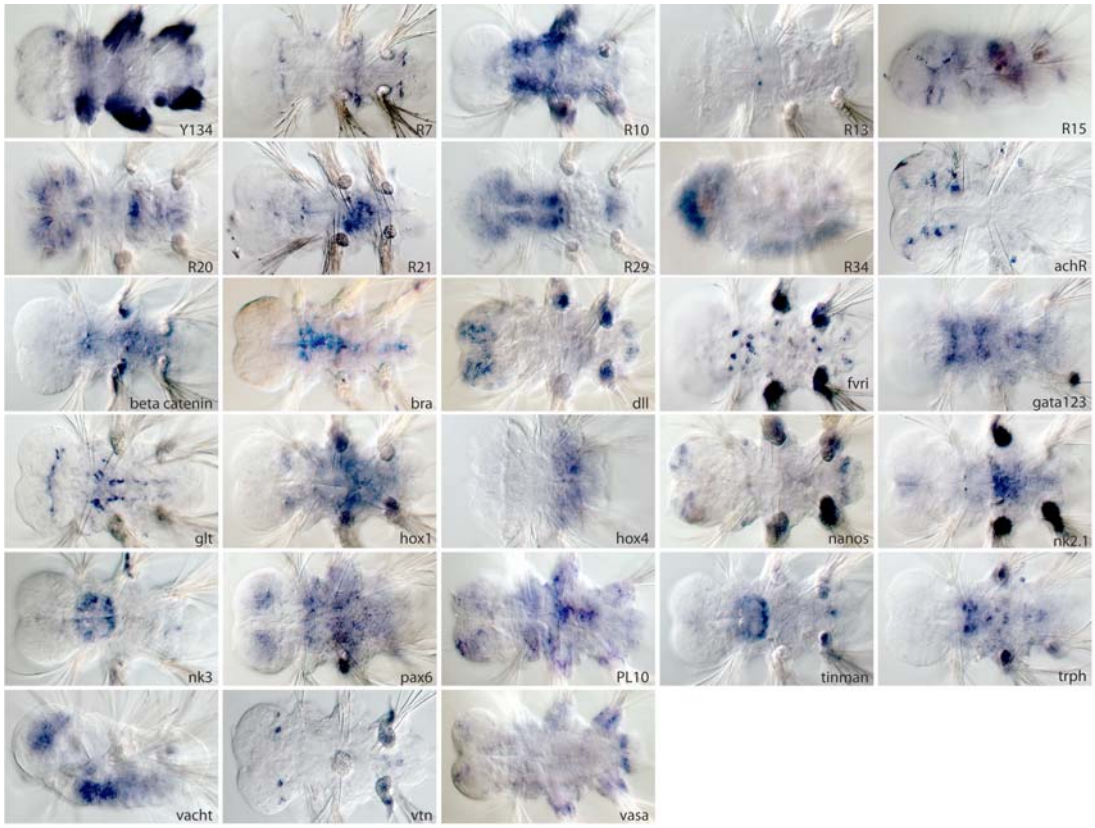


Figure 2.31: **Expression patterns of 28 target messengers in 5dpf worms.** CNS expression is documented from lateral view, peripheral sensory nervous system expression from ventral view, foregut, gut, musculature and brain expression from dorsal view and cilia from either lateral or ventral view. Detailed anatomical annotations for their expression patterns are provided in Table 2.5. Targets are named after the RNA probe IDs used for the in situ screen. Their genebank accession numbers are available in Appendix.

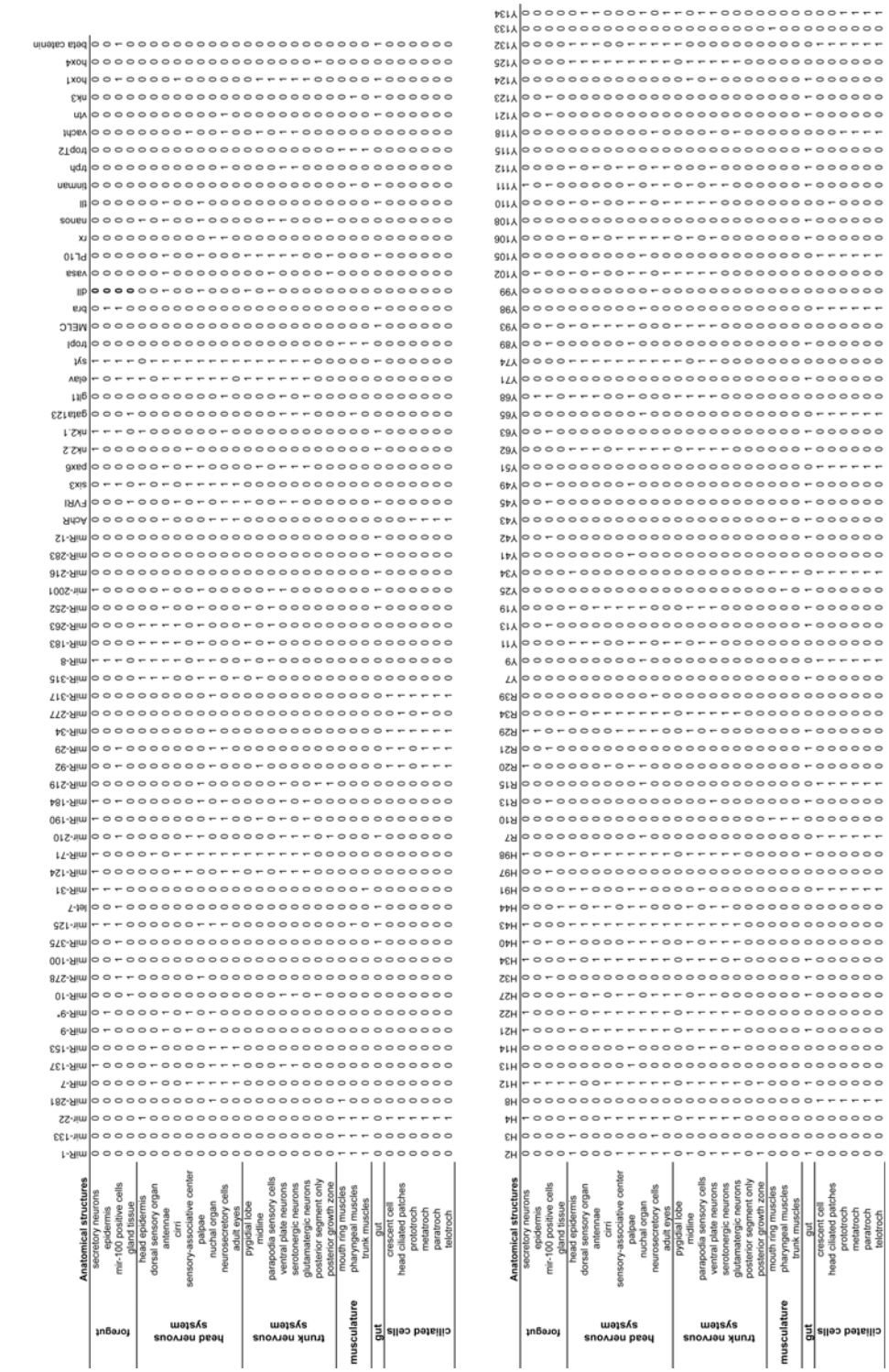


Table 2.5: **Detailed annotation of miRNAs and target transcripts expression patterns for five major body parts of the *Platynereis* 5 day worm.** Presence or absence of microRNA or target gene expression from a tissue/organ is scored with 1 or 0 respectively.

Tissue	miRNAs expressed in tissue: Names	miRNAs expressed in tissue: Counts of target-miRNA pairs: #coexpressed in the tissue \ total count	miRNAs expressed in tissue: Counts of target-miRNA pairs: Coexpression ratio	Randomization: Average coexpression ratio after randomizations	Randomization: Bootstrap percent support for: Fewer coexpression events	Randomization: Bootstrap percent support for: More coexpression events	Genes expressed in tissue: Names	Genes expressed in tissue: total #genes	fraction of all genes
secretory neurons	miR-31,miR-125,miR-R,miR-8,miR-184,miR-190,miR-71,miR-124,miR-137	4/32	0.05	0.16	99	1	H21,H34,R20,PDU-SYT,H43,Y111,H98,H12,PDU-NK2_1,R29,H4,H22,H40,PDU-ELAV	14	0.15
pharyngeal epidermis	miR-9*,miR-31,miR-125,miR-9,miR-8	3/47	0.06	0.09	67	15	PDU-BRA,Y102,PDU-SYT,PDU-SIX3,H12,PDU-NK2_1,R29,Y68	8	0.09
miR-100 positive cells	miR-31,miR-125,miR-278,miR-29,LET-7,miR-100,miR-375,miR-8,miR-184,miR-210,miR-92	22/66	0.26	0.25	33	54	PDU-BRA,R13,Y48,Y13,PDU-BETA_CATENIN,R21,H32,Y123,Y69,Y121,PDU-SYT,PDU-SIX3,Y111,Y45,PDU-HOX1,H12,Y63,PDU-NK2_1,Y93,H40,PDU-ELAV,Y42,Y68	23	0.25
gland tissue	miR-278,miR-10	0/7	0.00	0.06	39	0	H44,PDU-FVRI,PDU-GATA123,PDU-SYT,H12,R29,PDU-ELAV	7	0.08
head epidermis	miR-22,miR-315,miR-8	6/43	0.14	0.33	99	0	H44,H21,Y125,Y110,Y19,H91,H3,Y102,H34,H2,Y62,H27,PDU-NOS,Y112,Y74,Y132,H43,PDU-SIX3,R34,H98,H12,PDU-NK2_1,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68,Y11	30	0.33
antennae	miR-R,miR-252,miR-9*,miR-315,miR-9,miR-8,miR-183,miR-263	29/101	0.29	0.34	86	9	PDU-PL10A,H44,H21,Y125,Y110,Y19,Y134,Y102,PDU-TLX,PDU-ACETYLCHOLINE_RECEPTOR,H34,Y62,H27,PDU-VASA,PDU-NOS,Y112,Y74,Y132,PDU-PAX6,PDU-SYT-H43,PDU-SIX3,R34,H98,PDU-DLL,Y93,H22,PDU-ELAV,Y106,Y68,Y11	31	0.34
sensory-associative centre	miR-7,miR-9*,miR-8,miR-210,miR-71,miR-124	10/42	0.24	0.25	45	36	H21,Y125,Y19,PDU-VACHT,H34,H2,Y62,H27,Y112,Y74,PDU-PAX6,PDU-SYT,H43,PDU-SIX3,R34,H98,H12,R29,Y93,H4,H22,PDU-ELAV,Y106	23	0.25
palpae	miR-R,miR-125,miR-7,miR-252,miR-315,miR-8,miR-219,miR-183,miR-210,miR-263,miR-184,miR-71,miR-124	60/137	0.44	0.45	61	33	H14,Y49,PDU-PL10A,H44,H21,Y125,Y110,Y19,PDU-FVRI,H91,Y102,PDU-TLX,H34,H13,H2,Y62,H27,Y69,PDU-NOS,Y112,R20,Y74,Y132,PDU-PAX6,PDU-SYT,H43,PDU-SIX3,Y111,R34,H98,H12,PDU-DLL,Y41,R29,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68,Y11	42	0.46
cirri	miR-8,miR-183,miR-263,miR-71,miR-124	8/57	0.14	0.19	76	14	H44,H21,Y125,Y19,PDU-FVRI,H34,R20,Y74,PDU-SYT,H43,R34,PDU-HOX1,H98,H12,Y93,H4,H22,PDU-ELAV	18	0.20
nuchal organ	miR-281,miR-317,miR-125,miR-7,miR-29,miR-315,miR-153,miR-8,miR-190,miR-71,miR-34,miR-92,miR-137	42/147	0.29	0.37	99	1	R15,H44,H21,Y125,Y110,Y19,Y134,Y9,H91,Y105,Y102,PDU-ACETYLCHOLINE_RECEPTOR,R7,H34,H2,Y62,Y65,R20,Y74,Y132,PDU-SYT,H43,PDU-SIX3,R34,H98,H12,R29,H4,H22,H40,PDU-ELAV,Y98,Y106,Y11	34	0.37
neurosecretory cells	miR-125,miR-137,miR-7,miR-29,miR-153,miR-190,miR-184,miR-71,miR-124	23/71	0.32	0.37	75	15	PDU-GLT,H21,Y125,PDU-VACHT,PDU-FVRI,H3,PDU-VTN,PDU-ACETYLCHOLINE_RECEPTOR,H34,H2,Y62,H27,Y112,R39,Y98,Y118,Y74,Y132,PDU-SYT,H43,PDU-SIX3,Y111,R34,H98,H12,PDU-NK2_1,R29,H4,H22,H40,PDU-ELAV,Y106,PDU-TRPH	33	0.36
dorsal sensory organ	miR-153,miR-8,miR-71,miR-315,miR-7,miR-183,miR-263,miR-137	3/98	0.03	0.10	99	0	Y125,Y134,H91,Y62,Y74,Y132,PDU-SYT,H43,Y68,Y11	10	0.11
adult eyes	miR-315,miR-7,miR-153,miR-71,miR-124,miR-137	12/61	0.20	0.29	95	4	H21,Y125,Y110,PDU-FVRI,Y134,Y102,PDU-ACETYLCHOLINE_RECEPTOR,H34,H2,Y62,H27,Y112,H38,Y74,Y132,PDU-SYT,H43,PDU-SIX3,Y111,R34,H98,H12,H4,H22,Y106,Y11	26	0.28
ventral plate neurons	miR-R,miR-137,miR-10,miR-219,miR-190,miR-210,miR-184,miR-71,miR-124	24/69	0.35	0.46	95	4	H14,R13,PDU-GLT,PDU-PL10A,H44,H21,Y125,Y110,Y19,PDU-VACHT,PDU-FVRI,Y134,Y102,H34,H2,Y62,H27,PDU-NOS,Y112,PDU-GATA123,Y118,Y74,PDU-PAX6,PDU-SYT,Y124,H43,Y111,R34,PDU-HOX1,H98,H12,R29,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68,Y11	41	0.45
serotonergic neurons	miR-137,miR-210,miR-10,miR-71,miR-124	6/36	0.17	0.25	81	8	PDU-PL10A,H44,H21,Y125,PDU-VACHT,PDU-FVRI,H2,Y62,PDU-GATA123,PDU-PAX6,PDU-SYT,H43,Y111,R34,PDU-HOX1,H98,H12,H40,PDU-ELAV,PDU-TRPH	22	0.24
glutamatergic neurons	miR-190,miR-210,miR-184,miR-71,miR-124	5/34	0.15	0.22	78	9	H14,PDU-GLT,H21,Y125,H34,H2,PDU-GATA123,Y118,PDU-PAX6,PDU-SYT,H43,R34,PDU-HOX1,H98,H12,H4,H22,H40,PDU-ELAV	19	0.21
pygidial lobe	miR-252,miR-8,miR-183,miR-263,miR-71	9/68	0.13	0.17	69	20	PDU-PL10A,Y110,PDU-FVRI,Y134,Y102,H27,Y74,Y132,PDU-SYT,H43,PDU-DLL,Y93,PDU-ELAV,Y68,Y11	15	0.16
midline	miR-315,miR-190,miR-71,miR-92,miR-124	12/54	0.22	0.31	90	5	PDU-PL10A,H44,H21,Y125,Y110,Y19,PDU-VACHT,Y102,H34,H2,Y62,PDU-PAX6,PDU-SYT,Y124,H43,Y111,R34,PDU-HOX1,H98,H12,R29,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68	28	0.30
parapodia sensory cells	miR-R,miR-252,miR-8,miR-183,miR-263,miR-71	17/80	0.21	0.31	98	0	H14,PDU-PL10A,H21,Y125,Y110,Y19,Y134,H91,Y102,H34,Y62,H27,PDU-VASA,PDU-NOS,Y74,PDU-SYT,H43,R34,PDU-HOX1,H98,H12,PDU-DLL,Y93,H4,H22,PDU-ELAV,Y68,Y11	28	0.30
posterior segment only	miR-219,miR-10	0/14	0.00	0.01	9	0	PDU-HOX4	1	0.01
posterior growth zone	miR-210	1/7	0.14	0.07	9	62	PDU-PL10A,H2,PDU-VASA,PDU-NOS,H12	5	0.05
mouth ring muscle	miR-281,miR-133,miR-22,miR-1	1/32	0.03	0.04	46	25	TROP1,Y133,R10,Y34	4	0.04
pharyngeal muscles	miR-133,miR-125,miR-22,miR-1	0/29	0.00	0.09	92	0	TROP1,PDU-GATA123,Y25,PDU-NK3,Tinman,R10,Y34,Y43	8	0.09
trunk muscles	miR-133,miR-125,miR-22,miR-1	0/29	0.00	0.03	61	0	TROP1,Y133,R10,Y34	3	0.03
gut	miR-12,miR-R,miR-125,miR-252,miR-263,LET-7,miR-216,miR-375,miR-210	54/102	0.53	0.46	7	92	PDU-BRA,R13,Y71,Y48,Y13,PDU-PL10A,Y115,PDU-BETA_CATENIN,H21,Y19,PDU-FVRI,R21,H32,Y123,PDU-VTN,PDU-TLX,H34,H2,Y62,H27,Y69,Y121,Y112,R20,PDU-SYT,Y124,Y111,Y45,Y25,PDU-NK3,PDU-HOX1,Y7,H98,H12,Tinman,Y63,PDU-NK2_1,Y108,R29,MEL,C,Y43,Y106,Y42	43	0.47
crescent cell	miR-317,miR-29,miR-22,miR-34,miR-92	4/51	0.08	0.10	70	17	R15,Y9,H91,Y105,R7,Y65,H8,Y132,Y51,Y98	10	0.11
head ciliated patches	miR-317,miR-29,miR-22,miR-277,miR-34,miR-92	6/76	0.08	0.12	87	7	R15,Y110,Y9,H91,Y105,R7,Y65,H8,Y132,Y51,Y98	11	0.12
prototroch	miR-317,miR-22,miR-34	4/29	0.14	0.15	48	32	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15
metatroch	miR-317,miR-29,miR-22,miR-277,miR-34,miR-92	10/76	0.13	0.15	64	21	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15
paratroch	miR-317,miR-29,miR-22,miR-34,miR-92	6/51	0.12	0.15	72	14	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15
telotroch	miR-317,miR-29,miR-22,miR-34,miR-92	6/51	0.12	0.15	72	14	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15

Table 2.6: **Co-expression of miRNAs and target transcripts in different tissues using Targetscan predictions only.** The co-expression ratio is defined as the number of miRNA::target pairs which show overlap in the given tissue divided by the total number of miRNA::target pairs. Iterative randomization was done by reshuffling the predicted targets of miRNAs. Bootstrap for depletions and enrichment shows total count of iterations when the observed co-expression ratio for a particular tissue was below or above the randomly obtained value, respectively (see methods).

	miRNAs expressed in tissue			Randomization			Genes expressed in tissue		
		Counts of target-miRNA pairs			Bootstrap percent support for				
Tissue	Names	#coexpressed in the tissue \ total count	Coexpression ratio	Average coexpression ratio after randomizations	Fewer coexpression events	More coexpression events	Names	total #genes	fraction of all genes
secretory neurons	miR-8,miR-31,miR-125,miR-R,miR-184,miR-190,miR-71,miR-124,miR-137	2/45	0.04	0.15	97	1	H21,H34,R20,PDU-SYT,H43,Y111,H98,H12,PDU-NK2_1,R29,H4,H22,H40,PDU-ELAV	14	0.15
pharyngeal epidermis	miR-9*,miR-31,miR-125,miR-9,miR-8	2/31	0.06	0.09	54	24	PDU-BRA,Y102,PDU-SYT,PDU-SIX3,H12,PDU-NK2_1,R29,Y68	8	0.09
miR-100 positive cells	miR-92,miR-31,miR-125,miR-278,miR-29,LET-7,miR-100,miR-375,miR-8,miR-184,miR-210	16/56	0.29	0.25	23	66	PDU-BRA,R13,Y49,Y13,PDU-BETA_CATENIN,R21,H32,Y123,Y68,Y121,PDU-SYT,PDU-SIX3,Y111,Y45,PDU-HOX1,H12,Y63,PDU-NK2_1,Y93,H40,PDU-ELAV,Y42,Y68	23	0.25
gland tissue	miR-278,miR-10	0/5	0.00	0.06	30	0	H44,PDU-FVRI,PDU-GATA123,PDU-SYT,H12,R29,PDU-ELAV	7	0.08
head epidermis	miR-315,miR-22,miR-8	5/28	0.18	0.33	97	1	H44,H21,Y125,Y110,Y19,Y134,H91,Y105,PDU-SIX3,R34,H98,H12,H27,PDU-NOS,Y124,Y74,Y112,H43,PDU-SIX3,R34,H98,H12,PDU-NK2_1,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68,Y11	30	0.33
antennae	miR-252,miR-9*,miR-315,miR-R,miR-9,miR-8,miR-183,miR-263	23/62	0.37	0.34	21	66	PDU-PL10A,H44,H21,Y125,Y110,Y19,Y134,Y102,PDU-TLX,PDU-ACETYLCHOLINE_RECEPTOR,H34,Y62,H27,PDU-VASA,PDU-NOS,Y112,Y74,Y132,PDU-PAX6,PDU-SYT,H43,PDU-SIX3,R34,H98,H12,PDU-DLL,Y93,H22,PDU-ELAV,Y106,Y68,Y11	31	0.34
sensory-associative centre	miR-7,miR-9*,miR-9,miR-210,miR-71,miR-124	8/30	0.27	0.24	25	54	H21,Y125,Y19,PDU-VACHT,H34,H2,Y62,H27,Y112,Y74,PDU-PAX6,PDU-SYT,H43,PDU-SIX3,R34,H98,H12,R29,Y93,H4,H22,PDU-ELAV	23	0.25
palpae	miR-125,miR-7,miR-252,miR-315,miR-R,miR-8,miR-219,miR-183,miR-210,miR-263,miR-184,miR-71,miR-124	44/84	0.52	0.46	7	91	H14,Y49,PDU-PL10A,H44,H21,Y125,Y110,Y19,PDU-FVRI,H91,Y102,PDU-TLX,H34,H13,H2,Y62,H27,Y89,PDU-NOS,Y112,R20,Y74,Y132,PDU-PAX6,PDU-SYT,H43,PDU-SIX3,Y111,R34,H98,H12,PDU-DLL,Y41,R29,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68,Y11	42	0.46
cirri	miR-8,miR-183,miR-263,miR-71,miR-124	7/36	0.19	0.19	33	42	H44,H21,Y125,Y19,PDU-FVRI,H34,R20,Y74,PDU-SYT,H43,R34,PDU-HOX1,H98,H12,Y93,H4,H22,PDU-ELAV	18	0.20
nuchal organ	miR-281,miR-317,miR-125,miR-7,miR-29,miR-315,miR-153,miR-8,miR-190,miR-71,miR-34,miR-92,miR-137	20/85	0.24	0.37	100	0	R15,H44,H21,Y125,Y110,Y19,Y134,Y9,H91,Y105,Y102,PDU-ACETYLCHOLINE_RECEPTOR,R7,H34,H2,Y62,Y65,R20,Y74,Y132,PDU-SYT,H43,PDU-SIX3,R34,H98,H12,R29,H4,H22,H40,PDU-ELAV,Y68,Y106,Y11	34	0.37
neurosecretory cells	miR-125,miR-137,miR-7,miR-29,miR-153,miR-190,miR-184,miR-71,miR-124	16/48	0.33	0.36	62	30	ACETYLCHOLINE_RECEPTOR,H34,H2,Y62,H27,PDU-SYT,H43,PDU-SIX3,Y111,R34,H98,H12,PDU-DLL,Y93,H4,H22,H40,PDU-ELAV,Y106,PDU-TRPH	33	0.36
dorsal sensory organ	miR-153,miR-8,miR-71,miR-315,miR-7,miR-183,miR-263,miR-137	3/64	0.05	0.11	90	1	Y125,Y134,H91,Y62,Y74,Y132,PDU-SYT,H43,Y68,Y11	10	0.11
adult eyes	miR-315,miR-7,miR-153,miR-71,miR-124,miR-137	7/43	0.16	0.28	96	2	H21,Y125,Y110,PDU-FVRI,Y134,Y102,PDU-ACETYLCHOLINE_RECEPTOR,H34,H2,Y62,H27,Y112,H38,Y74,Y132,PDU-SYT,H43,PDU-SIX3,Y111,R34,H98,H12,H4,H22,Y106,Y11	26	0.28
ventral plate neurons	miR-137,miR-R,miR-10,miR-219,miR-190,miR-210,miR-184,miR-71,miR-124	14/41	0.34	0.44	88	6	H14,R13,PDU-GLT,PDU-PL10A,H44,H21,Y125,Y110,Y19,PDU-VACHT,PDU-FVRI,Y134,Y102,H34,H2,Y62,H27,PDU-NOS,Y112,PDU-GATA123,Y118,Y74,PDU-PAX6,PDU-SYT,Y124,H43,Y111,R34,PDU-HOX1,H98,H12,R29,Y93,H4,H22,H40,PDU-ELAV,Y106,PDU-TRPH,Y68,Y11	41	0.45
serotonergic neurons	miR-137,miR-210,miR-71,miR-10,miR-124	4/26	0.15	0.22	77	11	PDU-PL10A,H44,H21,Y125,PDU-VACHT,PDU-FVRI,H2,Y62,PDU-GATA123,PDU-PAX6,PDU-SYT,H43,Y111,R34,PDU-HOX1,H98,H12,H4,H22,H40,PDU-ELAV,PDU-TRPH	22	0.24
glutamatergic neurons	miR-71,miR-190,miR-210,miR-184,miR-124	4/22	0.18	0.20	48	32	H14,PDU-GLT,H21,Y125,H34,H2,PDU-GATA123,Y118,PDU-PAX6,PDU-SYT,H43,R34,PDU-HOX1,H98,H12,H4,H22,H40,PDU-ELAV	19	0.21
pyloidial lobe	miR-252,miR-71,miR-8,miR-183,miR-263	7/43	0.16	0.17	41	46	PDU-PL10A,Y110,PDU-FVRI,Y134,Y102,H27,Y74,Y132,PDU-SYT,H43,PDU-DLL,Y93,PDU-ELAV,Y68,Y11	15	0.16
midline	miR-71,miR-315,miR-190,miR-92,miR-124	8/35	0.23	0.29	69	14	PDU-PL10A,H44,H21,Y125,Y110,Y19,PDU-VACHT,Y102,H34,H2,Y62,PDU-PAX6,PDU-SYT,Y124,H43,Y111,R34,PDU-HOX1,H98,H12,R29,Y93,H4,H22,H40,PDU-ELAV,Y106,Y68	28	0.30
parapodia sensory cells	miR-252,miR-R,miR-8,miR-183,miR-263,miR-71	11/43	0.26	0.31	73	14	H14,PDU-PL10A,H21,Y125,Y110,Y19,Y134,H91,Y102,H34,Y62,H27,PDU-VASA,PDU-NOS,Y74,PDU-SYT,H43,R34,PDU-HOX1,H98,H12,PDU-DLL,Y93,H4,H22,PDU-ELAV,Y68,Y11	28	0.30
posterior segment only	miR-219,miR-10	0/11	0.00	0.01	7	0	PDU-HOX4	1	0.01
posterior growth zone	miR-210	1/5	0.20	0.06	2	73	PDU-PL10A,H2,PDU-VASA,PDU-NOS,H12	5	0.05
mouth ring muscle	miR-281,miR-133,miR-22,miR-1	0/22	0.00	0.05	70	0	TROP|Y133,R10,Y34	4	0.04
pharyngeal muscles	miR-133,miR-125,miR-22,miR-1	0/19	0.00	0.08	80	0	TROP|PDU-GATA123,Y25,PDU-NK3,Tinman,R10,Y34,Y43	8	0.09
trunk muscles	miR-133,miR-31,miR-22,miR-1	0/19	0.00	0.04	55	0	TROP|R10,Y34	3	0.03
gut	miR-12,miR-125,miR-252,miR-283,LET-7,miR-216,miR-375,miR-R,miR-210	38/66	0.58	0.47	2	96	PDU-BRA,R13,Y71,Y49,Y13,PDU-PL10A,Y115,PDU-BETA_CATENIN,H21,Y19,PDU-FVRI,R21,H32,Y123,PDU-VTN,PDU-TLX,H34,H2,Y62,H27,Y89,Y121,Y112,R20,PDU-SYT,Y124,Y111,Y45,Y25,PDU-NK3,PDU-HOX1,Y7,H98,H12,Tinman,Y63,PDU-NK2_1,Y108,R29,MEL/C,Y43,Y106,Y42	43	0.47
crescent cell	miR-317,miR-29,miR-22,miR-34,miR-92	3/22	0.14	0.11	24	56	R15,Y19,H91,Y105,R7,Y65,H8,Y132,Y51,Y98	10	0.11
head ciliated patches	miR-317,miR-277,miR-29,miR-22,miR-34,miR-92	3/22	0.14	0.12	27	51	R15,Y110,Y9,H91,Y105,R7,Y65,H8,Y132,Y51,Y98	11	0.12
prototroch	miR-317,miR-22,miR-34	1/8	0.13	0.17	39	23	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15
metatroch	miR-317,miR-277,miR-29,miR-22,miR-34,miR-92	3/22	0.14	0.16	50	28	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15
paratroch	miR-317,miR-29,miR-22,miR-34,miR-92	3/22	0.14	0.16	50	28	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15
telotroch	miR-317,miR-29,miR-22,miR-34,miR-92	3/22	0.14	0.16	50	28	R15,Y134,Y9,H91,Y105,PDU-ACETYLCHOLINE_RECEPTOR,R7,Y65,Y118,H8,Y132,Y51,Y34,Y98	14	0.15

Table 2.7: **Co-expression of miRNAs and target transcripts in different tissues using the intersection between Targetscan and PITA predictions.** The co-expression ratio is defined as the number of miRNA::target pairs which show overlap in the given tissue divided by the total number of miRNA::target pairs. Iterative randomization was done by reshuffling the predicted targets of miRNAs. Bootstrap for depletions and enrichment shows total count of iterations when the observed co-expression ratio for a particular tissue was below or above the randomly obtained value, respectively (see methods).

3 Discussion

3.1 Insights into microRNA phylogeny from the *Platynereis* repertoire

Phylogenetic distribution of *Platynereis* microRNAs The wide conservation of 30 microRNAs across bilateria had already been reported (Prochnik et al., 2007) before the deep sequencing of the *Platynereis* repertoire. With very few available deep-sequenced repertoires at the time, this study had to rely on comparative microRNA gene predictions which unavoidably missed some of the conserved ancient bilaterian microRNAs.

With time, the deep sequencing of more and more metazoan microRNA repertoires, including *Platynereis* (Ruby et al., 2006; Stark et al., 2007; Grimson et al., 2008; Chen et al., 2009; Wheeler et al., 2009; Shi et al., 2009), allowed cross phyla comparisons to assemble a more accurate list of conserved microRNAs. Of particular value was the release of basal deuterostomes microRNA repertoires (Heimberg et al., 2008; Wheeler et al., 2009; Chen et al., 2009) for the identification of microRNAs which had been lost in the vertebrate lineage.

For instance, *Platynereis* miR-278 and miR-71, miR-315 and miR-252, originally appeared to be shared among protostomes only as they were non-identifiable in any vertebrate (Table 2.1). When identified in sea urchin, lamprey and amphioxus respectively, it became clear that they were present in the protostome-deuterostome ancestor (Urbilateria). Similarly, until parts of the sea urchin and the cephalochordate Amphioxus microRNA repertoires were published (Wheeler et al., 2009; Shi et al., 2009), miR-2001 and -242 which got secondarily lost in the ecdysozoan lineage (Wheeler et al., 2009) were thought to be lophotrochozoan -specific microRNAs. However, it appears from their conservation between protostomes (including *Platynereis*) and basal deuterostomes (Table 2.1) that they are ancient.

Lastly, extensive sequencing of microRNA repertoires across bilateria revealed with higher confidence that some microRNA* sequences appear to be conserved and unlike previously thought (Bushati and Cohen, 2007), functionally important. The deep-sequenced *Platynereis* repertoire provided a high quality reference for phylogenetic comparisons

through which I assembled the most up-to date and accurate list of conserved bilaterian microRNAs which must have present in Urbilateria.

Thanks to the recent sequencing of other Lophotrochozoan microRNA repertoires (Wheeler et al., 2009; Friedländer et al., 2009), I could assemble an additional list of lophotrochozoan -specific microRNAs. Nevertheless, the 454 sequencing used by Wheeler et al. (2009) does not provide as deep coverage as solexa does and the *Platynereis* sequencing data propose an additional 19 lophotrochozoan or annelid -specific microRNA candidates (see appendix), which map to at least one lophotrochozoan genome in a locus predicted to give rise to precursor microRNA according to known criteria (Hofacker et al., 1994; Lu et al., 2008b).

The *Platynereis* microRNA repertoire has therefore contributed in assembling the list of conserved microRNAs in two additional important evolutionary nodes in the tree of bilateria, reflecting the repertoires of ecdysozoa-lophotrochozoa and that of Annelid-Mollusc ancestors.

The size of the *Platynereis* repertoire The *Platynereis* microRNA repertoire comprised a total of at least 66 microRNAs (Table 2.1), the majority of which were experimentally validated by whole mount *in situ* hybridization or northern blot analysis. This number could increase up to 85 microRNAs if all remaining 19 lophotrochozoan -specific microRNA candidates prove to be real upon experimental validation.

However, this number would still not reflect the true size of the *Platynereis* repertoire as there are still at least 32 *Platynereis* -specific microRNA candidates remaining to be validated. Judging from the number of Planarian *Schmidtea mediterannea* -specific microRNAs which sum to 45 (Friedländer et al., 2009), it would not be surprising if *Platynereis* reached a total repertoire size of 100 microRNAs after including all *Platynereis* -specific microRNAs. Only when the *Platynereis* genome gets sequenced will the size of its microRNA repertoire be really known, since mapping of microRNAs on their corresponding genomic loci will be possible and will allow subsequent computational (Lu et al., 2008b; Hofacker et al., 1994) and experimental validation.

Even with the current state of data, the *Platynereis* repertoire nicely fits in (together with Planaria) the hierarchy of complexity which uses as a basis the microRNA repertoires of metazoa to rank them accordingly in a phylogenetic tree (Figure 3.1). The *Platynereis* repertoire is certainly bigger than that of a simpler Cnidarian (exhaustively searched by Grimson et al., 2008 through deep sequencing) and even if the total size of the repertoire exceeds 100 it will never reach the 400 microRNAs found in, more complex, zebrafish. Once both *Platynereis* repertoire is confidently assembled to a final number

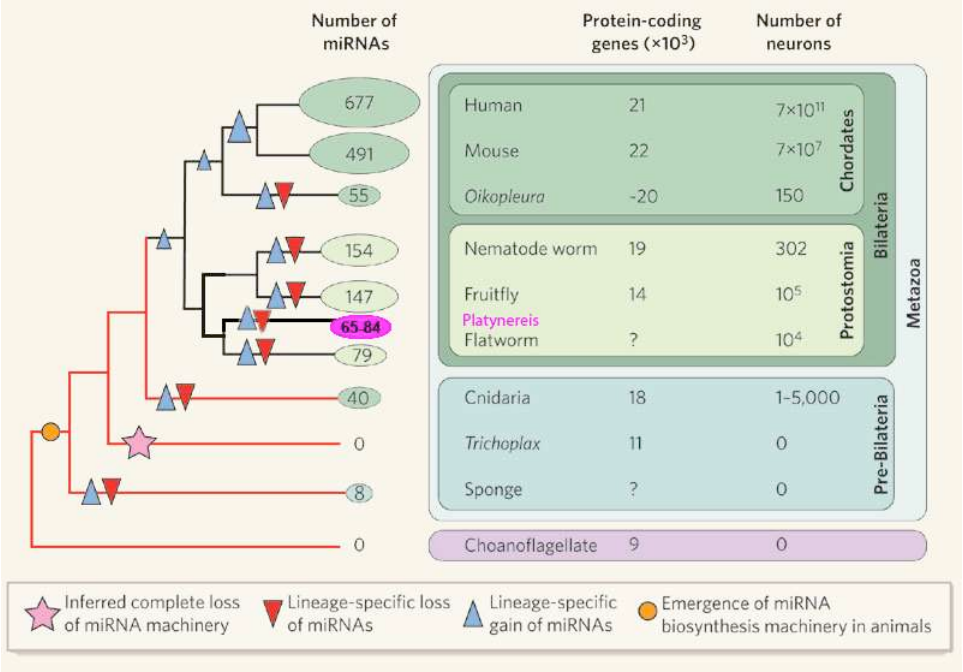


Figure 3.1: **Animal microRNA repertoires and morphological complexity.** Although the number of protein -coding genes does not change much, microRNA numbers correlate well with the total number of neurons found across metazoa. The *Platynereis* microRNA repertoire size fits in the clade of lophotrochozoa, exhibiting a similar number to Planaria. Figure adapted from Technau (2008)

and those of more metazoa also get exhaustively searched, it will be interesting to compare their sizes and see the link between repertoires and these animals' morphological complexities. So far such attempts look promising (Figure 3.1; Technau, 2008; Wheeler et al., 2009; Peterson et al., 2009) but include very few animal microRNA repertoires which have been exhaustively searched. In other words, the picture is still incomplete, but a trend is apparent and the incomplete repertoire of *Platynereis* fits this trend.

Sequence families within *Platynereis* microRNAs Even after imposing the strictest quality cutoff to *Platynereis* solexa sequencing data, some microRNAs have as many as 397 sequence variants (see "VARS" column in Table 2.1 on page 58). It is noteworthy that "younger" lophotrochozoan -specific microRNAs do not have too many sequence variants. This probably indicates that the microRNA genes from which they derive have had less time to evolve compared to ancient bilaterian microRNA genes. Concominant to this, evolutionarily "newly" acquired *Platynereis* -specific microRNAs do not show any sequence variants at all. In the future, the possibility to map all Platynereis microRNAs to a sequenced genome will give insight into the number of microRNA paralogs and their evolutionary history.

A sequenced genome would also allow comparison between the sequences of microRNA genomic loci and those of cloned microRNAs, in search for RNA editing events which could also potentially account for some of these sequence variants. At the moment I can only speculate on RNA editing being one additional reason, as it cannot be confirmed. Similarly, genome availability would allow to check for the possibility that multiple microRNA loci which belong to the same sequence family (as reported for miR-2 in *Drosophila*; Lai et al., 2003), give rise to a multitude of sequence related mature microRNAs.

The extent of naturally occurring sequence variants maybe intensified from polymorphisms which characterize our worm culture. In *Platynereis*, SNPs are not rare to find (1/45bp as calculated by O. Simakov) and part of this maybe due to mixed populations of worms in the Heidelberg breeding culture which were originally captured in north and Mediterranean sea and have since interbred for the past 60 years.

Finally, heterologous processing by Dicer and Drosha (Ruby et al., 2007; Landgraf et al., 2007) is also likely to occur in *Platynereis* as sequence variants of many different microRNAs had shortened 3' ends (Table 1.2.5 on page 21). In fewer cases 5' ends had a 2nt offset (Table 1.2.5 on page 21), something which in principle frame shifts the seed and functional target group of the microRNA (Wheeler et al., 2009).

Enriched microRNA sequences in *Platynereis* zygotic libraries Although the majority of *Platynereis* microRNAs were not detectable in zygotic libraries, miR-8, -22, -29, -2, -277 and -279 were detected with more than 100 reads reaching as many as 1817 reads. Since the freshly fertilized zygotes which were used for these libraries are transcriptionally inactive (unpublished data A. Fischer), these microRNAs must be maternally deposited.

The functional role of microRNAs in early embryonic development has been underlined by several independent studies in *Drosophila*, fish and mouse (Bushati et al., 2008; Giraldez, 2006; Tang et al., 2007a) but so far nothing is known about the role of maternally deposited microRNAs in animals like *Platynereis*, which exhibit spiral cleavage patterns during zygotic development (Fischer and Dorresteijn, 2004). Depletion of these "early" microRNAs by antagomiRs (complementary LNA oligos) or even a Dicer knockout should reveal the importance of the role they may play in early zygotic development.

3.2 The site of microRNA activity in important evolutionary nodes

Before involving *Platynereis* in microRNA expression comparative studies, the link between vertebrate and protostome microRNA expression patterns was only inferrable for very few microRNAs (miR-1, -124). This was due to the limited number of available whole mount *in situ* patterns first in *Drosophila* and later Planaria (Aboobaker et al., 2005; Gonzalez-Estevez et al., 2009) and due to the difficulties in establishing the protocol to work in *C.elegans*. Moreover, the distinct localisation patterns microRNAs have acquired in such fast evolving protostome species due to the concomitant functional diversification of microRNAs (Liu et al., 2008) posed a problem for comparisons to vertebrate data. Finally, comparisons between the localisation patterns of primary microRNA transcripts (pri-miRNAs) and mature microRNAs maybe inaccurate and limiting. Recent studies uncovered a mechanism which post-transcriptionally regulates microRNA maturation (Heo et al., 2008; Lehrbach et al., 2009) and may explain how a broad primary microRNA transcript (pri-miRNA) expression does not necessarily entail an equally broad expression of its mature microRNA (Obernosterer, 2006; Thomson et al., 2006). The *Platynereis* mature microRNA expression dataset constitutes the only of this kind among protostomes, since *Drosophila* and Planaria data focused on the expression of pri-miRNAs (Aboobaker et al., 2005; Gonzalez-Estevez et al., 2009). Maybe this is an additional reason why *Platynereis* could give a better insight into the site of expression of conserved bilaterian microRNAs.

3.2.1 Higher tissue specificity in *Platynereis* microRNA expression

In *Platynereis*, all expression patterns of ancient bilaterian microRNAs were highly specific to tissues or cell types. This high tissue specificity maybe explained by the fact that *Platynereis* is a slow evolving marine annelid species (Raible et al., 2005) which has not changed its habitat in millions of years and has therefore not accumulated too many changes from having to adapt to new ecological niches (Weistheide and Rieger, 1996). Thus, it is more likely for slow evolving species like *Platynereis* to retain the most "fundamental" expression site of a microRNA and for fast evolving species to functionally diversify their microRNAs and express them in additional sites. From the vertebrate data where microRNA expression patterns are broader and often ubiquitous (Wienholds et al., 2005; Ason et al., 2006), it is clear that more specific roles have evolved for these microRNAs depending on the functionality of these ancient urbilaterian tissues.

The comparisons to *Platynereis* whole mount *in situ* hybridization results, indicate that all conserved bilaterian microRNAs evolved in a strictly tissue-specific context, expressed in CNS, musculature, cilia, foregut or gut. Corroborating this, all brain- or musculature- specific microRNAs in *Platynereis* show the same specificity in fish, human and mouse expression profiling, with high confidence (Table 2.4 on page 89). Restricting the comparison of expression sites between *Platynereis* and fish only, reveals even more shared sites of microRNA expression (such as sensory organs and pancreas/gut; Table 2.4 on page 89). All the above, indicate that these highlighted tissue affinities were already in place in the protostome-deuterostome ancestor.

3.2.2 Revealing the site of ancient bilaterian microRNAs activity

The identification of ancient expression sites for conserved bilaterian microRNAs sheds new light on the evolution of animal body plans, because it implies that the following microRNA-defined tissues were in place already in the last common ancestor of protostomes and deuterostomes (Urbilateria):

miR-100 and the related miR-125 and let-7 (blue in Figure 3.2) may have acquired an ancient role in developmental timing: The very late onset of let-7 expression at 5dpf in *Platynereis* (before benthic settlement) is consistent with earlier observations in nematode (Reinhart et al., 2000), fly (Caygill and Johnston, 2008; Sokol et al., 2008), mollusc, zebrafish and in another annelid (Pasquinelli et al., 2000), indicative of a role in the control of late developmental transitions (Pasquinelli et al., 2000; Reinhart et al., 2000; Caygill and Johnston, 2008; Sokol et al., 2008). With time, the role of miR-100, -125 and let-7 in developmental timing may have spread to other tissues, such as target tissue

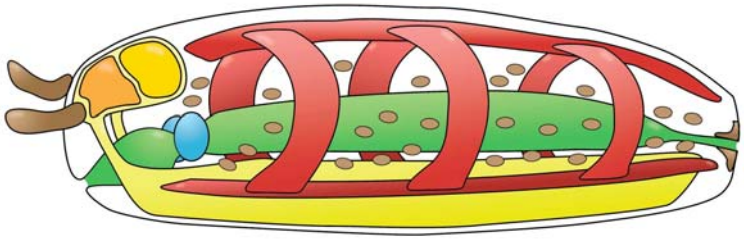


Figure 3.2: **Defining the identity of ancient tissues in the last common ancestor of protostomes and deuterostomes.** Schematised drawing of microRNA-defined tissues. CNS is drawn in yellow, neurosecretory brain tissue in dark yellow, sensory-associative brain parts in orange, musculature in red, PNS in brown, gut in green and the let-7-positive cells in blue. Sceme drawn by A. Fischer.

differentiating in the course of metamorphosis, as observed in fly (Caygill and Johnston, 2008; Sokol et al., 2008) and nematode (Pasquinelli et al., 2000; Reinhart et al., 2000).

In a similar manner, the affinity of miR-29, -34 and -92 to motile ciliated cells in both protostome and deuterostome larvae may shed new light on the evolution of ventricular neuron types expressing these microRNAs in the vertebrate CNS (Table 2.4 on page 89; Kapsimali et al., 2007), some of which known to bear motile cilia (Vigh et al., 2004). Prior to this finding, no link had ever been established by classical morphological studies of the ciliated cerebrospinal fluid (CSF)-contacting neurons (Vigh et al., 1983, 2004) and of motile ciliated cells of larval ciliary bands (Burke, 1978).

The comparative expression analysis using microRNAs also suggests that the Urbilateria already possessed a miR-124+ central nervous tissue as opposed to miR-8/-183/-263+ peripheral sensory nervous tissue (Denes et al., 2007; Rusten et al., 2002; yellow and brown in Figure 3.2), consistent with recent comparative developmental data suggesting that nervous system centralization predated the protostome/deuterostome ancestor (Arendt et al., 2008), see however (Holland, 2003). These ancestors possessed an ancient brain comprising miR-7+, - 137+ and -153+ neurosecretory parts, as recently suggested (Tessmar-Raible et al., 2007) and other miR-9/9*+ parts that may relate to sensory information processing (light and dark orange in Figure 3.2). Deep evolutionary conservation of "sensory" brain parts is currently unclear but the conserved activity of miR-9/-9* is an entry point to readdress this issue. The protostome/deuterostome ancestors had miR-1/-22/-133+ body musculature, in line with the proposed conservation of somatic muscle cell types (Denes et al., 2007) and a miR-12/-216/-283+ gut, in accordance with the view that the gut represents another conserved bilaterian trait, with (Arendt et al., 2001) or without (Hejnol and Martindale, 2008) anus (red and green in Figure 3.2).

At the present state of analysis, without having studied the functional role of any of

these ancient microRNAs in *Platynereis*, I can only speculate about the tissue-specific ancient roles of these microRNAs, by extrapolating from observations in other animal models. Still, if a microRNA remains highly specifically expressed in the same tissue over 600 million years it is likely to play a role there (what role this was should be subject to future functional analysis). The comparative expression data indicate that these ancestors had a biphasic life-cycle (Arendt et al., 2001), with populations of miR-100/-125/let-7/+ neurosecretory cells along the larval mouth controlling developmental timing (blue in Figure 3.2). These cells were part of the conserved larval brachyury+ foregut tissue (Arendt et al., 2001). Swimming larva had otx+ ciliary bands (Harada et al., 2000; Arendt et al., 2001) bearing miR-29, -34, -92+ locomotor cilia.

MicroRNAs have thereby been established as an important new tool for reconstructing ancient animal body plans at important evolutionary nodes, focusing here on the protostome-deuterostome divergence and confirming from a different direction that Urbilateria was indeed more complex than originally thought in line with previous studies (Denes et al., 2007; Tessmar-Raible et al., 2007; Robertis, 2008). More complete inventories of microRNAs and concomitant expression analysis will allow expanding this approach to other key events of animal evolution.

3.2.2.1 The targets of ancient bilaterian microRNAs in *Platynereis*

Previous analyses which compared the expression of microRNAs with that of their targets in *Drosophila* (Stark et al., 2005) and mammals (Sood et al., 2006; Farh, 2005) had come to contradicting conclusions which until recently (Shkumatava et al., 2009) remained unresolved (see introduction for details). Whole mount *in situ* data from *Drosophila* indicated that microRNAs and their targets are expressed in a mutually exclusive manner (Stark et al., 2005), while array data from mammals detected targets co-expressed in the same tissues as the microRNAs which regulate them (Farh, 2005). More recent work clarified that both mutual exclusion and co-expression might be observed between microRNAs and targets in a given tissue and that mRNAs with evolutionarily conserved target sites tend to be co-expressed with their targeting microRNAs (Shkumatava et al., 2009).

Expression analysis on ancient microRNAs' *Platynereis* targets confirmed findings of Shkumatava et al. (2009) by revealing tissues in which mutual exclusion was observed between the expression of microRNAs and of their targets (Stark et al., 2005) and tissues that showed the opposite trend with microRNAs and targets found co-expressed (Farh, 2005; Sood et al., 2006; Shkumatava et al., 2009). However, the lack of sequenced closely related species to *Platynereis dumerilii* with alignable 3' UTRs obstructed the distinction

between evolutionarily conserved versus non-conserved target sites. As a consequence, the *Platynereis* dataset can so far not contribute to support or disprove the current hypothesis put forward by Shkumatava et al. (2009), that microRNA-target interactions are predominantly of the tuning type for conserved/important target sites (co-expressed) and of switch or fail-safe type for non-conserved target sites (mutually exclusive).

However, a unique feature this analysis had to offer was the tissue based approach allowed by the strict tissue/cell type signatures microRNAs exhibit in *Platynereis*. Intriguingly, the tissues/cell types in which microRNAs show a clear tendency to target mRNAs that are not co-expressed, such as secretory neurons of the foregut, the nuchal organ and adult eyes (Table 2.6 on page 109), are already differentiated by 5dpf. On the contrary, the gut, an actively differentiating tissue at 5dpf, was the only example where microRNAs target co-expressed genes (Table 2.6 on page 109).

In an actively differentiating tissue, stoichiometry of various transcription factors and signalling molecules is essential for patterning and therefore, tuning microRNA interactions would be ideal to keep all targets at stable, optimal levels (Iovino et al., 2009). On the other hand, in a differentiated tissue, left-over transcripts or leaky transcripts from neighbouring developmental programs can disturb the adopted cell fate and in this case microRNAs could have a fail-safe/switch role in tissue clearance of unwanted messengers (Cohen et al., 2006; Hornstein and Shomron, 2006; Bushati and Cohen, 2007; Bartel, 2009). It is therefore tempting to speculate that aside from the evolutionary importance of a target site, the differentiation state of a tissue could also adjust microRNA-target interactions. It will be interesting to further elucidate the temporal dynamics of microRNA targeting during animal development.

Inclusion of more developmental stages to check for co-expression events before or after 5dpf could give a more dynamic range to this analysis. Future use of the WMISEP protocol that R. Tomer developed in the laboratory, could give an insight into targeting relationships within the 48h brain and trunk. With more than 100 genes already aligned in WMISEP database, single cell-resolution co-expression analysis could be performed for all aligned microRNAs and their predicted targets.

Finally, a comparison of targets in slow-evolving species should reveal the functional evolution of conserved bilaterian microRNAs and shed light into why microRNAs were co-opted by new tissues that did not express them before.

3.2.3 The site of lineage specific microRNAs activity

Conserved microRNAs among protostomes (bantam, miR-2, -12, -67, -87, -277, -279, -305, -317, -318, -981, -989, -993, -996), the lineage of lophotrochozoans (miR-36, -96,

-133*, -750, -1175, -1175*, -1986, -1989, -1992, -1993, -1994, -), the lineage of annelids (miR-1996, -1997, -1998, -L49) and finally *Platynereis* specific microRNAs (miR-C, -K), showed equally strict tissue/cell type specific localisation as ancient bilaterian microRNAs did. Protostome specific microRNAs for which expression data were available in at least one more protostome were always expressed in homologous cell types and tissues as the *Platynereis* site of expression, highlighting the evolutionary conservation of these expression sites (see miR-2, -87, -279, -36 and bantam in Figure 2.20 on page 92 and 2.23 on page 95). It therefore becomes apparent that strict tissue demarcation by microRNAs was not only a feature of the protostome-deuterostome ancestor but also of the ecdysozoa-lophotrochozoa ancestor as well as the annelids-moluscs ancestor.

The diversity of microRNA-demarcated tissues in the protostome-deuterostome and the ecdysozoa-lophotrochozoa ancestors, does not appear to vary a lot. Most of the examined protostome specific microRNAs were expressed in motile ciliated cells, CNS, PNS and gut with just a couple, bantam (Figure 2.20 on page 92) and miR-277 (Figure 2.22 on page 94), showing more restricted expression patterns in subset of CNS and ciliated cells respectively. Similarly, lophotrochozoan specific microRNAs demarcated cilia, CNS, PNS and gut with again, few exceptions like miR-1992 (Figure 2.24 on page 96), -1175 (Figure 2.25 on page 98) and -C (Figure 2.24 on page 96) showing more restricted affinity to a subset of CNS, gut and PNS respectively.

All the above observations do not support the hypothesis put forward by Peterson et al. (2009) whereby the newly acquired lineage specific microRNAs can contribute towards speciation by regulating gene expression in novel cell types and tissues. Had this hypothesis been true, I would expect to see a striking specialization in most lineage specific microRNAs and find them localized in very specific cell types and tissues which are characteristic of the lineage (such as a nerve ring underlying ciliated larval swimming organs of lophotrochozoa; Wanninger, 2009). This microRNA specialization was only visible for 2 out of the 7 examined protostome specific microRNAs and 3 out of the 11 lophotrochozoan specific microRNAs. However, it is important to bare in mind that I only examined the localisation in 53% of protostome specific microRNAs and 56% of lophotrochozoan microRNAs (without including the remaining 19 lophotrochozoan-specific candidates which if all real would change the percentage to just 25%). Therefore, the chance that I randomly chose to study the expression of the most broadly expressed microRNAs remains to be elucidated for the lophotrochozoan set.

3.3 Interesting cell types of *Platynereis* highlighted by microRNAs

Some of the microRNA-expressing tissues and cell types of *Platynereis* were not noticed or linked among themselves before this large microRNA *in situ* hybridization screen I performed. The conserved tissue affinity of the ancient bilaterian microRNAs together with any available bibliography on the function of tissue specific microRNAs in other animals, gave a hints to speculate about a cell type's/tissue identity and functional role in *Platynereis* development and behavior.

***Platynereis* foregut and the miR-100/let-7+ cells** Before the identification of miR-100+ cells in *Platynereis*, no special focus had been attributed to the foregut. In an attempt to characterize their position and molecular fingerprint, I unavoidably explored the identity of all neighbouring cells to realize that the foregut is a complex organ, populated by at least 4 distinct tissue types (secretory neurons, chemosensory neurons, miR-100/let-7 positive cells and muscle cells) expressing different microRNA markers in a non-overlapping fashion (Figure 2.4 on page 67) . This was a first insight into *Platynereis* stomatogastric nervous system which was earlier under the shade of brain (Tessmar-Raible et al., 2007) and ventral nerve cord (Denes et al., 2007).

The microRNA expression screen played a catalytic role in the discovery of miR-100/let-7+ cells, which had not been identified histologically. Of special interest is these cells' molecular fingerprint, which in the current state of analysis relates most to the pituitary (they are neurosecretory neurons co-expressing miR-375, a pituitary marker in vertebrates Wienholds et al., 2005, and preliminary experiments have localized the expression of gonadotropin release hormone receptor- Gnrhr- in these same cells). However, these cells position in *Platynereis* pharynx, together with their possible involvement in developmental timing of settlement (discussed below), may also draw links to the thyroid which plays a role in vertebrates metamorphosis (Buchholz and Hayes, 2005; Furlow and Neff, 2006).

Intriguingly, miR-100/let-7+ cells appear late in development (after 3dpf) and start expressing let-7 at 5dpf, which is when *Platynereis* young worm begin to feed on algae and eventually decide to settle on the benthic floor and spend the rest of their lives there (Fischer and Dorresteijn, 2004). By position, the cells are in direct contact to the digestive tract (Figure 2.4 on page 67) and their co-expression of cilia specific microRNA markers (Figure 2.6 on page 71) suggested they may bare ciliary protrusions pointing towards the digestive lumen. Ciliated cells in the pharynx of other lophotrochozoa have

already been reported (Tzetlin and Purschke, 2006) but i could not detect any stained cilia (by looking for acetylated tubulin staining adjacent to these cells) in *Platynereis* miR-100/let-7+ cells. Still, even in the absence of protruding cilia, these cells are in direct contact with ingested food. Given the role let-7 plays in developmental timing of other invertebrates (Reinhart et al., 2000; Caygill and Johnston, 2008; Sokol et al., 2008), it is tempting to speculate that *Platynereis* miR-100/let-7+ cells may "sense" the first food uptake of the young worm and generate a systemic response (through neuropeptide secretion) which could contribute to their decision on settlement.

Functional characterization of these cells may shed more light upon this inferred link to developmental timing of settlement. Since they are a visible organ pair in the 5dpf foregut, laser ablations can be done to test their putative involvement in settlement. Additionally, functional interference of miR-100 and let-7 should also give insight into the role these microRNAs may play in *Platynereis* developmental timing.

Gland tissue demarcated by miR-10 and miR-278 The V-shaped expression domain of these two microRNAs (Figure 2.3 on page 64), revealed a set of cells abutting the mouth opening, which were completely uncharacterized before in *Platynereis*. With a very small fraction of their molecular fingerprint assembled, it is clear these are secretory neurons (expressing *syt* and *phc2*). Their proximal position to the mouth opening and to miR-100/let-7+ cells, together with the reported functional role miR-278 has in *Drosophila* hormonal control of energy homeostasis (Teleman et al., 2006) foster speculations that their role in *Platynereis* might be equivalent to that of a gland tissue connecting to the stomatogastric system.

A subset of FVRI-amidergic cells expressing miR-219 Among the many FVRI-amidergic cells which can be found in the brain and trunk of 5dpf *Platynereis*, two previously unnoticed cells additionally express miR-219 in the palpae and another two in the posterior end of the trunk (Figure 2.14 on page 82). This microRNA plays a role in mammalian circadian rythmicity (Cheng et al., 2007) and is also involved in NMDA receptor signalling (Kocerha et al., 2009) by targeting key genes of underlying pathways. Interestingly, the expression of NMDA-receptors in the mollusk *Lymnaea stagnalis* is localised in the buccal ganglia (Ha et al., 2006), where FVRI-amidergic neurons have been identified with projections reaching and innervating buccal muscles (Filali et al., 2006). Although no direct link exists between these FVRI-amidergic neurons of mollusks buccal ganglia and miR-219, it would be interesting to localize the expression of *Platynereis* NMDA-receptors and see if they are co-expressed with miR-219 in the palpae (the closest clump

of neurons *Platynereis* has near the mouth).

Mouth ring muscle and nuchal organ expressing miR-281 The ring muscle of the mouth opening has always been reported by morphologists as part of the general musculature of other lophotrochozoa (Wanninger, 2009). However, no molecular marker had previously revealed that it has a unique molecular fingerprint compared to remaining differentiated muscles. Intriguingly, in *Platynereis*, miR-281 is not only expressed in the ring muscle but also in the chemosensory, nuchal organ (Purschke et al., 1997). Prior to this result, no general musculature gene marker has been additionally found expressed in the nuchal organ (see Table 2.5 on page 108). It would therefore be interesting to explore this putative link between the mouth and nuchal organ.

Of high interest is the gene in which *Platynereis* miR-281 maybe hosted, and whether this gene's transcription takes place in both cells of nuchal organ and ring muscle. From studies in *Drosophila*, it is known that the biogenesis of this particular microRNA is not related to the expression of its host gene ODA (Xiong et al., 2009). Functional interference against miR-281 should give a very specific readout in either or both cell types considering its highly restricted expression pattern as well as the easily recognizable structures it demarcates (Figure 2.18 on page 86).

Glutamatergic neurons expressing bantam The recent assignment of a role to bantam in circadian rythmicity of *Drosophila* (Kadener et al., 2009a), together with the known involvement of *Drosophila* glutamatergic neurons in rhythmic metabotropic glutamate signaling (Hamasaka et al., 2007), gave reasons to speculate over the specific and suggestive bantam expression observed in *Platynereis*. Bantam, which in *Platynereis* demarcates glutamatergic neurons of both brain and trunk as well as the adult eyes and sensory neurons of the mouth (Figure 2.20 on page 92), may have a conserved role contributing to the circadian rythmicity of *Platynereis* too. Although the full circadian circuit of *Platynereis* is not yet entirely resolved (work in progress by M. Tosches in the laboratory), the expression of some *Drosophila* circadian neurons' marker genes in the vicinity of *Platynereis* glutamatergic neurons suggests that the latter might also be involved in the *Platynereis* circadian system (M. Tosches, unpublished data). Functional interference against bantam should ultimately reveal if it is indeed involved in *Platynereis* circadian rythmicity.

Uknown epidermal cells expressing *Platynereis* -specific miR-C This *Platynereis* specific microRNA poses an example of a newly acquired microRNA which exhibits a more

specific and "novel" expression pattern when compared to ancient bilaterian microRNAs. However, it is only during larval stages (at 48h: see Figure 2.24 on page 96) when miR-C is highly specific and restricted to few unknown epidermal cells of the apical brain. Later in development it adopts a general PNS expression pattern (Figure 2.24 on page 96). Further marker gene analysis should elucidate the identity and molecular fingerprint of these apical larval brain epidermis cells and subsequently their role in larval development. Functional interference against such newly acquired and highly specific lineage specific microRNAs is highly interesting to understand their importance in speciation (Peterson et al., 2009).

Cells adjacent to the pharyngeal muscle sheath express miR-1175 Aside from its gut expression, miR-1175 is also found localized in single cells lining the external side of pharyngeal muscle (Figure 2.25 on page 98). These cells have not been characterized so far and they identity remains to be elucidated by further marker gene analysis. miR-1175 is one of the newly acquired microRNAs which exhibit a more specific expression pattern and it would interesting to interfere with its function and see if any lophotrochozoan specific trait is affected.

Photoreceptor cells expressing miR-1992 Photoreceptor cells have been thoroughly studied in *Platynereis* by current and previous members of our lab and many gene markers have been found to specifically label rhabdomeric or ciliary photoreceptors (Arendt et al., 2002, 2004; Tessmar-Raible et al., 2007). What makes miR-1992 a special marker gene is that it is expressed in all photoreceptor cells regardless on whether they are ciliary or rhabdomeric (Figure 2.24 on page 96). No other gene marker links these cells' photoreceptor identity by taking part in different photoreceptors molecular fingerprints except for miR-1992. It will be very interesting to see whether functional interference with miR-1992 can potentially affect larval, adult eye and ciliary photoreceptors development.

3.4 Post-transcriptional microRNA regulation is a conserved mechanism across bilateria

Out of the 4 localized primary microRNA transcripts (pri-miRNAs), it appears that the biogenesis of 3 (pri-miR-100-let-7, pri-miR-183-263 and pri-miR-1992) seems to be controlled at the post-transcriptional level (Figure 2.26 and 2.27 on page 100). Their mature microRNA counterparts were found localized in much more restricted patterns concominant with previous studies (Obernosterer, 2006; Thomson et al., 2006; Heo et

al., 2008; Lehrbach et al., 2009). Recent work in *C.elegans* (Lehrbach et al., 2009) and the above *Platynereis* expression data indicate that post-transcriptional control in microRNA biogenesis is a conserved mechanism which can affect the expression of both ancient bilaterian as well as younger lineage specific microRNAs. It therefore makes litle sense to try and compare mature microRNA expression patterns (detected using locked nucleic acids) to pri-miRNA expression patterns (detected using long riboprobes) since the pri-miRNA expression pattern does not necessarily reflect that of its mature microRNA.

So far, the underlying mechanism of microRNA post-transcriptional control has only been described for let-7. The precursor of let-7 gets recognized and bound by target Lin-28 which blocks its processing (Heo et al., 2008; Lehrbach et al., 2009). The extent of this mechanism beyond let-7 and Lin28 is very interesting because a second evolutionary mode of anti-targeting (besides target site depletion) could involve the ability of at least some target proteins to bind their targeting pre-miRNAs and control their maturation.

3.5 A model for the role of microRNAs in the evolution of tissue identity

Delegation of newly emerged microRNAs to specific tissues Many independent studies show that microRNAs are often found in genomic clusters (Altuvia et al., 2005; Megraw et al., 2007) or introns of protein coding genes (Rodriguez et al., 2004; Baskerville and Bartel, 2005), employing already available promoter-enhancer motifs for their expression (Ohler et al., 2004; Lee et al., 2007). Hence, the genomic locus of a microRNA is instrumental on the role it shall play in the evolution of a given tissue's identity, because the microRNA is likely to "follow" the potentially tissue -specific expression of its host gene (if it is intronic) or one of the genes in the surroundings of its genomic locus (Baskerville and Bartel, 2005; Wang et al., 2009).

Among vertebrates, the genomic architecture of microRNA loci is commonly conserved (observation after searching ensemble depicted in Figure 3.3) and in many cases, at least one functionally -related gene (when considering the tissue or cell type where the microRNA is expressed) is within 50kb distance from the microRNA. A good example is that of miR-375 (Figure 3.3), which in all vertebrate genomes lies within 10-20kb distance from the gene *crystallin beta A2* (*cryba2*), coding for a structural protein of the vertebrate eye lens. Aside from lens, cryba2 is found enriched in the pancreas and pituitary just like miR-375 (Hoffman et al., 2008; Poy et al., 2004; Wienholds et al., 2005), suggesting a common transcriptional regulation.

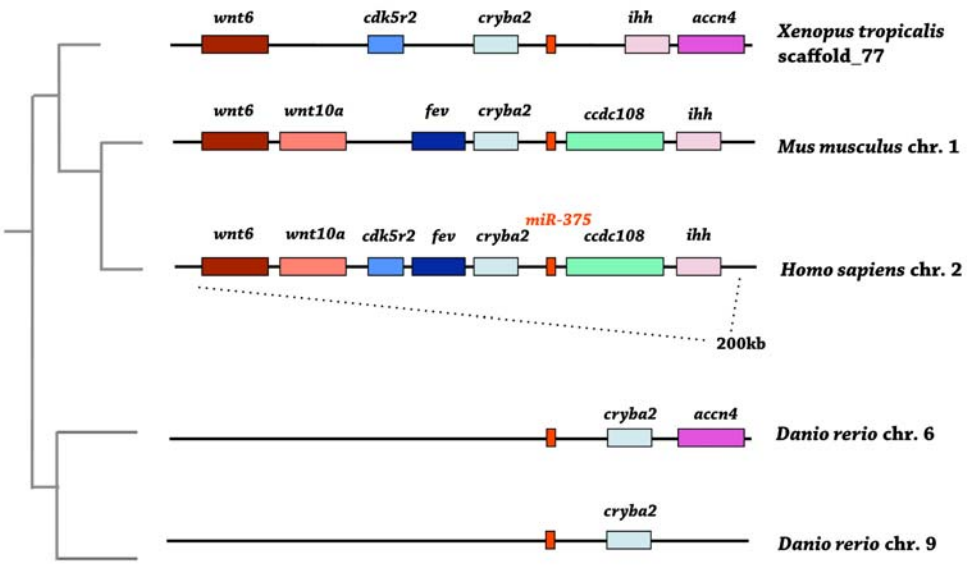


Figure 3.3: **Conserved genomic architecture of miR-375 genomic locus across vertebrates.** miR-375 is always found within 10-20kb distance from *crystallin beta A2* (*cryba2*) gene. Scheme summarizes a comparative survey i made using Ensemble.

It therefore seems that newly evolved microRNAs "find themselves" in charge of newly regulating several co-expressed transcripts in the tissue where their host gene or neighbouring gene is specifically expressed. Their "passive" delegation to newly regulate a tissue, becomes subject to natural selection which can "mold" the transcriptome so as to deplete deleterious target sites (anti-targeting) and stabilize beneficial target-microRNA interactions (Bartel, 2009; Shomron et al., 2009).

Tandem duplications of the same microRNA can lead to the creation of a microRNA cluster in that same locus (Chen and Rajewsky, 2007; Shabalina and Koonin, 2008). In this scenario, microRNA regulation of this tissue gets reinforced by the passive delegation of additional microRNAs (Bentwich et al., 2005). In *Platynereis*, I observed co-expression between microRNAs which do not appear to be clustered in the genome (miR-7, -137 and -153). It therefore appears that these microRNAs independently emerged in three distinct genomic loci in which transcriptional regulation drives expression to the neurosecretory brain. All three independent layers of microRNA regulation must have been positively selected to find them co-expressed across bilateria (Figure 2.9 and 2.10;Kapsimali et al., 2007). It would be very interesting to test how many common targets these three microRNAs share in order to assess whether they complement or enhance each other as a functional role in defining neurosecretory brain identity (attempted but the number of targets is too low to have a significant statistical result).

Speculations over the role of newly acquired microRNAs in the evolution of tissues

In *Platynereis*, acquisition of lineage specific microRNAs seems to impose additional layers of regulation in the protein output of a handful of tissues (as most lineage specific microRNAs are expressed in CNS, PNS, gut, ciliated cells). A major evolutionary implication of controlling protein output within a tissue in such a multilayer manner is robustness. The more microRNAs are recruited to buffer/stabilize the output of different developmental programs which take place in a tissue, the higher the programs' precision can be. In this manner, the tissue's phenotypic variation is stabilized within a population, thus increasing its heritability and "evolvability" by natural selection (Peterson et al., 2009).

Since newly acquired microRNAs appear to join the expression of ancient ones in the same tissues they demarcate, it is fair to speculate that tissues are already formed by the time microRNAs come into play. If this is true, then hypotheses which involve microRNAs in the morphological innovations of lineages (Peterson et al., 2009) may be misleading, because the microRNAs do not seem to induce but rather stabilize the pre-existing tissues and cell types. Therefore, the role of cis regulatory control over spatial

gene expression is what assembles gene networks in tissues (Davidson, 2006) and then microRNAs are secondarily, "passively", recruited by the same cis elements, to control the stability of these networks output and enhance their evolvability.

It still remains to be seen to what extent, additional layers of regulation by new microRNA acquisitions target co-expressed mRNAs of the tissue in an additive, complementary or both manners. Studying the targets of co-expressed microRNAs in their tissue context may shed new light in the role of microRNAs may play in tissue evolution.

3.6 Open questions and future outlook

To better understand the role of microRNAs in the evolution of animal tissues and cell types, an exhaustive and comparative target analysis will be necessary at least for microRNAs expressed in a particular homologous organ/tissue across bilateria. Despite the wide conservation of protein coding genes across metazoa (Putnam et al., 2007), their 3'UTRs (where microRNAs predominantly target) are not conserved between different phyla, resulting to very few conserved targeting interactions across bilaterian phyla (Chen and Rajewsky, 2006a; Karres et al., 2007; Li et al., 2009). In theory, independent acquisition of the same microRNA by two distinct evolutionary lineages could have a different effect in defining the identity of the same tissue (given that in both lineages the microRNA is expressed in the same tissue) because it may target different transcripts and affect the output of different proteins. An insight in the evolution of target interactions in conserved tissue contexts will therefore be very informative and my work has set the basis by revealing the ancient bilateria tissues which have been under the control of microRNAs.

Concominant studies over the functional role of newly acquired, lineage specific, microRNAs may also crystallize speculations over their role in evolution of lineage specific morphological traits (Peterson et al., 2009). As a first step, further localisation of lophotrochozoan and *Platynereis* specific microRNAs may shed some light by revealing their conserved site of action. In cases of highly restricted expression, interference against the microRNA regulation should reveal how important it is for the development of this specific cell type/tissue it demarcates and possibly suggest a role in speciation.

Once the genome of *Platynereis* gets published, mapping all *Platynereis* microRNAs to their genomic loci should give a wealth of information over their transcriptional control. It will be interesting to see to what extent they get co-transcribed with their host genes or genes in the local vicinity, by performing a large *in situ* screen which includes both microRNAs and co-transcribed genes. This should also reveal the extent of post-

transcriptional regulation of microRNAs and give insights into its evolutionary implications. As a start, such a screen could be done using the closely related annelid *Capitella spI* which our laboratory grows in culture and for which the genome is available.

4 Materials and methods

4.1 microRNA cloning and sequencing

Small RNA cloning was performed separately for three different developmental stages of *Platynereis*: 20 minutes post fertilization, swimming larvae (15hpf-51hpf) and young worms (72hpf-11dpf). For each stage 4 batches were pooled. RNA extraction and cloning of the small RNA fraction (19-24 nucleotides) was performed as described in(Pfeffer et al., 2005). The resulting small RNA libraries were sequenced using the Illumina platform.

4.2 Sequence processing and analysis

Sequence processing and analysis For each of the three libraries, raw Illumina reads were processed as follows: first, adaptor sequences (5'-CTGTAGGCACCATCAAT-3') were clipped off by searching for the motif CTGTAGGC (allowing one degenerate position). Subsequently, sequences were selected for length (16 <= x <= 29) and quality (base scores >=35 out of a maximum of 40). Sequences of the individual libraries were then pooled into one non-redundant data set. Finally, around 4% of these sequences were excluded, as database searches revealed significant similarity to known mitochondrial or ribosomal sequences.

The resulting non-redundant sequence dataset comprised 111,575 sequences, representing 1,259,482 individual sequence reads. Known miRNA sequences were identified from this set by sequence searches with sequences retrieved from the miRBase repository v10.18.). Comparisons required 80% overall identity of the query with the target and 7 out of 8 nucleotides in the 5'end. Work done by F. Raible.

4.2.1 Sequence variants

3,735 sequences of the dataset represent ~89% (1,114,984) of all the reads, whereas the remaining reads occur less than 10 times. This general disproportion in distribution reflects a combination of (a) methodological artifacts, (b) original heterogeneity of alleles present in the batches pooled for the sampling, and (c) differences in miRNA processing,

as previously observed in large miRNA datasets generated by massive parallel sequencing (Lu et al., 2008b; Morin et al., 2008). Without information about the complete *Platynereis* genome, it is not possible to reliably distinguish between these possibilities. However, in several cases, sequence variants account for significant fractions of a given miRNA family. This is illustrated for the case of miR-9 (Table 1.2.5). 71% of the respective RNAs represent the canonical sequence, whereas the remainder includes length variants and point mutations, including a 5' variant that would be predicted to shift the seed region involved in target recognition. miR-22/-745, showed similar seed shifting (Table 1.2.5). In this case, a recent study (Wheeler et al., 2009) indeed supports the notion that the respective sub-classes of miRNAs reflect genetic divergence of the family. Work done by F. Raible.

4.3 *Platynereis* and *Capitella* whole-mount *in situ* hybridization with LNA probes

Custom miRCURY LNA™(Valoczi et al., 2004) Detection Probes (Exiqon) for *Platynereis* were labeled with digoxigenin (DIG) 3'-End labeling Kit-2nd generation (Roche) and then purified with Sephadex G25 MicroSpin columns (GE Healthcare) as done in (Wienholds et al., 2005). *Platynereis* and *Capitella* embryos were fixed at room temperature in 4% paraformaldehyde in 2x PBS + 0.1% Tween-20 and stored in methanol at -20°C. In situ hybridizations were performed by fusing and modifying fish and *Platynereis* whole-mount *in situ* (WMISH) protocols (Wienholds et al., 2005; Jékely and Arendt, 2007) as follows: Stepwise 5 minute rehydration of fixed embryos in PBS+01% Tween (PTW) to dilute methanol from 100%, 75%, 50%, 25% to 0%. Digestion times of embryos in 10µg/ml Proteinase K PTW: *Capitella* larvae (stage 3- 7): 5min. *Platynereis* embryos 24hpf: 7min; 48hpf: 30min; 72hpf: 60min; 5dpf: 75min. Two 5min washes in 2mg/ml glycine PTW followed by 20min postfixation in 4%PFA PTW. Five 5min washes in PTW. Prehybridization in high stringency hyb- mix (HM) for 1 hour at 37 ° C (70% formamide, 5x SSC, 50µg/ml Heparin, 5mg/ml tRNA, 0.1% Tween 20) followed by overnight hybridization at 37 ° C in 200µl of high stringency HM+4pmol labeled LNA probe. Posthybridization washes were done in standard HM (50% formamide, 5x SSC, 50µg/ml Heparin, 5mg/ml tRNA, 0.1% Tween 20) at 37 ° C stepwise: 15min 75% HM/ 25% 2x SSC, 15min 50% HM/ 2x SSC, 15min 25% HM/ 2x SSC followed by two 15min 2xSSC and two 30min 0.2xSSC washes. Then at room temperature: 10min 75% 0.2x SSC/25% PTW, 10min 50% 0.2x SSC/50% PTW, 10min 25% 0.2x SSC/25% PTW, 10min PTW and blocking for 90min in 2% sheep serum in PTW. AP-coupled anti-DIG-Fab' frag-

ments (Roche) were diluted 1:2000, mouse anti- acetylated tubulin (Sigma, St. Louis, MO,USA) 1:500 and rabbit anti-serotonin (5HT) 1:250 (Immunostar Inc. USA) in PTW for overnight incubation at 4 ° C. Embryos were washed six times 15min in PTW, twice in staining buffer (100 mM Tris, pH 9.5, 100 mM NaCl, 50 mM MgCl2, 0.1% Tween 20), and stained with 337.5 µg/ml nitroblue tetrazolium (NBT) and 175 µg/ml 5-bromo-4-chloro-3-indolyl phosphate (BCIP) in staining buffer for several hours. After staining, embryos were rinsed in 100% ethanol for 15min, washed in PTW and immunostaining proceeded as in (Jékely and Arendt, 2007) to be finally mounted in glycerol containing 2.5 mg/ml DABCO. The same *in situ* protocol was used for the hybridization of long DIG-11-UTP-labeled RNA probes.

4.4 *Nematostella* and sea urchin whole-mount *in situ* hybridization with LNA probes

Fixations and hybridizations of *Nematostella* carried out as described in (Rentzsch et al., 2006) and for sea urchin larvae as described in (Arenas-Mena et al., 2000) with the following adaptations: Content of formamide in hybridization mix increased to 70% instead of 50%. Pre-hybridization and hybridization carried out at 37 ° C instead of 60 ° C.

4.5 Northern Blot analysis

Total RNA was isolated from *Platynereis* embryos at 6hpf, 12hpf, 18hpf, 24hpf, 48hpf and 5dpf using peqGOLD TriFast reagent (peqlab Biotechnologie). 15 µg of total RNA was separated on 15% denaturing polyacrylamide gel and blotted via chemical cross-linking as in (Pall and Hamilton, 2008). For detection, oligonucleotide probes were designed using the known sequences for each *Platynereis* miRNA (Table 2.1) and were end-labeled with 32P-ATP using T4 polynucleotide kinase (NEB).

4.6 Microscopy

NBT/BCIP stained embryos were imaged using a Leica TCS SPE confocal microscope through reflection imaging as described by (Jékely and Arendt, 2007). A 40x oil-immersion objective was used. White light pictures were taken under Nomarsky optics using a Zeiss Axiophot microscope equipped with a Leica DC500 camera.

4.7 Image processing

Confocal stacks images were processed using ImageJ 1.40g. All confocal images displayed are products of z-projections of stacks. Image reconstructions and sections were generated using Imaris 6.2.1. Contrast was adjusted uniformly across the entire image.

4.8 Primer sequences for miRNA cluster cloning

Gene specific cDNA libraries were prepared using the 3' prime most miRNA sequence of miR-100-let-7 cluster, miR-12-miR-216 cluster and miR-183-miR-263 cluster. Pdu-miR-100-let-7 cluster (FJ838789) resulted from a PCR reaction using the forward primer AACCCGTACAACCGAACTTGTG and the reverse primer ACTATACAACCTACTACCTCA. Pdu-miR-12-miR-216 cluster (FJ838790) resulted from a PCR reaction using the forward primer UGAGTATTACATCAGGTACTGA and the reverse primer CTCACTTTTGCCAGCTGAGATTA. Pdu-miR-183-miR-263 cluster resulted from a PCR reaction using the forward primer AATGGCACTGGTAGAATTCACGG and the reverse primer CTTGGCACTGGTAGAATTCACTGA. The RNAfold program (Hofacker, 2003) was used to find the hairpin structure within the cluster.

4.9 miRNA target prediction analysis

92 3'-UTRs were surveyed for 8mer, 7mer-A1 and 7mer-m8 binding sites (Bartel, 2009) using Targetscan (Lewis et al., 2005) and for 8mer, 7mer and 6mer perfect seed binding sites using PITA (Kertesz et al., 2007) with default settings. Targetscan yielded 433 and PITA 741 binding sites. The intersection between the two datasets contains 293 binding sites. A sliding window of either 1or 2nt was allowed in order to capture binding sites that were predicted as a 6mer in PITA but as 7mer or 8mer in Targetscan. Done by K. Trachana.

4.10 Calculation of significance for co-expression of miRNAs and target transcripts

For tissues, the co-expression ratio is defined as the number of miRNA::target pairs that show overlap in the given tissue divided by the total number of miRNA::target pairs. Iterative randomization was done by reshuffling the predicted targets of miRNAs. Bootstrap for depletions and enrichment shows total count of iterations when the observed

coexpression ratio for a particular tissue was below or above the randomly obtained value, respectively.

4.11 3'-UTR retrieval

The majority of the EST sequences used to find 3' UTRs were 3' Sanger reads of a full length cDNA library. BLAST was used to search the sequences against swissprot protein database to identify the correct open reading frame, and thus the 3' UTR. In few cases where there was no significant BLAST hit, we assumed it to be mostly composed of UTR, as they contained frequent stop codons in all the reading frames. Done by R. Tomer.

4.12 SNP assessment in the 3'UTRs

Total RNA extracted from heads of adult animals and different developmental stages was sequenced with the Solexa platform at EMBL yielding 15 million 76bp reads, 21000 of which could be mapped to 3' UTRs by BLASTN. Reads were truncated to the first 40bp in order to reduce sequencing errors. Every position in the 3' UTR was analyzed for the presence of a mismatch and a SNP was called when less than 90% of nucleotides assigned to the position showed conservation. Whereas on average we observed SNP rate of 1/45bp in the non-coding regions, the predicted miRNA binding sites showed a rate of 1/60bp. Still, the overall SNP frequency was too low in order to distinguish between conserved and non-conserved miRNA binding sites for the entire set of predicted targets. Done by O. Simakov.

Bibliography

Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC (2005) Drosophila microrna exhibit diverse spatial expression patterns during embryonic development. *Proc Natl Acad Sci USA* 102:18017–22.

Abraham MC, Lu Y, Shaham S (2007) A morphologically conserved nonapoptotic program promotes linker cell death in caenorhabditis elegans. *Dev Cell* 12:73–86.

Acampora D, Postiglione MP, Avantaggiato V, Bonito MD, Vaccarino FM, Michaud J, Simeone A (1999) Progressive impairment of developing neuroendocrine cell lineages in the hypothalamus of mice lacking the orthopedia gene. *Genes & Development* 13:2787–800.

Affymetrix (2009) Post-transcriptional processing generates a diversity of 5'-modified long and short rnas. *Nature* 457:1028–32.

Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387:489–493.

Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H (2005) Clustering and conservation patterns of human micrornas. *Nucleic Acids Res* 33:2697–2706.

Ambros V, Horvitz HR (1987) The lin-14 locus of caenorhabditis elegans controls the time of expression of specific postembryonic developmental events. *Genes Dev* 1:398–414.

Ambros V (2008) The evolution of our thinking about micrornas. *Nat Med* 14:1036–40.

Andersson MG, Haasnoot PCJ, Xu N, Berenjian S, Berkhout B, Akusjärvi G (2005) Suppression of rna interference by adenovirus virus-associated rna. *J Virol* 79:9556–65.

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

Arenas-Mena C, Cameron AR, Davidson EH (2000) Spatial expression of hox cluster genes in the ontogeny of a sea urchin. *Development* 127:4631–4643.

Arendt D, Nubler-Jung K (1999) Comparison of early nerve cord development in insects and vertebrates. *Development* 126:2309–2325.

Arendt D, Technau U, Wittbrodt J (2001) Evolution of the bilaterian larval foregut. *Nature* 409:81–5.

Arendt D (2005) Genes and homology in nervous system evolution: comparing gene functions, expression patterns, and cell type molecular fingerprints. *Theory Biosci* 124:185–197.

Arendt D (2008) The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* 9:868–882.

Arendt D, Denes AS, Jékely G, Tessmar-Raible K (2008) The evolution of nervous system centralization. *Philos Trans R Soc Lond, B, Biol Sci* 363:1523–8.

Arendt D, Tessmar K, de Campos-Baptista MIM, Dorresteijn A, Wittbrodt J (2002) Development of pigment-cup eyes in the polychaete platynereis dumerilii and evolutionary conservation of larval eyes in bilateria. *Development* 129:1143–54.

Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodt J (2004) Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* 306:869–71.

Ason B, Darnell DK, Wittbrodt B, Berezikov E, Kloosterman WP, Wittbrodt J, Antin PB, Plasterk RHA (2006) Differences in vertebrate microrna expression. *Proc Natl Acad Sci USA* 103:14385–9.

Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC (2008) Characterization of endogenous human argonautes and their mirna partners in rna silencing. *Proc Natl Acad Sci USA* 105:7964–9.

Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microrna on protein output. *Nature* 455:64–71.

Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, Eachus R, Pasquinelli AE (2005) Regulation by let-7 and lin-4 mirnas results in target mrna degradation. *Cell* 122:553–563.

Bak M, Silahtaroglu A, Moller M, Christensen M, Rath MF, Skryabin B, Tommerup N, Kauppinen S (2008) Microrna expression in the adult mouse central nervous system. *RNA* 14:432–444.

Bartel DP (2004) Micrornas: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297.

Bartel DP (2009) Micrornas: target recognition and regulatory functions. *Cell* 136:215–33.

Baskerville S, Bartel DP (2005) Microarray profiling of micrornas reveals frequent coexpression with neighboring mirnas and host genes. *RNA* 11:241–247.

Beckstead RB, Lam G, Thummel CS (2005) The genomic response to 20-hydroxyecdysone at the onset of drosophila metamorphosis. *Genome Biol* 6:R99.

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E (2006) mrna degradation by mirnas and gw182 requires both ccr4:not deadenylase and dcp1:dcp2 decapping complexes. *Genes Dev* 20:1885–1898.

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z (2005) Identification of hundreds of conserved and nonconserved human micrornas. *Nat Genet* 37:766–770.

Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC (2007) Mammalian mirtron genes. *Mol Cell* 28:328–36.

Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RHA (2006) Diversity of micrornas in human and chimpanzee brain. *Nat Genet* 38:1375–1377.

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ (2003) Dicer is essential for mouse development. *Nat Genet* 35:215–217.

Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microrna.org resource: targets and expression. *Nucleic Acids Res* 36:D149–D153.

Boekhoff-Falk G (2005) Hearing in drosophila: development of johnston's organ and emerging parallels to vertebrate ear development. *Dev Dyn* 232:550–8.

Bohnsack MT, Czaplinski K, Gorlich D (2004) Exportin 5 is a rangtp-dependent dsrna-binding protein that mediates nuclear export of pre-mirnas. *RNA* 10:185–91.

Borchert GM, Lanier W, Davidson BL (2006) Rna polymerase iii transcribes human micrornas. *Nat Struct Mol Biol* 13:1097–101.

Brennecke J, Cohen SM (2003) Towards a complete description of the microrna complement of animal genomes. *Genome Biol* 4:228.

Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM (2003) bantam encodes a developmentally regulated microrna that controls cell proliferation and regulates the proapoptotic gene hid in drosophila. *Cell* 113:25–36.

Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microrna-target recognition. *Plos Biol* 3:e85.

Britten RJ, Kohne DE (1968) Repeated sequences in dna. hundreds of thousands of copies of dna sequences have been incorporated into the genomes of higher organisms. *Science* 161:529–540.

Buchholz DR, Hayes TB (2005) Variation in thyroid hormone action and tissue content underlies species differences in the timing of metamorphosis in desert frogs. *Evol Dev* 7:458–67.

Burke R (1978) Structure of nervous-system of pluteus larva of strongylocentrotus purpuratus. *Cell Tissue Res* 191:233–247.

Bushati N, Cohen SM (2007) microrna functions. *Annu Rev Cell Dev Biol* 23:175–205.

Bushati N, Stark A, Brennecke J, Cohen SM (2008) Temporal reciprocity of mirnas and their targets during the maternal-to-zygotic transition in drosophila. *Curr Biol* 18:501–6.

Carthew RW, Sontheimer EJ (2009) Origins and mechanisms of mirnas and sirnas. *Cell* 136:642–655.

Caygill EE, Johnston LA (2008) Temporal regulation of metamorphic processes in drosophila by the let-7 and mir-125 heterochronic micrornas. *Curr Biol* 18:943–50.

Cayirlioglu P, Kadow IG, Zhan X, Okamura K, Suh GSB, Gunning D, Lai EC, Zipursky SL (2008) Hybrid neurons in a microrna mutant are putative evolutionary intermediates in insect co2 sensory systems. *Science* 319:1256–60.

Chen CZ, Li L, Lodish HF, Bartel DP (2004) Microrna modulate hematopoietic lineage differentiation. *Science* 303:83–6.

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang DZ (2006) The role of microrna-1 and microrna-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 38:228–233.

Chen K, Rajewsky N (2006a) Deep conservation of microrna-target relationships and 3'utr motifs in vertebrates, flies, and nematodes. *Cold Spring Harb Symp Quant Biol* 71:149–56.

Chen K, Rajewsky N (2006b) Natural selection on human microrna binding sites inferred from snp data. *Nat Genet* 38:1452–6.

Chen K, Rajewsky N (2007) The evolution of gene regulation by transcription factors and microrna. *Nat Rev Genet* 8:93–103.

Chen X, Li Q, Wang J, Guo X, Jiang X, Ren Z, Weng C, Sun G, Wang X, Liu Y, Ma L, Chen JY, Wang J, Zen K, Zhang J, Zhang CY (2009) Identification and characterization of novel amphioxus microrna by solexa sequencing. *Genome Biol* 10:R78.

Chen Y, Segarini P, Raoufi F, Bradham D, Leask A (2001) Connective tissue growth factor is secreted through the golgi and is degraded in the endosome. *Exp Cell Res* 271:109–117.

Chendrimada TP, Finn KJ, Ji X, Baillat D, Gregory RI, Liebhaber SA, Pasquinelli AE, Shiekhattar R (2007) Microrna silencing through risc recruitment of eif6. *Nature* 447:823–8.

Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R (2005) Trbp recruits the dicer complex to ago2 for microrna processing and gene silencing. *Nature* 436:740–4.

Cheng HYM, Papp JW, Varlamova O, Dziema H, Russell B, Curfman JP, Nakazawa T, Shimizu K, Okamura H, Impey S, Obrietan K (2007) microrna modulation of circadian-clock period and entrainment. *Neuron* 54:813–29.

Chuang JC, Mathers PH, Raymond PA (1999) Expression of three rx homeobox genes in embryonic and adult zebrafish. *Mech Dev* 84:195–8.

Cohen SM, Brennecke J, Stark A (2006) Denoising feedback loops by thresholding–a new role for microrna. *Genes & Development* 20:2769–2772.

Corbin JG, Rutlin M, Gaiano N, Fishell G (2003) Combinatorial function of the homeodomain proteins nkx2.1 and gsh2 in ventral telencephalic patterning. *Development* 130:4895–906.

Cui Q, Yu Z, Purisima EO, Wang E (2007) Microrna regulation and interspecific variation of gene expression. *Trends Genet* 23:372–375.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J (2008) An endogenous small interfering rna pathway in drosophila. *Nature* 453:798–802.

Darnell DK, Kaur S, Stanislaw S, Davey S, Konieczka JH, Yatskievych TA, Antin PB (2007) Geisha: an in situ hybridization gene expression resource for the chicken embryo. *Cytogenet Genome Res* 117:30–5.

Davidson EH (2006) *The regulatory genome. Gene regulatory networks in development and evolution* Academic Press Elsevier.

Denes AS, Jékely G, Steinmetz PRH, Raible F, Snyman H, Prud'homme B, Ferrier DEK, Balavoine G, Arendt D (2007) Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. *Cell* 129:277–88.

Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ (2004) Processing of primary micrornas by the microprocessor complex. *Nature* 432:231–235.

Deo M, Yu JY, Chung KH, Tippens M, Turner DL (2006) Detection of mammalian microrna expression by in situ hybridization with rna oligonucleotides. *Dev. Dyn.* 235:2538–2548.

Diederichs S, Haber DA (2007) Dual role for argonautes in microrna processing and posttranscriptional regulation of microrna expression. *Cell* 131:1097–108.

Ding XC, Grosshans H (2009) Repression of c. elegans microrna targets at the initiation level of translation requires gw182 proteins. *EMBO J* 28:213–22.

Doench JG, Sharp PA (2004) Specificity of microrna target selection in translational repression. *Genes Dev* 18:504–511.

Duan R, Pak C, Jin P (2007) Single nucleotide polymorphism associated with mature mir-125a alters the processing of pri-mirna. *Hum Mol Genet* 16:1124–1131.

Easow G, Teleman AA, Cohen SM (2007) Isolation of microrna targets by mirnp immunopurification. *RNA* 13:1198–1204.

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.

Erkner A, Roure A, Charroux B, Delaage M, Holway N, Coré N, Vola C, Angelats C, Pagès F, Fasano L, Kerridge S (2002) Grunge, related to human atrophin-like proteins, has multiple functions in drosophila development. *Development* 129:1119–29.

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E (2007) P-body formation is a consequence, not the cause, of rna-mediated gene silencing. *Mol Cell Biol* 27:3970–3981.

Eulalio A, Huntzinger E, Nishihara T, Rehwinkel J, Fauser M, Izaurralde E (2009) Deadenylation is a widespread effect of mirna regulation. *RNA* 15:21–32.

Eulalio A, Rehwinkel J, Stricker M, Huntzinger E, Yang SF, Doerks T, Dorner S, Bork P, Boutros M, Izaurralde E (2007) Target-specific requirements for enhancers of decapping in mirna-mediated gene silencing. *Genes Dev* 21:2558–2570.

Fabian MR, Mathonnet G, Sundermeier T, Mathys H, Zipprich JT, Svitkin YV, Rivas F, Jinek M, Wohlschlegel J, Doudna JA, Chen CYA, Shyu AB, Yates JR, Hannon GJ, Filipowicz W, Duchaine TF, Sonenberg N (2009) Mammalian mirna risc recruits caf1 and pabp to affect pabp-dependent deadenylation. *Mol Cell* 35:868–880.

Farh KKH (2005) The widespread impact of mammalian micrornas on mrna repression and evolution. *Science* 310:1817–1821.

Fedonkin M, Waggoner B (1997) The late precambrian fossil kimberella is a mollusc-like bilaterian organism. *Nature* 388:868–871.

Feinbaum R, Ambros V (1999) The timing of lin-4 rna accumulation controls the timing of postembryonic developmental events in caenorhabditis elegans. *Developmental Biology* 210:87–95.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397–405.

Filali ZE, Minnen JV, Liu WK, Smit AB, Li KW (2006) Peptidomics analysis of neuropeptides involved in copulatory behavior of the mollusk lymnaea stagnalis. *J Proteome Res* 5:1611–7.

Fischer A, Dorresteijn A (2004) The polychaete platynereis dumerilii (annelida): a laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *Bioessays* 26:314–325.

Flynt AS, Li N, Thatcher EJ, Solnica-Krezel L, Patton JG (2007) Zebrafish mir-214 modulates hedgehog signaling to specify muscle cell fate. *Nat Genet* 39:259–63.

Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD (2007) Drosophila micrornas are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* 130:287–97.

Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD (2005) Normal microrna maturation and germ-line stem cell maintenance requires loquacious, a double-stranded rna-binding domain protein. *PLoS Biol* 3:e236.

Friedländer MR, Adamidi C, Han T, Lebedeva S, Isenbarger TA, Hirst M, Marra M, Nusbaum C, Lee WL, Jenkin JC, Alvarado AS, Kim JK, Rajewsky N (2009) High-resolution profiling and discovery of planarian small rnas. *Proc Natl Acad Sci USA* 106:11546–51.

Friedman RC, Farh KKH, Burge CB, Bartel DP (2009) Most mammalian mrnas are conserved targets of micrornas. *Genome Res* 19:92–105.

Furlow JD, Neff ES (2006) A developmental switch induced by thyroid hormone: Xenopus laevis metamorphosis. *Trends Endocrinol Metab* 17:40–47.

Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M (2007) Inference of mirna targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* 8:69.

Giraldez AJ (2006) Zebrafish mir-430 promotes deadenylation and clearance of maternal mrnas. *Science* 312:75–79.

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF (2005) Microrna regulate brain morphogenesis in zebrafish. *Science* 308:833–838.

Gonzalez-Estevez C, Arseni V, Thambyrajah RS, Felix DA, Aboobaker AA (2009) Diverse mirna spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *Int. J. Dev. Biol.* 52:1–14.

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human risc couples microrna biogenesis and posttranscriptional gene silencing. *Cell* 123:631–40.

Griffiths-Jones S, Saini HK, Dongen SV, Enright AJ (2007) mirbase: tools for microrna genomics. *Nucleic Acids Research* 36:D154–D158.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) mirbase: tools for microrna genomics. *Nucleic Acids Res* 36:D154–D158.

Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) Microrna targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27:91–105.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP (2008) Early origins and evolution of micrornas and piwi-interacting rnas in animals. *Nature* 455:1193–1197.

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC (2001) Genes and mechanisms related to rna interference regulate expression of the small temporal rnas that control c. elegans developmental timing. *Cell* 106:23–34.

Gwizdek C, Ossareh-Nazari B, Brownawell AM, Doglio A, Bertrand E, Macara IG, Dargemont C (2003) Exportin-5 mediates nuclear export of minihelix-containing rnas. *J Biol Chem* 278:5505–8.

Ha I, Wightman B, Ruvkun G (1996) A bulged lin-4/lin-14 rna duplex is sufficient for caenorhabditis elegans lin-14 temporal gradient formation. *Genes Dev* 10:3041–3050.

Ha TJ, Kohn AB, Bobkova YV, Moroz LL (2006) Molecular characterization of nmda-like receptors in aplysia and lymnaea: relevance to memory mechanisms. *Biol Bull* 210:255–70.

Hamasaka Y, Rieger D, Parmentier ML, Grau Y, Helfrich-Forster C, Nassel DR (2007) Glutamate and its metabotropic receptor in drosophila clock neuron circuits. *J Comp Neurol* 505:32–45.

Hammell CM, Lubin I, Boag PR, Blackwell TK, Ambros V (2009) nhl-2 modulates microrna activity in caenorhabditis elegans. *Cell* 136:926–38.

Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN (2006) Molecular basis for the recognition of primary micrornas by the drosha-dgcr8 complex. *Cell* 125:887–901.

Harada Y, Okai N, Taguchi S, Tagawa K, Humphreys T, Satoh N (2000) Developmental expression of the hemichordate otx ortholog. *Mech Dev* 91:337–339.

Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ (2008) Micrornas and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci USA* 105:2946–50.

Hejnol A, Martindale MQ (2008) Acoel development indicates the independent evolution of the bilaterian mouth and anus. *Nature* 456:382–386.

Heo I, Joo C, Cho J, Ha M, Han J, Kim VN (2008) Lin28 mediates the terminal uridylation of let-7 precursor microrna. *Molecular Cell* 32:276–284.

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF, of Bioinformatics Computer Labs 2004 S, 2005 (2006) The expansion of the metazoan microrna repertoire. *BMC Genomics* 7:1–26.

Hill SD, Boyer BC (2003) Hnk-1/n-cam immunoreactivity correlates with ciliary patterns during development of the polychaete capitella sp. i. *Biol Bull* 205:182–4.

Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH (2001) Quality control of mrna 3'-end processing is linked to the nuclear exosome. *Nature* 413:538–542.

Hofacker I, Fontana W, Stadler P, Bonhoeffer L, Tacker M, Schuster P (1994) Fast folding and comparison of rna secondary structures. *Monatsh Chem* 125:167–188.

Hofacker IL (2003) Vienna rna secondary structure server. *Nucleic Acids Res* 31:3429–3431.

Hoffman BG, Zavaglia B, Witzsche J, de Algara TR, Beach M, Hoodless PA, Jones SJM, Marra MA, Helgason CD (2008) Identification of transcripts with enriched expression in the developing and adult pancreas. *Genome Biol* 9:R99.

Holland ND (2003) Early central nervous system evolution: an era of skin brains? *Nat Rev Neurosci* 4:617–627.

Hornstein E, Mansfield JH, Yekta S, Hu JKH, Harfe BD, Mcmanus MT, Baskerville S, Bartel DP, Tabin CJ (2005) The microrna mir-196 acts upstream of hoxb8 and shh in limb development. *Nature* 438:671–674.

Hornstein E, Shomron N (2006) Canalization of development by micrornas. *Nat Genet* 38 Suppl:S20–4.

Hristova M, Birse D, Hong Y, Ambros V (2005) The caenorhabditis elegans heterochronic regulator lin-14 is a novel transcription factor that controls the developmental timing of transcription from the insulin/insulin-like growth factor gene ins-33 by direct dna binding. *Mol Cell Biol* 25:11059–11072.

Hughes NC (2007) Strength in numbers: High phenotypic variance in early cambrian trilobites and its evolutionary implications. *Bioessays* 29:1081–4.

Humphreys DT, Westman BJ, Martin DIK, Preiss T (2005) Micrornas control translation initiation by inhibiting eukaryotic initiation factor 4e/cap and poly(a) tail function. *Proc Natl Acad Sci U S A* 102:16961–16966.

Hutvagner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD (2001) A cellular function for the rna-interference enzyme dicer in the maturation of the let-7 small temporal rna. *Science* 293:834–8.

Iovino N, Pane A, Gaul U (2009) mir-184 has multiple roles in drosophila female germline development. *Dev Cell* 17:123–33.

Jafar-Nejad H, Acar M, Nolo R, Lacin H, Pan H, Parkhurst SM, Bellen HJ (2003) Senseless acts as a binary switch during sensory organ precursor selection. *Genes Dev* 17:2966–2978.

Jakymiw A, Lian S, Eystathioy T, Li S, Satoh M, Hamel JC, Fritzler MJ, Chan EKL (2005) Disruption of gw bodies impairs mammalian rna interference. *Nat Cell Biol* 7:1267–1274.

Jékely G, Arendt D (2007) Cellular resolution expression profiling using confocal detection of nbt/bcip precipitate by reflection microscopy. *Biotech.* 42:751–755.

Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q (2005) Dicer-1 and r3d1-l catalyze microrna maturation in drosophila. *Genes Dev* 19:1674–1679.

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ (2005) Ras is regulated by the let-7 microrna family. *Cell* 120:635–647.

Johnston RJ, Chang S, Etchberger JF, Ortiz CO, Hobert O (2005) Micrornas acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc Natl Acad Sci USA* 102:12449–54.

Jones-Rhoades MW, Bartel DP, Bartel B (2006) Micrornas and their regulatory roles in plants. *Annu Rev Plant Biol* 57:19–53.

Kadener S, Menet JS, Sugino K, Horwich MD, Weissbein U, Nawathean P, Vagin VV, Zamore PD, Nelson SB, Rosbash M (2009a) A role for micrornas in the drosophila circadian clock. *Genes & Development* 23:2179–91.

Kadener S, Rodriguez J, Abruzzi KC, Khodor YL, Sugino K, Marr MT, Nelson S, Rosbash M (2009b) Genome-wide identification of targets of the drosha-pasha/dgcr8 complex. *RNA* 15:537–45.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) Rna maps reveal new rna classes and a possible function for pervasive transcription. *Science* 316:1484–8.

Kapsimali M, Kloosterman WP, Bruijn ED, Rosa F, Plasterk RHA, Wilson SW (2007) Micrornas show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol* 8:R173.

Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G, Hannon GJ (2007) A biochemical approach to identifying microrna targets. *Proc Natl Acad Sci U S A* 104:19291–19296.

Karres JS, Hilgers V, Carrera I, Treisman J, Cohen SM (2007) The conserved microrna mir-8 tunes atrophin levels to prevent neurodegeneration in drosophila. *Cell* 131:136–45.

Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, Hatzigeorgiou AG, Nishikura K (2008) Frequency and fate of microrna editing in human brain. *Nucleic Acids Research* 36:5270–80.

Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K (2007) Redirection of silencing targets by adenosine-to-inosine editing of mirnas. *Science* 315:1137–40.

Kedde M, Strasser MJ, Boldajipour B, Vrielink JAFO, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, ¯rom UA, Lund AH, Perrakis A, Raz E, Agami R (2007) Rna-binding protein dnd1 inhibits microrna access to target mrna. *Cell* 131:1273–1286.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microrna target recognition. *Nat Genet* 39:1278–1284.

Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH (2001) Dicer functions in rna interference and in synthesis of small rna involved in developmental timing in c. elegans. *Genes Dev* 15:2654–2659.

Khvorova A, Reynolds A, Jayasena SD (2003) Functional sirnas and mirnas exhibit strand bias. *Cell* 115:209–216.

Kim VN, Han J, Siomi MC (2009) Biogenesis of small rnas in animals. *Nat Rev Mol Cell Biol* 10:126–139.

Kim VN, Nam JW (2006) Genomics of microrna. *Trends Genet* 22:165–73.

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JGI, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D (2008) The genome of the choanoflagellate monosiga brevicollis and the origin of metazoans. *Nature* 451:783–788.

Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A (2004) A combined computational-experimental approach predicts human microrna targets. *Genes Dev* 18:1165–1178.

Kloosterman WP (2006) Cloning and expression of new micrornas from zebrafish. *Nucleic Acids Research* 34:2558–2569.

Kloosterman WP, Wienholds E, Ketting RF, Plasterk RHA (2004) Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* 32:6284–6291.

Kocerha J, Faghihi MA, Lopez-Toledano MA, Huang J, Ramsey AJ, Caron MG, Sales N, Willoughby D, Elmen J, Hansen HF, Orum H, Kauppinen S, Kenny PJ, Wahlestedt C (2009) Microrna-219 modulates nmda receptor-mediated neurobehavioral dysfunction. *Proc Natl Acad Sci USA* 106:3507–12.

Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N (2005) Combinatorial microrna target predictions. *Nat Genet* 37:495–500.

Kulakova M, Bakalenko N, Novikova E, Cook CE, Eliseeva E, Steinmetz PRH, Kostyuchenko RP, Dondua A, Arendt D, Akam M, Andreeva T (2007) Hox gene expression in larval development of the polychaetes nereis virens and platynereis dumerilii (annelida, lophotrochozoa). *Dev Genes Evol* 217:39–54.

Lacalli T (2003) Evolutionary biology: Body plans and simple brains. *Nature* 424:263–264.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed rnas. *Science* 294:853–858.

Lai EC (2002) Micro rnas are complementary to 3' utr sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* 30:363–364.

Lai EC, Tomancak P, Williams RW, Rubin GM (2003) Computational identification of drosophila microrna genes. *Genome Biol* 4:R42.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foà R, Schliwka J, Fuchs U, Novosel A, Müller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, Vita GD, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Lauro RD, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T (2007) A mammalian microrna expression atlas based on small rna library sequencing. *Cell* 129:1401–14.

Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF (2009) Evidence for human microrna-offset rnas in small rna sequencing data. *Bioinformatics* 25:2298–301.

Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny rnas with probable regulatory roles in caenorhabditis elegans. *Science* 294:858–862.

Lee J, Li Z, Brower-Sinning R, John B (2007) Regulatory circuit of human microrna biogenesis. *PLoS Comput Biol* 3:e67.

Lee RC, Ambros V (2001) An extensive class of small rnas in caenorhabditis elegans. *Science* 294:862–864.

Lee RC, Feinbaum RL, Ambros V (1993) The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *Cell* 75:843–54.

Lee YS, Dutta A (2009) Microrna s in cancer. *Annual review of pathology* 4:199–227.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, Kim VN (2003) The nuclear rnase iii drosha initiates microrna processing. *Nature* 425:415–9.

Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN (2006) The role of pact in the rna silencing pathway. *EMBO J* 25:522–32.

Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) Microrna maturation: stepwise processing and subcellular localization. *EMBO J* 21:4663–70.

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) Microrna genes are transcribed by rna polymerase ii. *EMBO J* 23:4051–60.

Legendre M, Lambert A, Gautheret D (2005) Profile-based detection of microrna precursors in animal genomes. *Bioinformatics* 21:841–5.

Lehrbach N, Armisen J, Lightfoot H, Murfitt K, Bugaut A, Balasubramanian S, Miska E (2009) Lin-28 and the poly(u) polymerase pup-2 regulate let-7 microrna processing in caenorhabditis elegans. *Nat Struct Mol Biol* .

Leiss D, Hinz U, Gasch A, Mertz R, Renkawitz-Pohl R (1988) Beta 3 tubulin expression characterizes the differentiating mesodermal germ layer during drosophila embryogenesis. *Development* 104:525–31.

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. *Cell* 120:15–20.

Lewis BP, hung Shih I, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microrna targets. *Cell* 115:787–98.

Li X, Carthew RW (2005) A microrna mediates egf receptor signaling and promotes photoreceptor differentiation in the drosophila eye. *Cell* 123:1267–1277.

Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW (2009) A microrna imparts robustness against environmental fluctuation during development. *Cell* 137:273–282.

Li Y, Wang F, Lee JA, Gao FB (2006) Microrna-9a ensures the precise specification of sensory organ precursors in drosophila. *Genes & Development* 20:2793–2805.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some micrornas downregulate large numbers of target mrnas. *Nature* 433:769–773.

Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP (2003) The micrornas of caenorhabditis elegans. *Genes Dev* 17:991–1008.

Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the drosophila argonaute 2 paz domain. *Nature* 426:465–469.

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian rnai. *Science* 305:1437–41.

Liu J, Rivas FV, Wohlschlegel J, Yates JR, Parker R, Hannon GJ (2005a) A role for the p-body component gw182 in microrna function. *Nat Cell Biol* 7:1261–1266.

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005b) Microrna-dependent localization of targeted mrnas to mammalian p-bodies. *Nat Cell Biol* 7:719–723.

Liu N, Okamura K, Tyler DM, Phillips MD, Chung WJ, Lai EC (2008) The evolution and functional diversification of animal microrna genes. *Cell Res* 18:985–996.

Love GD, Grosjean E, Stalvies C, Fike DA, Grotzinger JP, Bradley AS, Kelly AE, Bhatia M, Meredith W, Snape CE, Bowring SA, Condon DJ, Summons RE (2009) Fossil steroids record the appearance of demospongiae during the cryogenian period. *Nature* 457:718–721.

Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ (2005) Elucidation of the small rna component of the transcriptome. *Science* 309:1567–9.

Lu J, Fu Y, Kumar S, Shen Y, Zeng K, Xu A, Carthew R, Wu CI (2008a) Adaptive evolution of newly emerged micro-rna genes in drosophila. *Mol Biol Evol* 25:929–938.

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI (2008b) The birth and death of microrna genes in drosophila. *Nat Genet* 40:351–355.

Lu YC, Smielewska M, Palakodeti D, Lovci MT, Aigner S, Yeo GW, Graveley BR (2009) Deep sequencing identifies new and regulated microrna in schmidtea mediterranea. *RNA* pp. 1–10.

Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microrna precursors. *Science* 303:95–8.

Luo Y, Zhang S (2009) Computational prediction of amphioxus microrna genes and their targets. *Gene* 428:41–6.

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.

Ma JB, Ye K, Patel DJ (2004) Structural basis for overhang-specific small interfering rna recognition by the paz domain. *Nature* 429:318–322.

Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5'-end-specific recognition of guide rna by the a. fulgidus piwi protein. *Nature* 434:666–70.

Makeyev EV, Zhang J, Carrasco MA, Maniatis T (2007) The microrna mir-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mrna splicing. *Mol Cell* 27:435–48.

Maniataki E, Mourelatos Z (2005) A human, atp-independent, risc assembly machine fueled by pre-mirna. *Genes Dev* 19:2979–2990.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–80.

Martindale MQ (2004) Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone nematostella vectensis (phylum, cnidaria; class, anthozoa). *Development* 131:2463–2474.

Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJM (2008) Genome-scale spatiotemporal analysis of caenorhabditis elegans microrna promoter activity. *Genome Research* 18:2005–15.

Mathonnet G, Fabian MR, Svitkin YV, Parsyan A, Huck L, Murata T, Biffo S, Merrick WC, Darzynkiewicz E, Pillai RS, Filipowicz W, Duchaine TF, Sonenberg N (2007) Microrna inhibition of translation initiation in vitro by targeting the cap-binding complex eif4f. *Science* 317:1764–1767.

Mayr C, Bartel DP (2009) Widespread shortening of 3'utrs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138:673–684.

Mcdougall C, Chen WC, Shimeld SM, Ferrier DEK (2006) The development of the larval nervous system, musculature and ciliary bands of pomatoceros lamarckii (annelida): heterochrony in polychaetes. *Front Zool* 3:16.

Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG (2007) mirgen: a database for the study of animal microrna genomic organization and function. *Nucleic Acids Research* 35:D149–D155.

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T (2004) Human argonaute2 mediates rna cleavage targeted by mirnas and sirnas. *Mol Cell* 15:185–97.

Miska EA, Alvarez-Saavedra E, Townsend M, Yoshii A, Sestan N, Rakic P, Constantine-Paton M, Horvitz HR (2004) Microarray analysis of microrna expression in the developing mammalian brain. *Genome Biol* 5:R68.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA (2008) Application of massively parallel sequencing to microrna profiling and discovery in human embryonic stem cells. *Genome Research* 18:610–21.

Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ (2008) Primary microrna transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* 15:902–9.

Moss EG, Lee RC, Ambros V (1997) The cold shock domain protein lin-28 controls developmental timing in c. elegans and is regulated by the lin-4 rna. *Cell* 88:637–646.

Mourelatos Z (2002) mirnps: a novel class of ribonucleoproteins containing numerous micrornas. *Genes & Development* 16:720–728.

Muhlrad D, Decker CJ, Parker R (1995) Turnover mechanisms of the stable yeast pgk1 mrna. *Mol Cell Biol* 15:2145–2156.

Murakami Y, Ogasawara M, Sugahara F, Hirano S, Satoh N, Kuratani S (2001) Identification and expression of the lamprey pax6 gene: evolutionary origin of the segmented brain of vertebrates. *Development* 128:3521–31.

Nakajima Y (1986) Development of the nervous system of sea urchin embryos: Formation of ciliary bands and the appearance of 2 types of ectoneural cells in the pluteus. *Dev Growth Differ* 28:531–542.

Newman MA, Thomson JM, Hammond SM (2008) Lin-28 interaction with the let-7 precursor loop mediates regulated microrna processing. *RNA* 14:1539–49.

Nielsen C (2001) *Animal Evolution. Interrelationships of the Living Phyla., 2nd ed.* Oxford University press, Oxford, 2nd edition.

Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB (2007) Determinants of targeting by endogenous and exogenous microrna and sirnas. *RNA* 13:1894–910.

Nolo R, Abbott LA, Bellen HJ (2000) Senseless, a zn finger transcription factor, is necessary and sufficient for sensory organ development in drosophila. *Cell* 102:349–362.

Nomura T, Kimura M, Horii T, Morita S, Soejima H, Kudo S, Hatada I (2008) Mecp2-dependent repression of an imprinted mir-184 released by depolarization. *Human Molecular Genetics* 17:1192–1199.

Nottrott S, Simard MJ, Richter JD (2006) Human let-7a mirna blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* 13:1108–1114.

Obernosterer G (2006) Post-transcriptional regulation of microrna expression. *RNA* 12:1161–1167.

O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005) c-myc-regulated microrna modulate e2f1 expression. *Nature* 435:839–843.

Ohler U, Yekta S, Lim LP, Bartel DP, Burge CB (2004) Patterns of flanking sequence conservation and a characteristic upstream motif for microrna gene identification. *RNA* 10:1309–22.

Ohman M (2007) A-to-i editing challenger or ally to the microrna process. *Biochimie* 89:1171–1176.

Ohno S (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. *Semin Cell Dev Biol* 10:517–522.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC (2007) The mirtron pathway generates microrna-class regulatory rnas in drosophila. *Cell* 130:89–100.

Okamura K, Ishizuka A, Siomi H, Siomi MC (2004) Distinct roles for argonaute proteins in small rna-directed rna cleavage pathways. *Genes & Development* 18:1655–66.

Okamura K, Phillips MD, Tyler DM, Duan H, ting Chou Y, Lai EC (2008) The regulatory activity of microrna* species has substantial influence on microrna and 3' utr evolution. *Nat Struct Mol Biol* 15:354–63.

Olsen PH, Ambros V (1999) The lin-4 regulatory rna controls developmental timing in caenorhabditis elegans by blocking lin-14 protein synthesis after the initiation of translation. *Developmental Biology* 216:671–80.

Palakodeti D (2006) Micrornas from the planarian schmidtea mediterranea: A model system for stem cell biology. *RNA* 12:1640–1649.

Pall GS, Hamilton AJ (2008) Improved northern blot method for enhanced detection of small rna. *Nat Protoc* 3:1077–1084.

Parker JS, Roe SM, Barford D (2005) Structural insights into mrna recognition from a piwi domain-sirna guide complex. *Nature* 434:663–6.

Parker R, Song H (2004) The enzymes and control of eukaryotic mrna turnover. *Nat Struct Mol Biol* 11:121–127.

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory rna. *Nature* 408:86–9.

Pawlicki JM, Steitz JA (2008) Primary microrna transcript retention at sites of transcription leads to enhanced microrna production. *The Journal of Cell Biology* 182:61–76.

Pena JTG, Sohn-Lee C, Rouhanifard SH, Ludwig J, Hafner M, Mihailovic A, Lim C, Holoch D, Berninger P, Zavolan M, Tuschl T (2009) mirna in situ hybridization in formaldehyde and edc-fixed tissues. *Nat Methods* 6:139–141.

Petersen CP, Bordeleau ME, Pelletier J, Sharp PA (2006) Short rnas repress translation after initiation in mammalian cells. *Mol Cell* 21:533–542.

Peterson KJ, Dietrich MR, Mcpeek MA (2009) Micrornas and metazoan macroevolution: insights into canalization, complexity, and the cambrian explosion. *Bioessays* 31:736–47.

Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grässer FA, van Dyk LF, Ho CK, Shuman S, Chien M, Russo JJ, Ju J, Randall G, Lindenbach BD, Rice CM, Simon V, Ho DD, Zavolan M, Tuschl T (2005) Identification of micrornas of the herpesvirus family. *Nat Methods* 2:269–76.

Philippe H, Lartillot N, Brinkmann H (2005) Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. *Mol Biol Evol* 22:1246–1253.

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W (2005) Inhibition of translational initiation by let-7 microrna in human cells. *Science* 309:1573–1576.

Piriyapongsa J, Marino-Ramirez L, Jordan IK (2006) Origin and evolution of human micrornas from transposable elements. *Genetics* 176:1323–1337.

Piriyapongsa J, Jordan IK (2007) A family of human microrna genes from miniature inverted-repeat transposable elements. *PLoS One* 2:e203.

Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P, Stoffel M (2004) A pancreatic islet-specific microrna regulates insulin secretion. *Nature* 432:226–230.

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* 3:827–837.

Prochnik SE, Rokhsar DS, Aboobaker AA (2007) Evidence for a microrna expansion in the bilaterian ancestor. *Dev Genes Evol* 217:73–77.

Prud'homme B, de Rosa R, Arendt D, Julien JF, Pajaziti R, Dorresteijn AWC, Adoutte A, Wittbrodt J, Balavoine G (2003) Arthropod-like expression patterns of engrailed and wingless in the annelid platynereis dumerilii suggest a role in segment formation. *Curr Biol* 13:1876–1881.

Purschke G (2005) Sense organs in polychaetes (annelida). *HYDROBIOLOGIA* 535:53–78 Symposium on Morphology, Molecules, Evolution and Phylogeny in Polychaeta and Related Taxa, Osnabruck, GERMANY, SEP, 2002.

Purschke G, Wolfrath F, Westheide W (1997) Ultrastructure of the nuchal organ and cerebral organ in onchnesoma squamatum (sipuncula, phascolionidae). *Zoomorphology* 117:23–31.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.

Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balavoine G, Ferrier D, Benes V, de Jong P, Weissenbach J, Bork P, Arendt D (2005) Vertebrate-type intron-rich genes in the marine annelid platynereis dumerilii. *Science* 310:1325–1326.

Rao PK, Kumar RM, Farkhondeh M, Baskerville S, Lodish HF (2006) Myogenic factors that regulate expression of muscle-specific micrornas. *Proc Natl Acad Sci USA* 103:8721–6.

Raser JM (2005) Noise in gene expression: Origins, consequences, and control. *Science* 309:2010–2013.

Rastogi S, Liberles DA (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol* 5:28.

Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E (2005) A crucial role for gw182 and the dcp1:dcp2 decapping complex in mirna-mediated gene silencing. *RNA* 11:1640–1647.

Rehwinkel J, Natalin P, Stark A, Brennecke J, Cohen SM, Izaurralde E (2006) Genome-wide analysis of mrnas regulated by drosha and argonaute proteins in drosophila melanogaster. *Mol Cell Biol* 26:2965–75.

Reichert H, Simeone A (2001) Developmental genetic evidence for a monophyletic origin of the bilaterian brain. *Philos Trans R Soc Lond B Biol Sci* 356:1533–1544.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 rna regulates developmental timing in caenorhabditis elegans. *Nature* 403:901–6.

Rentzsch F, Anton R, Saina M, Hammerschmidt M, Holstein TW, Technau U (2006) Asymmetric expression of the bmp antagonists chordin and gremlin in the sea anemone nematostella vectensis: implications for the evolution of axial patterning. *Dev Biol* 296:375–387.

Ro S, Park C, Young D, Sanders KM, Yan W (2007) Tissue-dependent paired expression of mirnas. *Nucleic Acids Research* 35:5944–5953.

Robertis EMD (2008) Evo-devo: variations on ancestral themes. *Cell* 132:185–95.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) Identification of mammalian microrna host genes and transcription units. *Genome Research* 14:1902–10.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of drosophila micrornas. *Genome Research* 17:1850–1864.

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP (2006) Large-scale sequencing reveals 21u-rnas and additional micrornas and endogenous sirnas in c. elegans. *Cell* 127:1193–207.

Ruby JG, Jan CH, Bartel DP (2007) Intronic microrna precursors that bypass drosha processing. *Nature* 448:83–6.

Rusten TE, Cantera R, Kafatos FC, Barrio R (2002) The role of tgf beta signaling in the formation of the dorsal nervous system is conserved between drosophila and chordates. *Development* 129:3575–3584.

Ruvkun G, Ambros V, Coulson A, Waterston R, Sulston J, Horvitz HR (1989) Molecular genetics of the caenorhabditis elegans heterochronic gene lin-14. *Genetics* 121:501–516.

Ruvkun G, Giusto J (1989) The caenorhabditis elegans heterochronic gene lin-14 encodes a nuclear protein that forms a temporal developmental switch. *Nature* 338:313–319.

Ryan DG, Oliveira-Fernandes M, Lavker RM (2006) Micrornas of the mammalian eye display distinct and overlapping tissue specificity. *Mol Vis* 12:1175–84.

Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, Wulczyn FG (2008) A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nature* 10:987–93.

Sacheli R, Nguyen L, Borgs L, Vandenbosch R, Bodson M, Lefebvre P, Malgrange B (2009) Expression patterns of mir-96, mir-182 and mir-183 in the development inner ear. *Gene Expression Patterns* 9:364–70.

Saini HK, Enright AJ, Griffiths-Jones S (2008) Annotation of mammalian primary micrornas. *BMC Genomics* 9:564.

Saito K, Ishizuka A, Siomi H, Siomi MC (2005) Processing of pre-micrornas by the dicer-1-loquacious complex in drosophila cells. *PLoS Biol* 3:e235.

Sandmann T, Cohen SM, Gonzalez C (2007) Identification of novel drosophila melanogaster micrornas. *PLoS ONE* 2:e1265.

Schlotzer-Schrehardt U (1986) Ultrastructural investigation of the nuchal organs of pygospio elegans (polychaeta). i. larval nuchal organs. *Helgoland Marine Research* 40:397–417.

Scholz CB, Technau U (2003) The ancestral role of brachyury: expression of nembra1 in the basal cnidarian nematostella vectensis (anthozoa). *Dev Genes Evol* 212:563–70.

Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME (2006) A brain-specific microrna regulates dendritic spine development. *Nature* 439:283–9.

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD (2003) Asymmetry in the assembly of the rnai enzyme complex. *Cell* 115:199–208.

Seidah NG, Chrétien M (1999) Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res* 848:45–62.

Seitz H, Ghildiyal M, Zamore PD (2008) Argonaute loading improves the 5' precision of both micrornas and their mirna strands in flies. *Curr Biol* 18:147–51.

Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by micrornas. *Nature* 455:58–63.

Sempere LF, Cole CN, Mcpeek MA, Peterson KJ (2006) The phylogenetic distribution of metazoan micrornas: insights into evolutionary complexity and constraint. *J. Exp. Zool.* 306B:575–588.

Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V (2004) Expression profiling of mammalian micrornas uncovers a subset of brain-expressed

microrнas with possible roles in murine and human neuronal differentiation. *Genome Biol* 5:R13.

Sen GL, Blau HM (2005) Argonaute 2/risc resides in sites of mammalian mrna decay known as cytoplasmic bodies. *Nat Cell Biol* 7:633–636.

Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic rna interference. *Trends Ecol Evol (Amst)* 23:578–87.

Shi W, Hendrix D, Levine M, Haley B (2009) A distinct class of small rnas arises from pre-mirna–proximal regions in a simple chordate. *Nat Struct Mol Biol* 16:183–189.

Shkumatava A, Stark A, Sive H, Bartel DP (2009) Coherent but overlapping expression of microrнas and their targets during vertebrate development. *Genes & Development* 23:466–481.

Shomron N, Golan D, Hornstein E (2009) An evolutionary perspective of animal microrнas and their targets. *J. Biomed. Biotech.* 2009:594738.

Smalheiser NR, Torvik VI (2005) Mammalian microrнas derived from genomic repeats. *Trends Genet* 21:322–326.

Sokol NS, Xu P, Jan YN, Ambros V (2008) Drosophila let-7 microrna is required for remodeling of the neuromusculature during metamorphosis. *Genes & Development* 22:1591–1596.

Sokol NS, Ambros V (2005) Mesodermally expressed drosophila microrna-1 is regulated by twist and is required in muscles during larval growth. *Genes & Development* 19:2343–54.

Sood P, Krek A, Zavolan M, Macino G, Rajewsky N (2006) Cell-type-specific signatures of microrнas on target mrna expression. *Proc Natl Acad Sci USA* 103:2746–51.

Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M (2007) Systematic discovery and characterization of fly microrнas using 12 drosophila genomes. *Genome Research* 17:1865–1879.

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM (2005) Animal microrнas confer robustness to gene expression and have a significant impact on 3'utr evolution. *Cell* 123:1133–46.

Steiner FA, Hoogstrate SW, Okihara KL, Thijssen KL, Ketting RF, Plasterk RHA, Sijen T (2007) Structural features of small rna precursors determine argonaute loading in caenorhabditis elegans. *Nat Struct Mol Biol* 14:927–33.

Strathmann RR (1971) The feeding behavior of planktotrophic echinoderm larvae: Mechanisms, regulation, and rates of suspension feeding. *J. Exp. Mar. Biol. Ecol.* 6:109–160.

Sullivan JC, Ryan JF, Watson JA, Webb J, Mullikin JC, Rokhsar D, Finnerty JR (2006) Stellabase: the nematostella vectensis genomics database. *Nucleic Acids Res* 34:D495–D499.

Sun Y, Zhao YN, Wang DY (2006) Computational analysis of genetic loci required for synapse structure and function and their corresponding micrornas in c. elegans. *Neuroscience bulletin* 22:339–49.

Szafranska AE, Davison TS, John J, Cannon T, Sipos B, Maghnouj A, Labourier E, Hahn SA (2007) Microrna expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. *Oncogene* 26:4442–4452.

Tang F, Kaneda M, O'carroll D, Hajkova P, Barton SC, Sun YA, Lee C, Tarakhovsky A, Lao K, Surani MA (2007a) Maternal micrornas are essential for mouse zygotic development. *Genes & Development* 21:644–648.

Tang X, Gal J, Zhuang X, Wang W, Zhu H, Tang G (2007b) A simple array platform for microrna analysis and its application in mouse tissues. *RNA* 13:1803–1822.

Tanzer A, Stadler PF (2004) Molecular evolution of a microrna cluster. *J Mol Biol* 339:327–35.

Technau U (2008) Evolutionary biology: Small regulatory rnas pitch in. *Nature* 455:1184–5.

Teleman AA, S M, SM. C (2006) Drosophila lacking microrna mir-278 are defective in energy homeostasis. *Genes & Development* 20:417–422.

Tessmar-Raible K, Arendt D (2003) Emerging systems: between vertebrates and arthropods, the lophotrochozoa. *Curr Opin Genet Dev* 13:331–340.

Tessmar-Raible K, Raible F, Christodoulou F, Guy K, Rembold M, Hausen H, Arendt D (2007) Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell* 129:1389–400.

Thai TH, Calado DP, Casola S, Ansel KM, Xiao C, Xue Y, Murphy A, Frendewey D, Valenzuela D, Kutok JL, Schmidt-Supprian M, Rajewsky N, Yancopoulos G, Rao A, Rajewsky K (2007) Regulation of the germinal center response by microrna-155. *Science* 316:604–8.

Thatcher EJ, Bond J, Paydar I, Patton JG (2008) Genomic organization of zebrafish microrna s. *BMC Genomics* 9:253.

Thermann R, Hentze MW (2007) Drosophila mir2 induces pseudo-polysomes and inhibits translation initiation. *Nature* 447:875–878.

Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM (2006) Extensive post-transcriptional regulation of microrna s and its implications for cancer. *Genes & Development* 20:2202–7.

Till S, Lejeune E, Thermann R, Bortfeld M, Hothorn M, Enderle D, Heinrich C, Hentze MW, Ladurner AG (2007) A conserved motif in argonaute-interacting proteins mediates functional interactions through the argonaute piwi domain. *Nat Struct Mol Biol* 14:897–903.

Tomari Y, Du T, Zamore PD (2007) Sorting of drosophila small silencing rna s. *Cell* 130:299–308.

Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD (2004) A protein sensor for sirna asymmetry. *Science* 306:1377–80.

Tourrière H, Chebli K, Tazi J (2002) mrna degradation machines in eukaryotic cells. *Biochimie* 84:821–37.

Tzetlin AB, Purschke G (2006) Fine structure of the pharyngeal apparatus of the pelagosphera larva in phascolosoma agassizii (sipuncula) and its phylogenetic significance. *Zoomorphology* 125:109–117.

Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO (2007) Exosome-mediated transfer of mrna s and microrna s is a novel mechanism of genetic exchange between cells. *Nature* 9:654–659.

Valoczi A, Hornyik C, Varga N, Burgyan J, Kauppinen S, Havelda Z (2004) Sensitive and specific detection of microrna s by northern blot analysis using lna-modified oligonucleotide probes. *Nucleic Acids Res* 32:e175.

van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN (2007) Control of stress-dependent cardiac growth and gene expression by a microrna. *Science* 316:575–579.

Varga ZM, Wegner J, Westerfield M (1999) Anterior movement of ventral diencephalic precursors separates the primordial eye field in the neural plate and requires cyclops. *Development* 126:5533–46.

Velasco BD, Shen J, Go S, Hartenstein V (2004) Embryonic development of the drosophila corpus cardiacum, a neuroendocrine gland with similarity to the vertebrate pituitary, is controlled by sine oculis and glass. *Developmental Biology* 274:280–294.

Vigh B, e Silva MJM, Frank CL, Vincze C, Czirok SJ, Szabo A, Lukats A, Szel A (2004) The system of cerebrospinal fluid-contacting neurons. its supposed role in the nonsynaptic signal transmission of the brain. *Histol Histopathol* 19:607–628.

Vigh B, Vigh-Teichmann I, Röhlich P, Oksche A (1983) Cerebrospinal fluid-contacting neurons, sensory pinealocytes and landolt's clubs of the retina as revealed by means of an electron-microscopic immunoreaction against opsin. *Cell Tissue Res* 233:539–48.

Visvanathan J, Lee S, Lee B, Lee JW, Lee SK (2007) The microrna mir-124 antagonizes the anti-neural rest/scp1 pathway during embryonic cns development. *Genes & Development* 21:744–749.

Viswanathan SR, Daley GQ, Gregory RI (2008) Selective blockade of microrna processing by lin28. *Science* 320:97–100.

Wada Y, Mogami Y, Baba S (1997) Modification of ciliary beating in sea urchin larvae induced by neurotransmitters: beat-plane rotation and control of frequency fluctuation. *J Exp Biol* 200:9–18.

Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150:563–565.

Wang B, Yanez A, Novina CD (2008) Microrna-repressed mrnas contain 40s but not 60s components. *Proc Natl Acad Sci USA* 105:5343–8.

Wang D, Lu M, Miao J, Li T, Wang E, Cui Q, Khanin R (2009) Cepred: Predicting the co-expression patterns of the human intronic microrna with their host genes. *PLoS ONE* 4:e4421.

Wang Z, Kiledjian M (2001) Functional link between the mammalian exosome and mrna decapping. *Cell* 107:751–762.

Wanninger A (2009) Shaping the things to come: ontogeny of lophotrochozoan neuromuscular systems and the tetraneuralia concept. *Biol Bull* 216:293–306.

Webster M (2007) A cambrian peak in morphological variation within trilobite species. *Science* 317:499–502.

Weistheide W, Rieger R (1996) *Spezielle Zoologie Erster Teil: Einzeller und Wirbellose Tiere.* Gustav Fischer Verlag.

Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ (2009) The deep evolution of metazoan micrornas. *Evol Dev* 11:50–68.

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, Bruijn ED, Horvitz HR, Kauppinen S, Plasterk RHA (2005) Microrna expression in zebrafish embryonic development. *Science* 309:310–1.

Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in c. elegans. *Cell* 75:855–62.

Winter F, Edaye S, Huttenhofer A, Brunel C (2007) Anopheles gambiae mirnas as actors of defence reaction against plasmodium invasion. *Nucleic Acids Research* 35:6953–6962.

Winter J, Jung S, Keller S, Gregory RI, Diederichs S (2009) Many roads to maturity: microrna biogenesis pathways and their regulation. *Nature* 11:228–34.

Xiong H, Qian J, He T, Li F (2009) Independent transcription of mir-281 in the intron of oda in drosophila melanogaster. *Biochemical and Biophysical Research Communications* 378:883–9.

Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D (2007) Microrna (mirna) transcriptome of mouse retina and identification of a sensory organ-specific mirna cluster. *J Biol Chem* 282:25053–66.

Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K (2006) Modulation of microrna processing and expression through rna editing by adar deaminases. *Nat Struct Mol Biol* 13:13–21.

Yekta S (2004) Microrna-directed cleavage of hoxb8 mrna. *Science* 304:594–596.

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CCG, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC (2006) Analysis of the c. elegans argonaute family reveals that distinct argonautes act sequentially during rnai. *Cell* 127:747–57.

ying Chu C, Rana TM (2006) Translation repression in human cells by microrna-induced gene silencing requires rck/p54. *PLoS Biol* 4:e210.

Yu Z, Raabe T, Hecht NB (2005) Microrna mirn122a reduces expression of the posttranscriptionally regulated germ cell transition protein 2 (tnp2) messenger rna (mrna) by mrna cleavage. *Biol Reprod* 73:427–33.

Yuan A, Farber EL, Rapoport AL, Tejada D, Deniskin R, Akhmedov NB, Farber DB (2009) Transfer of micrornas by embryonic stem cell microvesicles. *PLoS ONE* 4:e4722.

Zdanowicz A, Thermann R, Kowalska J, Jemielity J, Duncan K, Preiss T, Darzynkiewicz E, Hentze MW (2009) Drosophila mir2 primarily targets the m7gpppn cap structure for translational repression. *Mol Cell* 35:881–8.

Zeng Y, Cullen BR (2004) Structural requirements for pre-microrna binding and nuclear export by exportin 5. *Nucleic Acids Research* 32:4776–85.

Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W (2004) Single processing center models for human dicer and bacterial rnase iii. *Cell* 118:57–68.

Zhang R, Peng Y, Wang W, Su B (2007) Rapid evolution of an x-linked microrna cluster in primates. *Genome Research* 17:612–617.

Zhang S, Xu L, Lee J, Xu T (2002) Drosophila atrophin homolog functions as a transcriptional corepressor in multiple developmental processes. *Cell* 108:45–56.

Zhang Y, Zhang R, Su B (2009) Diversity and evolution of microrna gene clusters. *Sci China, C, Life Sci* 52:261–6.

Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, Muth AN, Tsuchihashi T, Mcmanus MT, Schwartz RJ, Srivastava D (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking mirna-1-2. *Cell* 129:303–17.

APPENDIX

Platynereis miRNA alignments.

a, Alignments of *Platynereis* miRNA sequences with other animal microRNAs. If more than one *Platynereis* sequence variant exists, the most abundant one is displayed. Numbers in brackets refer to absolute number of occurrences and the respective frequency in the analysed dataset. Bold print indicates differences from the *Platynereis* sequence.

b, Taxonomic abbreviations: **Aqu**– *Amphimedon queenslandica*; **Bfl** – *Branchiostoma floridae*; **Cel** – *Caenorhabditis elegans*; **Csp**– *Capitella sp*.; **Dup** – *Daphnia pulex*; **Dme** – *Drosophila melanogaster*; **Dre** – *Danio rerio*; **Hma** – *Hydra magnipapillata*; **Hsa** – *Homo sapiens*; **Hsn** – *Henricia sanguinolenta*; **Isc** – *Ixodesscapularis*; **Lgi** – *Lottia gigantea*; **Nve** – *Nematostella vectensis*; **Sko** – *Saccoglossuskowalevskii*; **Spu** – *Strongylocentrotus purpuratus*. Alignment file adapted from Wheeler et al 2009.

Ancient bilaterian microRNAs alignements

microRNA Alignments (adapted file from Wheeler et al 2009)

miR-1

Pdu UGGAAUGUAAAGAAGUAUGUAG (57587 / 4.57%)
Dme UGGAAUGUAAAGAAGUAUG**G**AG
Dre UGGAAUGUAAAGAAGUAUGUA-
Hsa UGGAAUGUAAAGAAGUAUGUA-
Bfl UGGAAUGUAAAGAAGUAUGUA**U**
Sko UGGAAUGUAAUGAAGUAUGUA**U**
Spu UGGAAUGUAAAGAAGUAUGUA**U**
Hsn UGGAAUGUAAAGAAGUAUGUA**U**
Dpu UGGAAUGUAAAGAAGUAUG**G**AG
Isc UGGAAUGUAAAGAAGUAUG**G**AG
Pca UGGAAUGUAAAGAAGUAUG**G**AG
Hru UGGAAUGUAAAGAAGUAUGUA**U**
Lgi UGGAAUGUAAAGAAGUAUGUA**U**
Csp UGGAAUGUAAAGAAGUAUGUAG
Ndi UGGAAUGUAAAGAAGUAUGUAG
Cla UGGAAUGUAAAGAAGUAUGUAG
Sro UGGAAUGU**U**AAG**G**AGUA**CAAU**-

miR-7

Pdu UGGAAGACUAGUGAUUUUGUUGUU (5881 / 0.47%)
Dme UGGAAGACUAGUGAUUUUGUUGU-
Dre UGGAAGACUAGUGAUUUUGUUGU-
Hsa UGGAAGACUAGUGAUUUUGUUGU-
Bfl UGGAAGACUAGUGAUUUUGUUGU-
Sko UGGAAGACUAGUGAUUUUGUUGU-
Spu UGGAAGACUAGUGAUUUUGUUGU-
Hsn UGGAAGACUAGUGAUUUUGUUGU-
Dpu UGGAAGACUAGUGAUUUUGUUGU-
Isc UGGAAGACUAGUGAUUUUGUUGU-
Pca UGGAAGACUAGUGAUUUUGUUGU-
Hru UGGAAGACUAGUGAUUUUGUUGUU
Lgi UGGAAGACUAGUGAUUUUGUUGUU
Csp UGGAAGACUAGUGAUUUU**A**UUGUU
Ndi UGGAAGACUAGUGAUUUUGUUGUU
Cla UGGAAGACUAGUGAUUUUGUUGU-

miR-8, -141, -200

Pdu UAAUACUGUCAGGUAAAGAUGUU (111203 / 8.83%)
Dme UAAUACUGUCAGGUAAAGAUGU**C**
Dre UAA**C**ACUGUC**U**GGUAA**C**GAUGU- 200a
Hsa UAA**C**ACUGUC**U**GGUAA**C**GAUGU- 200a
Bfl **A**AAUACUG**C**C**U**GGUAA**U**GAUG**C**-
Sko UAAUACUGUC**U**GGUAA**U**GAUGUU
Spu UAAUACUGUC**U**GGU**G**A**U**GAUGUU
Hsn UAAUACUGUC**U**GGUAA**U**GAUGU-
Dpu UAAUACUGUCAGGUAAAGAUGU**C**
Isc UAAUACUGUCAGGUAAAGAUGU**C**
Pca UAAUACUGUCAGGUAAAGAUGU**C**
Hru UAAUACUGUCAGGUAAAGAUGU**C**
Lgi UAAUACUGUCAGGUAAAGAUGU**C**
Csp UAAUACUGUCAGGUAAAGAUGU**C**
Ndi UAAUACUGUCAGGUAAAGAUGUU
Cla UAAUACUGUCAGGUAAAGAUGU**C**

miR-9

Pdu UCUUUGGUUAUCUAGCUGUAUGA (943 / 0.07%)
Dme UCUUUGGUUAUCUAGCUGUAUGA
Dre UCUUUGGUUAUCUAGCUGUAUGA
Hsa UCUUUGGUUAUCUAGCUGUAUGA
Bfl UCUUUGGUUAUCUAGCUGUAUGA
Sko UCUUUGGUUAUCUAGCUGUAU--
Spu UCUUUGGUUAUCUAGCUGUAUG-
Hsn UCUUUGGUUAUCUAGCUGUAUGA
Dpu UCUUUGGUUAUCUAGCUGUAUGA
Isc UCUUUGGUUAUCUAGCUGUAUGA
Pca UCUUUGGUUAUCUAGCUGUAUG-
Hru UCUUUGGUUAUCUAGCUGUAUGA
Lgi UCUUUGGUUAUCUAGCUGUAUGA
Csp UCUUUGGUUAUCUAGCUGUAUGA
Ndi UCUUUGGUUAUCUAGCUGUAUGA
Cla UCUUUGGUUAUCUAGCUGUAUGA

miR-10

Pdu -UACCCUGUAGAUCCGAAUUUGU (2891 / 0.23%)
Dme --ACCCUGUAGAUCCGAAUUUGU--
Dre –UACCCUGUAGAUCCGAAUUUGU-- a
Hsa –UACCCUGUAGAUCCGAAUUUGU--
Bfl **G**UACCCU-GUAGAUCCGAAUUUGU**G**-
Sko –UACCCU-GUAGAUCCGAAUUUGU**G**-
Spu –**A**ACCCU-GUAGAUCCGAAUUUGU**G**-
Hsn –**A**ACCCU-GUAGAUCCGAAUUUGU--
Dpu –UACCCU-GUAGAUCCGAAUUUGU--
Isc -UACCCU-GUAGAUCCGAAUUUGU--
Pca -UACCCU-GUAG**U**UCCG**G**AUUUGU--
Hru -UACCCU-GUAGAUCCGAAUUUGU--
Lgi –UACCCU-GUAGAUCCGAAUUUGU--
Csp -UACCCU-GUAG**U**UCCG**G**AUUUGU-
Ndi -**C**ACCCU-GUAGA**A**CCGA**GC**UUGU--
Cla -UACCCU-GUAGAUCCGAAUUUGU--

miR-22, -745, -980

Pdu --AGCUGCCUGGUGAAGAGCUGUC (26280 / 2.09%)
Dme –**U**AGCUGCCU**U**GUGAAG**G**GCU**UA**-
Dre ––AGCUGCC**AG**U**U**GAAGAGCUGU-
Bfl -**A**AGCUGCC**AGA**UGAAGAGCUGU-
Spu **UC**AGCUGCC**C**GGUGAAG**U**G**UAU**--
Hsn **UC**AGCUGCC**C**GGUGAAG**U**G**UA**G--
Dpu -**G**AGCUGCC**CA**GUGAAGG**G**CU**U**U-
Isc -**C**AGCUGCCU**U**GUGAAG**G**GCU**UG**-
Pca –**G**AGCUGCC**CA**GUGAAG**G**GCU---
Hru --AGCUGCCUG**A**UGAAGAGCUGU-
Lgi --AGCUGCCUG**A**UGAAGAGCUGU-
Csp --AGCUGCCUGGUGAAGAGCUGUC
Ndi --AGCUGCCUGGUGAAGAGCUGUC
Cla --AGCUGCCU**G**GUGAAGAGCUGU**U**

miR-29, -83, -285

Pdu UAGCACCAUUUGAAAUCAGUUU (6429 / 0.51%)
Dme UAGCACCAUU**C**GAAAUCAGU**GC**-
Dre UAGCACCAUUUGAAAUCAGU**G**U
Hsa UAGCACCAUUUGAAAUCAGU**G**U
Bfl UAGCACCAU**A**UGAAAUCAGUU**AU**
Sko UAGCACCAU**A**UGAAAUCAGUUU-
Spu **A**AGCACCA**G**UUGAAAUCAG**AGC**-
Hsn **A**AGCACCA**G**UUGAAAUCAG**AGC**-
Dpu UAGCACCAUU**GG**AAUUCAGUUU-
Isc UAGCACCAUUUGAAUUCAGUU**C**-
Pca UAGCACCAUUUGAAAUCAGU**GC**-
Hru UAGCACCAUUUGAAAUCAGUUU-
Lgi UAGCACCAUUUGAAAUCAGUUU-
Csp UAGCACCAUUUGAAAUCAGUUU-
Ndi UAGCACCAUUUGAAAUCAGU**GC**-
Cla UAGCACCAUUUGAAAUCAGUU**C**-

miR-31

Pdu AGGCAAGAUGUUGGCAUAGCUGA (10177 / 0.81%)
Dme **U**GGCAAGAUGU**C**GGCAUAGCUGA
Dre -GGCAAGAUGUUGGCAUAGCUG-
Hsa AGGCAAGAUGCUGGCAUAGCU--
Bfl **U**GGCAAGAUGUUGGCAUAGCUG**U**
Sko AGGCAAGAUGUUGGCAUAGCUG-
Spu AGGCAAGAUGUUGGCAUAGCU--
Hsn AGGCAAGAUGUUGGCAUAGCU--
Dpu AGGCAAGAUGU**C**GGCAUAGCUGA
Pca AGGCAAGAU**UC**UGGCAUAGCUG-
Hru AGGCAAGAUGUUGGCAUAGCU--
Lgi AGGCAAGAUGUUGGCAUAGCU--
Csp AGGCAAGAUGUUGGCAUAGCU--
Ndi AGGCAAGAUGUUGGCAUAGCUGA
Cla AGGCAAGAUGUUGGCAUAGCUGA
Sro AGG-CAAGAUGUUGGCAUA**A**CU--

miR-33

Pdu --GUGCAUUGUAGUUGCAUUGCA (2 / 0%)
Dme AGGUGCAUUGUAGUCGCAUUG----
Hsa --GUGCAUUGUAGUUGCAUUGCA
Bfl --GUGCAUUGUAGUUGCAUUGCA**U**-
Spu --GUGCAUUGU**C**GUUGCAUUGCA**UU**
Hsn --GUGCAUUGUAGUUGCAUUGCA**U**
Dpu --GUGCAUUGUAGUUGCAUUGCA**C**
Isc --GUGCAUUG**C**AGUUGCAUUGCA**C**
Pca --GUGCAUUGUAGUUGCAUUGCA-
Hru --GUGCAUUGUAGUUGCAUUGC**CU**
Lgi --GUGCAUUGUAGUUGCAUUGC**CU**
Csp –-GUGCAUUGUAGUUGCAUUGCA-
Ndi --GUGCAUUGUAGUUGCAUUGCA**A**
Cla --GUGCAUUGUAGUUGCAUUGCA-

miR-34

Pdu UGGCAGUGUGGUUAGCUGGUUGU (2316 / 0.18%)
Dme UGGCAGUGUGGUUAGCUGGUUGUG-
Dre UGGCAGUGU**C**-UUAGCUGGUUGU
Has UGGCAGUGU**C**-UUAGCUGGUUGU
Bfl UGGCAGUGUGG**A**UAGCUGG**CC**GUUU
Sko UGGCAGUGUGGUUAGCUGGUUG---
Spu **C**GGCAGUGU**A**GUUAGCUGGUUG---
Hsn UGGCAGUGUGGUUAGCUGGUUG---
Dpu UGGCAGUGUGGUUAGCUGGUUGUG-
Hru UGGCAGUGUGGUUAGCUGGUUGU--
Lgi UGGCAGUGUGGUUAGCUGGU**A**GU--
Csp UGGCAGUGUGGUUAGCUGGUUGU--
Ndi UGGCAGUGUGGUUAGCUGGUUGU--
Cla UGGCAGUGUGGUUAGCUGGUUGU--

miR-71

Pdu UGAAAGACAUGGGUAGUGAGAUG (4251 / 0.34%)
Cel UGAAAGACAUGGGUAGUGA----
Bfl UGAAAGACAUGGGUAGUGAGAU-
Sko UGAAAGACA**CA**GGUAGUGAGAU-
Spu UGAAAGACAUGGGUAGUGAGAU**U**
Hsn UGAAAGACAUGGGUAGUGAGAU-
Dpu UGAAAGACAUGGGUAGUGAGAUG
Isc UGAAAGACAUGGGUAGUGAGAUG
Pca UGAAAGACAUGGGUAGUGAGAU-
Hru UGAAAGACAUGGGUAGUGAGAUG
Lgi UGAAAGACA**A**GGGUAGUGAGAUG
Csp UGAAAGACAUGGGUAGUGAGAUG
Ndi UGAAAGACAUGGGUAGUGAGAUG
Cla U-GAAAGACAUGGGUAGUGAGA--

miR-9*, -79

Pdu AUAAAGCUAGGUUACCAAAGCU- (366 / 0.03%)
Dme -UAAAGCUAG**A**UUACCAAAGC**A**U
Bfl AUAAAGCUAGGU**A**ACCAAA**AC**A**A**
Sko AUAAAGCUAGGUUACCAAAG**ACA**
Spu AUAAAGCUAGGUUACCAAAG**A**U-
Hsn –UAAAGCUAGGUUACCAAAG**A**U-
Dpu AUAAAGCUAGGUUACCAAAG**U**U**A**
Isc AUAAAGCUAGGUUACCAAAG**U**U**A**
Pca -UAAAGCUAGGUUACCAAAG**G**U**A**
Hru AUAAAGCUAGGUUACCAAAG**GC**-
Lgi AUAAAGCUAGGUUACCAAAG**GC**-
Csp AUAAAGCUAGGUUACCAAAGCU-
Ndi -UAAAGCUAGGUUACCAAAGCU**A**
Cla AUAAAGCUAGGUUACCAAAG**G**U-
Sro –UAAAGCUAG**AAA**ACCAAAG**A**U-

miR-92

Pdu AAUUGCACUUGUCCCGGCCUGC- (1605 / 0.13%)
Dme AAUUGCACU**A**GUCCCGGCCUGC-
Dre **U**AUUGCACUUGUCCCGGCCUG**U-** a
Hsa **U**AUUGCACUUGUCCCGGCCUG**U-** a
Bfl **U**AUUGCACUUGUCCCGGCCU**UU**-
Sko **U**AUUGCACUUGUCCCGGCCU**AA**-
Spu **U**AUUGCACUUGUCCCGGCCU**AC**U
Hsn **U**AUUGCACUUGUC**U**CGGCCUGC-
Dpu AAUUGCACU**C**GUCCCGGCCUGC-
Pca **U**AUUGCACU**C**GUCCCGGCCU**UU**-
Hru AAUUGCACU**C**GUCCCGGCCUGC-
Lgi AAUUGCACUUGUCCCGGCCUGC-
Csp **G**AUUGCACU**A**GUCCCGGCCU**U**C-
Ndi AAUUGCACU**C**GUCCCGGCCUGC-
Cla AAUUGCACU**C**GUCCCGGCCUGC-
Sro **U**AUUGCACUUG**G**C**U**CGGCCU**CA**-

miR-100

Pdu AACCCGUACAACCGAACUUGUG (24 / 0%)
Dme AACCCGUA**A**A**U**CCGAACUUGUG-
Dre AACCCGUA**G**A**U**CCGAACUUGUG
Hsa AACCCGUA**G**A**U**CCGAACUUGUG
Bfl AACCCGUA**G**A**U**CCGAACUUGUG**U**
Sko AACCCGUA**G**A**U**CCGAACUUGUG-
Hsn AACCCGUA**G**A**U**CCGAA**U**UUGU--
Dpu AACCCGUA**G**A**U**CCGAACUUGUG**U**
Isc AACCCGUA**G**A**U**CCGAACUUGUG**U**
Pca AACCCGUA**G**A**U**CCGAACUUGUG-
Hru AACCCGUA**G**A**U**CCGAACUUGUG-
Lgi AACCCGUA**G**A**U**CCGAACUUGUG-
Csp AACCCGUACAACCGAACUUGUG-
Ndi AACCCGUACAACCGAACUUGUG-
Cla AACCCGUA**G**A**U**CCGAACUUGUG-
Nve **G**ACCCGUA**G**A**U**CCGAACUUGUG-

miR-124

Pdu UAAGGCACGCGGUGAAUGCCA-- (7217 / 0.57%)
Dme UAAGGCACGCGGUGAAUGCCAAG
Dre UAAGGCACGCGGUGAAUGCCAA-
Hsa UAAGGCACGCGGUGAAUGCC---
Bfl UAAGGCACGCGGUGAAUGCCAA-
Sko UAAGGCACGCGGUGAAUGCCAA-
Spu UAAGGCACGCGGUGAAUGCCA--
Hsn UAAGGCACGCGGUGAAUGCCA--
Dpu UAAGGCACGCGGUGAAUGCCAAG
Hru UAAGGCACGCGGUGAAUGCCA--
Lgi UAAGGCACGCGGUGAAUGCCA--
Csp UAAGGCACGCGGUGAAUGCCA--
Ndi UAAGGCACGCGGUGAAUGCCA--
Sro UAAGGCACGCGGUGAAUGC**A**A--

miR-125

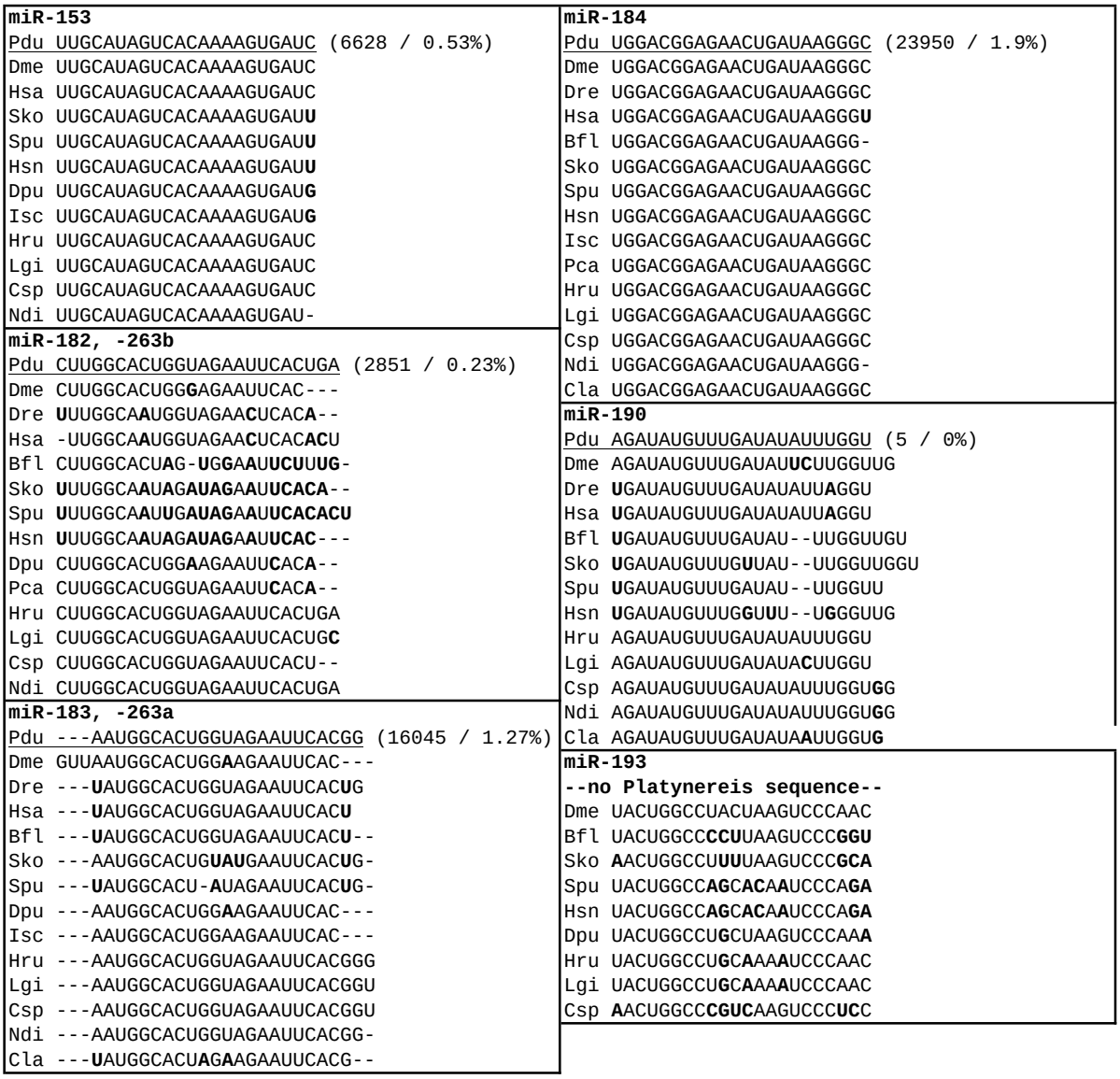
Pdu UCCCUGAGACCCUAACUUGUGA (384 / 0.03%)
Dme UCCCUGAGACCCUAACUUGUGA
Dre UCCCUGAGACCCUAACUUGUGA b
Hsa UCCCUGAGACCCUAACUUGUGA b
Bfl UCCCUGAGACCCUAACUUGUGA
Sko UCCCUGAGACCCUAACUUGUGA
Spu UCCCUGAGACCCUAACUUGUGA
Hsn UCCCUGAGACCCUAACUUGUGA
Dpu UCCCUGAGACCCUAACUUGUGA
Isc UCCCUGAGACCCUAACUUGUGA
Pca UCCCUGAGACCCUAACUUGUGA
Hru UCCCUGAGACC**A**UAACUUGUG**C**
Csp UCCCUGAGACCCUAACUUGUGA
Ndi UCCCUGAGACCCUAACU**CU**UGA
Cla U-CCCUGAGACCCUAACUUGUGA

miR-133

Pdu -UUGGUCCCCUUCAACCAGCUGU (17186 / 1.36%)
Dme -UUGGUCCCCUUCAACCAGCUGU
Dre **U**UUGGUCCCCUUCAACCAGCUG- a
Hsa **U**UUGGUCCCCUUCAACCAGCUG- a
Bfl -UUGGUCCCCUUCAACCAGCUGU
Sko -UUGGUCCCCUUCAACCAGCUGU
Spu **U**UUGGUCCCCUUCAACCAGCCGU
Hsn **U**UUGGUCCCCUUCAACCAGCCGU
Dpu -UUGGUCCCCUUCAACCAGCUGU
Isc -UUGGUCCCCUUCAACCAGCUGU
Pca -UUGGUCCCCUUCAACCAGCUGU
Hru -UUGGUCCCCUUCAACCAGCUGU
Lgi -UUGGUCCCCUUCAACCAGCUGU
Csp -UUGGUCCCCUUCAACCAGCUGU
Ndi -UUGGUCCCCUUCAACCAGCUGU
Cla -UUGGUCCCCUUCAACCAGCUGU

miR-137

Pdu -UAUUGCUUGAGAAUACACGUAG (336 / 0.03%)
Dme -UAUUGCUUGAGAAUACACGUAG
Dre UUAUUGCUU**A**AGAAUAC**G**CGUA-
Hsa UUAUUGCUU**A**AGAAUAC**G**CGUAG
Bfl -UAUUGCUUGAGAAUACACGU**GA**
Sko -UAUUGCUUGAGAAUACACGUAG
Spu -UAUUGCUUGAGAAUACACGUAG
Hsn -UAUUGCUUGAGAAUACACGUAG
Dpu -UAUUGCUUGAGAAUACACGU**U**G
Isc -UAUUGCUUGAGAAUACACGU**C**G
Hru -UAUUGCUUGAGAAUACACGUA**A**
Lgi -UAUUGCUUGAGAAUACACGUA**A**
Csp -UAUUGCUUGAGAAUACACGUAG
Ndi -UAUUGCUUGAGAAUACACGUAG


miR-153
Pdu UUGCAUAGUCACAAAAGUGAUC (6628 / 0.53%)
Dme UUGCAUAGUCACAAAAGUGAUC
Hsa UUGCAUAGUCACAAAAGUGAUC
Sko UUGCAUAGUCACAAAAGUGAUU
Spu UUGCAUAGUCACAAAAGUGAUU
Hsn UUGCAUAGUCACAAAAGUGAUU
Dpu UUGCAUAGUCACAAAAGUGAUG
Isc UUGCAUAGUCACAAAAGUGAUG
Hru UUGCAUAGUCACAAAAGUGAUC
Lgi UUGCAUAGUCACAAAAGUGAUC
Csp UUGCAUAGUCACAAAAGUGAUC
Ndi UUGCAUAGUCACAAAAGUGAU-
miR-182, -263b
Pdu CUUGGCACUGGUAGAAUUCACUGA (2851 / 0.23%)
Dme CUUGGCACUGGGAGAAUUCAC---
Dre UUUGGCAAUGGUAGAACUCACA--
Hsa -UUGGCAAUGGUAGAACUCACACU
Bfl CUUGGCACUAG-UGGAAUUCUUUG-
Sko UUUGGCAAUAGAUAGAAUUCACA--
Spu UUUGGCAAUUGAUAGAAUUCACACU
Hsn UUUGGCAAUAGAUAGAAUUCAC---
Dpu CUUGGCACUGGAAGAAUUCACA--
Pca CUUGGCACUGGUAGAAUUCACA--
Hru CUUGGCACUGGUAGAAUUCACUGA
Lgi CUUGGCACUGGUAGAAUUCACUGC
Csp CUUGGCACUGGUAGAAUUCACU--
Ndi CUUGGCACUGGUAGAAUUCACUGA
miR-183, -263a
Pdu ---AAUGGCACUGGUAGAAUUCACGG (16045 / 1.27%)
Dme GUUAAUGGCACUGGAAGAAUUCAC---
Dre ---UAUGGCACUGGUAGAAUUCACUG
Hsa ---UAUGGCACUGGUAGAAUUCACU
Bfl ---UAUGGCACUGGUAGAAUUCACU--
Sko ---AAUGGCACUGUAUGAAUUCACUG-
Spu ---UAUGGCACU-AUAGAAUUCACUG-
Dpu ---AAUGGCACUGGAAGAAUUCAC---
Isc ---AAUGGCACUGGAAGAAUUCAC---
Hru ---AAUGGCACUGGUAGAAUUCACGGG
Lgi ---AAUGGCACUGGUAGAAUUCACGGU
Csp ---AAUGGCACUGGUAGAAUUCACGGU
Ndi ---AAUGGCACUGGUAGAAUUCACGG-
Cla ---UAUGGCACUAGAAGAAUUCACG--
miR-184
Pdu UGGACGGAGAACUGAUAAGGGC (23950 / 1.9%)
Dme UGGACGGAGAACUGAUAAGGGC
Dre UGGACGGAGAACUGAUAAGGGC
Hsa UGGACGGAGAACUGAUAAGGGU
Bfl UGGACGGAGAACUGAUAAGGG-
Sko UGGACGGAGAACUGAUAAGGGC
Spu UGGACGGAGAACUGAUAAGGGC
Hsn UGGACGGAGAACUGAUAAGGGC
Isc UGGACGGAGAACUGAUAAGGGC
Pca UGGACGGAGAACUGAUAAGGGC
Hru UGGACGGAGAACUGAUAAGGGC
Lgi UGGACGGAGAACUGAUAAGGGC
Csp UGGACGGAGAACUGAUAAGGGC
Ndi UGGACGGAGAACUGAUAAGGG-
Cla UGGACGGAGAACUGAUAAGGGC
miR-190
Pdu AGAUAUGUUUGAUAUAUUUGGU (5 / 0%)
Dme AGAUAUGUUUGAUAUUCUUGGUUG
Dre UGAUAUGUUUGAUAUAUUAGGU
Hsa UGAUAUGUUUGAUAUAUUAGGU
Bfl UGAUAUGUUUGAUAU--UUGGUUGU
Sko UGAUAUGUUUGUUAU--UUGGUUGGU
Spu UGAUAUGUUUGAUAU--UUGGUU
Hsn UGAUAUGUUUGGUUU--UGGGUUG
Hru AGAUAUGUUUGAUAUAUUUGGU
Lgi AGAUAUGUUUGAUAUACUUGGU
Csp AGAUAUGUUUGAUAUAUUUGGUGG
Ndi AGAUAUGUUUGAUAUAUUUGGUGG
Cla AGAUAUGUUUGAUAUAAUUGGUG
miR-193
--no Platynereis sequence--
Dme UACUGGCCUACUAAGUCCCAAC
Bfl UACUGGCCCCUUAAGUCCCGGU
Sko AACUGGCCUUUUAAGUCCCGCA
Spu UACUGGCCAGCACAAUCCCAGA
Hsn UACUGGCCAGCACAAUCCCAGA
Dpu UACUGGCCUGCUAAGUCCCAAA
Hru UACUGGCCUGCAAAAUCCCAAC
Lgi UACUGGCCUGCAAAAUCCCAAC
Csp AACUGGCCCGUCAAGUCCCUCC

miR-210

Pdu CUUGUGCGUGUGACAGUGACAAU (160 / 0.01%)
Dme –UUGUGCGUGUGACAG**C**G**GC**U**A**--
Dre –**C**UGUGCGUGUGACAG**C**G**GC**U**AA**
Hsa –**C**UGUGCGUGUGACAG**C**G**GC**U**GA**
Bfl -**C**UGUGCGUGUGACAG**C**G**GC**U**GA**-
Sko -UUGUGCGUG**C**GACAG**C**GAC**UUC**-
Spu -UUGUGCGUG**C**GACAG**C**GAC**UGA**-
Hsn -UUGUGCGUG**C**GACAG**C**GAC**UGA**-
Dpu -UUGUGCGUGUGACAG**C**G**GC**U**A**U-
Isc -UUGUGCGUGUGACAG**C**G**GC**U**A**U-
Hru -**A**UGUGCGUG**A**GACAG**C**GAC**CA**C**U**
Lgi -UUGUGCGUG**G**GACAG**C**GA**UCGAU**
Csp **U**UUGUGCGUGUGACAGUGACAAU-
Ndi CUUGUGCGUGUGACAGUGACAAU-
Cla -UUGUGCGUG**A**GACAG**C**GAC**U**AU-

miR-216, -283, -747

Pdu -UAAUCUCAGCUGGCAAAAGUGAG (1 / 0%)
Dme UAAAU**A**UCAGCUGG**U**AA**UUC**U---
Dre -UAAUCUCAGCUGGCAA**CU**GUGA
Hsa -UAAUCUCAGCUGGCAA**CU**GUGA
Bfl -UAAUCUCAGCUGGCAA**CU**GUGAG
Dpu -**A**AAU**A**UCAGC**A**GG**U**AA**UUC**U---
Pca -UAAUCUCAGCUGG**U**AA**UU**GUGA-
Hru -UAAUCUCAGCUGG**U**AA**UUCC**GAG
    -UAAU**A**UCAGCUGG**U**AA**UCC**UGAG
Lgi -UAAUCUCAGCUGG**U**AA**UUC**UGAG
Csp -UAAUCUCAG**U**UGG**U**AA**UUCA**GA-
Ndi -UAAUCUCAGCUGG**U**AAAAGUGAG

miR-219

Pdu UGAUUGUCCAAACGCAUUUCUUG (10 / 0%)
Dme UGAUUGUCCAAACGCA**A**UUCUUG
Dre UGAUUGUCCAAACGCA**A**UUCUU
Hsa UGAUUGUCCAAACGCA**A**UUCU
Bfl UGAUUGUCCAAACGCA**A**UUCUU-
Sko UGAUUGUCCAAACGCA**A**UUCUU-
Spu UGAUUGUCC**G**AACGCA**A**UUCUU-
Dpu UGAUUGUCCAAACGCA**A**UUCUU-
Isc UGAUUGUCCAAACGCA**A**UUCUU-
Hru UGAUUGUCCAAACGCA**A**UUCUU-
Csp UGAUUGUCCAAACGCA**A**UUCUU-
Ndi UGAUUGUCCAAACGCAUUUCU--
Cla UGAUUGUCCAAACGCA**A**UUCUU-
Sro UGAUUGUCCAAACGCA**A**UUCUU-

miR-252a

Pdu CUAAGUACUAGCGCCGCAGGA (4995 / 0.4%)
Dme CUAAGUACUAG**U**GCCGCAGGAG-
Bfl CUAAGUACUAG**U**GCCGCAGG**U**GU
Sko CUAAGUACUAG**U**GCCGCAGGAGU
Spu CUAAGUACUAG**U**GCCG**U**AGG**U**U-
Hsn CUAAGUACUAG**U**GCCGCAG-**U**U-
Dpu CUAAGUACU**C**G**U**GCCGCAGGAG-
Hru CUAAGUACU**GGU**GCCGCAGGA--
Lgi CUAAGUACU**GGU**GCCGC**G**GGA--
Csp CUAAGUACUAG**U**GCCGCAGGAG-
Ndi CUAAGUACUAG**C**GCCGCAGGA--
Cla CUAAGUACUAG**U**GCCGCAGGAG-

miR-278

Pdu UCGGUGGGACUUUCGUUCGUUC (2685 / 0.21%)
Dme UCGGUGGGACUUUCGU**C**CGUU**U**
Bfl UCGGUGGG**GU**UUUCGUUCG**AGU**
Sko UCGGUGGGACUUUCGUUCGUU**U**
Spu UCGGUGGGACUUUCGUUCG**AU**
Hsn UCGGUGGGACUUUCGUUCG**A**U-
Dpu UCGGUGGGACUUUCGU**C**CGU**GU**
Hru UCGGUGGGACUUUCGUUCGUU**U**
Lgi UCGGUGGGACUUUCGUUCGU**CU**
Csp UCGGUGGGACUUUCGUUCGUUC
Ndi UCGGUGGGACUUUCGUUCGUUC
Cla UCGGUGGGACUUUCGUUCG**CU**

miR-281

Pdu UGUCAUGGAGUUGCUCCCUUUA (5 / 0%)
Dme UGUCAUGGA**A**UUGCUC**U**CUUU**G**U
Bfl UGUCAUGGAGUUGCUC**U**CUUU**U**-
Sko UGUCAUGGAGUUGCUC**U**CUU**A**A-
Spu UGUCAUGGAGUUGCUC**U**CUUU**U**-
Dpu UGUCAUGGAGCUGCUC**U**CUUUAU
Hru UGUCAUGGAGUUGCUC**U**CUUUA-
Lgi UGUCAUGGAGUUGCUC**U**CUUUA-
Csp UGUCAUGGAGUUGCUC**U**CUUUA-
Ndi UGUCAUGGAGUUGCUC**U**CUUUA-
Cla UGUCAUGGAGUUGCUC**U**CUUUA-

miR-315

Pdu UUUUGAUUGUUGCUCAGAAAGCC (15996 / 1.27%)
Dme UUUUGAUUGUUGCUCAGAAAGC-
Sko UUUUGAUUGUUGCUCAGAAAG--
Dpu UUUUGAUUGUUGCUCAGAAAGC-
Isc UUUUGAUUGUUGCUCAGAAGGC-
Pca UUUUGAUUGUUGCUCAGAAAGC-
Hru UUUUGAUUGUUGCUCAGAAAGCC
Lgi UUUUGAUUGUUGCUCAGAAAGCC
Csp UUUUGAUUGUUGCUCAGAAAGCC
Ndi UUUUGAUUGUUGCUCAGAAAGCC
Cla UUUUGAUUGUUGCUCAGAAAGCC

miR-375

Pdu UUUGUUCGUCCGGCUCGCGUUA (279 / 0.02%)
Dme UUUGUUCGU**UU**GGCU**UAA**GUUA-
Dre UUUGUUCGU**U**CGGCUCGCGUUA
Has UUUGUUCGU**U**CGGCUCGCGU**G**A
Bfl UUUGUUCGU**U**CGGCUCGCGUUAU
Sko UUUGUUCGU**U**CGGCUCGCG**CG**A-
Spu UUUGUUCGU**U**CGGCUCGCGU**C**A-
Hsn UUUGUUCGU**U**CGGCUCGCGU**C**A-
Dpu UUUGUUCGU**U**CGGCU**UGA**GUUA-
Isc UUUGUUCGU**U**CGGCUCG**A**GUUA-
Hru UUUGUUCGU**U**CGGCUCGCGUUA-
Lgi UUUGUUCGU**U**CGGCUCGCGUUA-
Csp UUUGUUCGUCCGGCUCGCGUUA-
Ndi UUUGUUCGUCCGGCUCGCGUU--
Cla UUUGUUCGU**U**CGGCUCGCGUUA-

miR-2001

Pdu UUGUGACCGUUACAAUGGGCA (1818 / 0.14%)
Sko UUGUGACCGUUA**U**AAUGGGCAU-
Spu **A**UGUGACCG**A**UA**U**AAUGGGCAU-
Hsn **A**UGUGACCGUUACAAUGGGCAU-
Isc UUGUGACCGUUACAAUGGGCAU-
Hru UUGUGACCGUUA**U**AAUGGGCAUU
Lgi UUGUGACCGUUA**U**AAUGGGCAUU
Csp UUGUGACCGUUA**U**AAUGGGCAU-

let-7

Pdu UGAGGUAGUAGGUUGUAUAGU (71 / 0.01%)
Dme UGAGGUAGUAGGUUGUAUAGU-
Dre UGAGGUAGUAGGUUGUAUAGUU
Hsa UGAGGUAGUAGGUUGUAUAGUU
Bfl UGAGGUAGUAGGUUGUAUAGUU
Sko UGAGGUAGUAGGUUGUAUAGUU
Spu UGAGGUAGUAGGUUGUAUAGUU
Hsn UGAGGUAGU**C**GGUUGUA**A**AG**A**-
Isc UGAGGUAGUAGGUUGUAUAGU-
Pca UGAGGUAGUAGGUUGUAUAGU**U**
Hru UGAGGUAGUAGGUUGUAUAGU**U**
Lgi UGAGGUAGUAGGUUGUAUAGU**U**
Csp UGAGGUAGUAGGUUGUAUAGU-
Ndi UGAGGUAGUAGGUUGUAUAGU**U**
Cla UGAGGUAGUAGGUUGUAUAGU**U**

Protostome specific microRNAs alignements

microRNA Alignments (adapted file from Wheeler et al 2009)

Bantam

Pdu UGAGAUCAUUGUGAAAACUGAUU
Dme UGAGAUCAUUUUGAAAGCUGAUU
Dpu UGAGAUCAUUGUGAAAGCUGAUU
Isc UGAGAUCAUUGUGAAAGCUGAUU
Pca UGAGAUCAUUAUGAAAGCUGAUU
Hru UGAGAUCAUUGUGAAAACUGAUU
Csp UGAGAUCAUUGUGAAAACUAAUC
Ndi UGAGAUCAUGGUGAAAACUAAU-
Cla UGAGAUCAUUGUGAAAACUGAUU

miR-2

Pdu UAUCACAGCCAGCUUUGAUGAGC
Dme UAUCACAGCCAGCUUUGAUGAGC
Dpu UAUCACAGCCAGCUUUGAUGAGC
Isc UAUCACAGCCAGCUUUGAUGAGC
Pca UAUCACAGCCAGCUUUGAUGAGU
Hru UAUCACAGCCUGCUUGGAUCAGU
Lgi UAUCACAGCCAGCUUUGAUGAGUU
Csp UAUCACAGCCCGCUUUGUUGACU
Ndi UAUCACAGCCUGCUUGGAUCAUA
Cla UAUCACAGCCUGCUUGGAUUAGA

miR-12

Pdu UGAGUAUUACAUCAGGUACUGA
Dme UGAGUAUUACAUCAGGUACUGGU
Dpu UGAGUAUUACAUCAGGUACUGGU
Isc UGAGUAUUACAUCAGGUACUGGU
Pca UGAGUAUUACUUCAGGUACUGA-
Hru UGAGUAUUACAUCAGGUACUGA-
Lgi UGAGUAUUACAUCAGGUACUGA-
Csp UGAGUAUUACAUCAGGUACUGA-
Csp UGAGUAUUACAUCAGGUACUGA-
Csp UGAGUAUUACAUCAGGUACUGA-

miR-67, -307

Pdu UCACAACCUGCAUGAACGAGGUAA
Dme UCACAACCUCCUUGAGUGAG----
Dpu UCACAACCUCCUUGAGUGAG----
Isc UCACAACCUCCUUGAGUGAG----
Pca UCACAACCUGCAUGAAUGAGGAC-
Hru UCACAACCUGCAUGAAUGAGGAC-
Lgi UCACAACCUGCAUGAAUGAGGAC-
Csp UCACAACCUGCAUGAAUGAGGU--
Ndi UCACAACCUGCAUGAAUGAGGUAA

miR-87

Pdu GUGAGCAAAGUUUCAGGUGUGU
Dme UUGAGCAAAAUUUCAGGUGUG-
Dpu UUGAGCAAAAUUUCAGGUGUG-
Isc GUGAGCAAAGUUUCAGGUGUG-
Hru GUGAGCAAAGUUUCAGGUGUAU
Lgi GUGAGCAAAGUUUCAGGUGUAU
Csp GUGAGCAAAGUUUCAGGUGUGC
Ndi GUGAGCAAAGUUUCAGGUGUGU
Cla GUGAGCAAAGUUUCAGGUGUGA

miR-277

Pdu UAAAUGCAUUAUCUGGUAUGUA
Dme UAAAUGCACUAUCUGGUACGACA
Pca UAAAUGCAUAAUCUGGUAUGAA-
Hru UAAAUGCAUCAUCUGGUAUCUGA
Csp UAAAUGCAUUAUCUGGUAUGUA-
Ndi UAAAUGCAUUAUCUGGUAUGUAA
Cla UAAAUGCAUUAUCUGGUAUGUAA

miR-279

Pdu UGACUAGAUCCACACUCAUCC
Dme UGACUAGAUCCACACUCAUUAA
Dpu UGACUAGAUCCACACUCAUCCA
Isc UGACUAGAUCCACACUCAUCCA
Pca UGACUAGAUCCACACUCAUCCA
Hru UGACUAGAUCCACACUCAUCCA
Lgi UGACUAGAUCCACACUCAUCCA
Csp UGACUAGAUCCACACUCAUCC
Ndi UGACUAGAUCCACACUCAUCC
Cla UGACUAGAUCCACACUCAUCC

miR-317

Pdu UGAACACAGCUGGUGGUAUCUCUUU
Dme UGAACACAGCUGGUGGUAUC-CAGU
Dpu UGAACACAGCUGGUGGUAUCUCAGU
Isc UGAACACAGCUGGUGGUAUAUCAGU
Hru UGAACACAGCUGGUGGUAUCUUCU--
Lgi UGAACACAGCUGGUGGUAUCUUCU--
Csp UGAACACAGCUGGUGGUAUCUCUUU
Ndi UGAACACAGCUGGUGGUAUCUCU--
Cla UGAACACAGCUGGUGGUAUCUCUUU

miR-981

Pdu UUCGUUGUCGUCGAAACCUGC
Dme UUCGUUGUCGACGAAACCUGC
Csp UUCGUUGUCGUCGAAACCUGC
Lgi UUCGUUGUCGACGAAACCUGC

miR-750

Pdu UCAGAUCUAACUCUUCCAGCUCA
Sme UCAGAUCUAACUCUUCCAGUUC**U**
Isc **C**CAGAUCUAACUCUUCCAGCUCA
Pca **C**CAGAUCUAACUCUUCCAGCUCA
Hru **C**CAGAUCUAACUCUUCCAGCUCA
Lgi UCAGAUCUAACUCUUCCAGCUCA
Csp **C**CAGAUCUAACUCUUCCAGCUCA
Ndi **C**CAGAUCUAACUCUUCCAGCUCA
Cla **C**CAGAUCUAACUCUUCCAGCU**U**-

miR-746

Pdu UAGCACCAGAUGAUAUCAGGG
Sme UAGCACCAG**GGU**AUAUC**G**GG**AU**

miR-1996b

Pdu AUCAAGUGAGGUCAGAUCUUGG
Csp AUCAAGUGAGGUCAGAUCUUGG
Ndi AUCAAGUGAGGUCAGAUCUUGG

miR-1992
Pdu UCAGCAGUUGUACCACUGAUGUG
Hru UCAGCAGUUGUACCACUGAUGUG
Lgi UCAGCAGUUGUACCACUGAU**U**UG
Csp UCAGCAGUUGUACCACUGAUGUG
Ndi UCAGCAGUUGUACCACUGAUGUG
Cla UCAGCAGUUGUACCACUGAU**G**UG

miR-1998
Pdu UUGAACGCAGAGAUGUACAUCA
Csp UUGAACGCAGAGAUGUACAUCA
Ndi UUGAACGCAGAGAUGUACAUCA

miR-1989
Pdu UCAGCUGUC**GC**GAUGCCUUCUU
Hru UCAGCUGUCAUGAUGCCUUC**GA**
Lgi UCAGCUGUCAUGAUGCCUUC**C**U
Csp UCAGCUGUCAUGAUGCCUUCUU

miR-36
Pdu UCACCGGGUAUACAUUCAUCCG
Sme UCACCGGGUAGACAUUCAU---

miR-1997
Pdu UCUGCAGGUUCACAUCAGCCCCA
Csp UCUGCAGGUUCACAUCAGCCCCA
Ndi UCUGCAGGUUCACAUCAGCCCCA

miR-1175*
Pdu AGUGGAGAGAGUUCUAUCUCAUC
Csp AGUGGAGAGAGUU**U**UAUCUCAU-
Lgi AGUGGAGAGAGUU**U**UAUCUCAU-

miR-1175
Pdu UGAGAUUCAACUCCUCCAACUGC
Csp UGAGAUUCAACUCCUCCAACUGC
Lgi UGAGAUUCAACUCCUCCAACUGC

miR-1986
Pdu UGGAUUUCCCAUGAUCCGUAAC
Hru UGGAUUUCCCA**A**GAUCCGU**G**A-
Lgi UGGAUUUCCCA**A**GAUCCGU**G**A-

miR-1994
Pdu UGAGACAGUGUGUCCUCCCUCG
Hru UGAGACAGUGUGUCCUCCCU**U**G
Lgi UGAGACAGUGUGUCCUCCCU**U**G
Csp UGAGACAGUGUGUCCUCCCUC**UGG**
Cla UGAGACAGUGUGUCCUCCCUCG

miR-133*
Pdu AGCUGGUUGAAAUAGGGCCAAAU
Lgi AGCUGGUUGAAAU**U**GGGCCAAAU

miR-1993
Pdu UAUUAUGCUGUUAUUCACGAGA
Pca UAUUAUGCUG**A**UAUUCACGAGA
Hru UAUUAUGCUG**C**UAUUCACGAGA
Lgi UAUUAUGCUG**A**UAUUCACGAGA
Csp UAUUAUGCUG**A**UAUUCACGAGA
Cla UAUUAUGCUGUUAUUCACGAGA
Sme UAUUAUGCUGUUAUUCA**U**GAG-

miR-96
Pdu CUUGGCACUGGCGGAAUAAUCAC
Csp CUUGGCACUGGCGGAAUUAUCA-
Lgi CUUGGCACUGGCGGAAUAAUCA-

miR-993
Pdu AGAAGCUCGGUUCUACAGGUAU
Dme -GAAGCUCGUCUCUACAGGUAUCU
Lgi -GAAGCUCGUUUCUACAGGUAUCU
Isc -GAAGCUCGUUUCUACAGGUAUCU
Pca -GAAGCUCGUUUCUACAGGUAGUU

miR-305
Pdu AUUGUACUUCAUCAGGUGCUCUGG
Dme AUUGUACUUCAUCAGGUGCUCUG-

miR-318
Pdu UCACUGGGCUUUGUUUAUCUCA
Dme UCACUGGGCUUUGUUUAUCUCA

miR-989
Pdu UGUGAUGUGACGUAGUGGAACA
Dme UGUGAUGUGACGUAGUGGAAC-

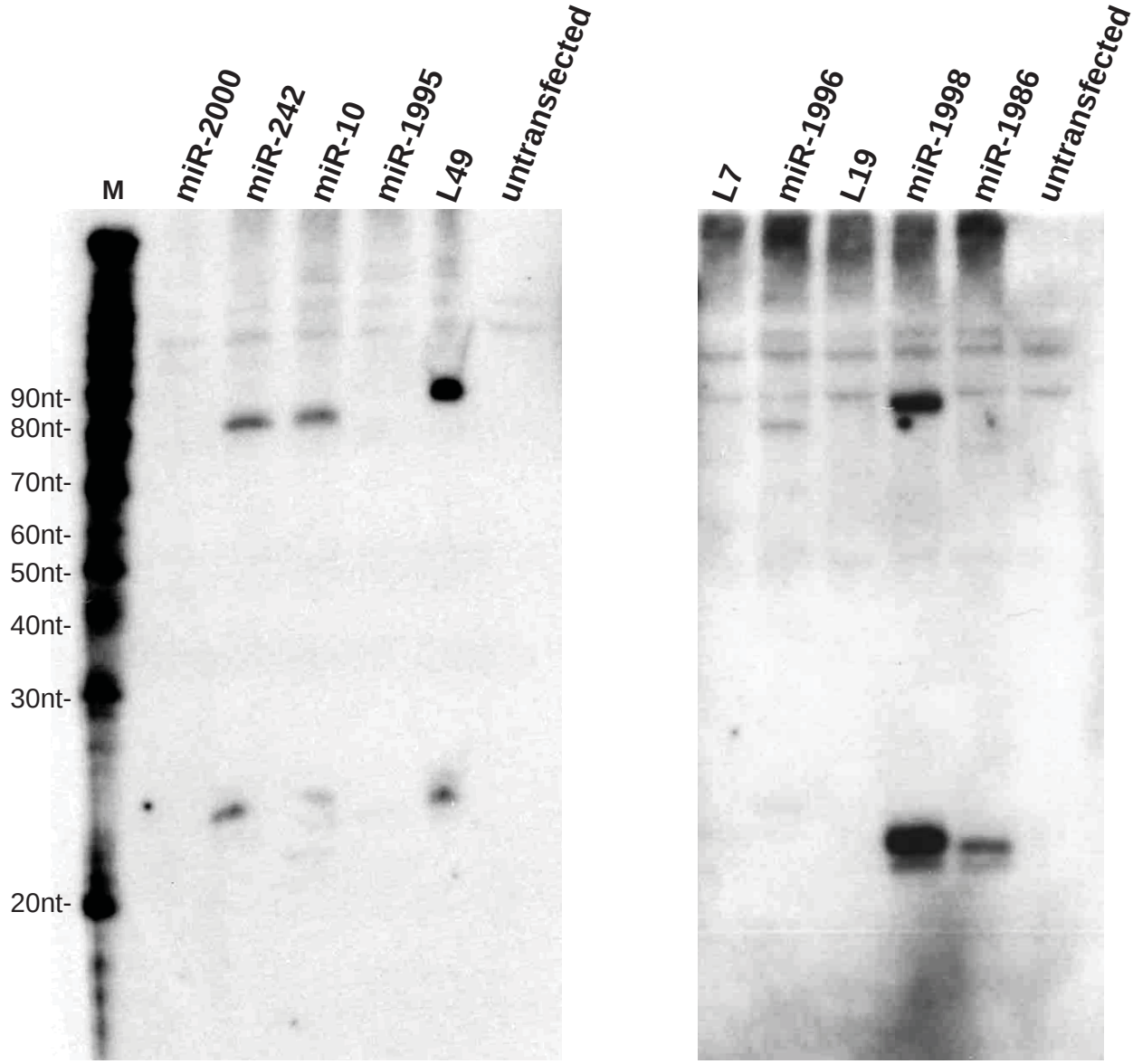
miR-996
Pdu UGACUAGAUUUCAUGCUCGUCUA
Dme UGACUAGAUUUCAUGCUCGUCU-

Lophotrochozoan specific microRNA candidates (identified in at least one lophotrochozoan genome)

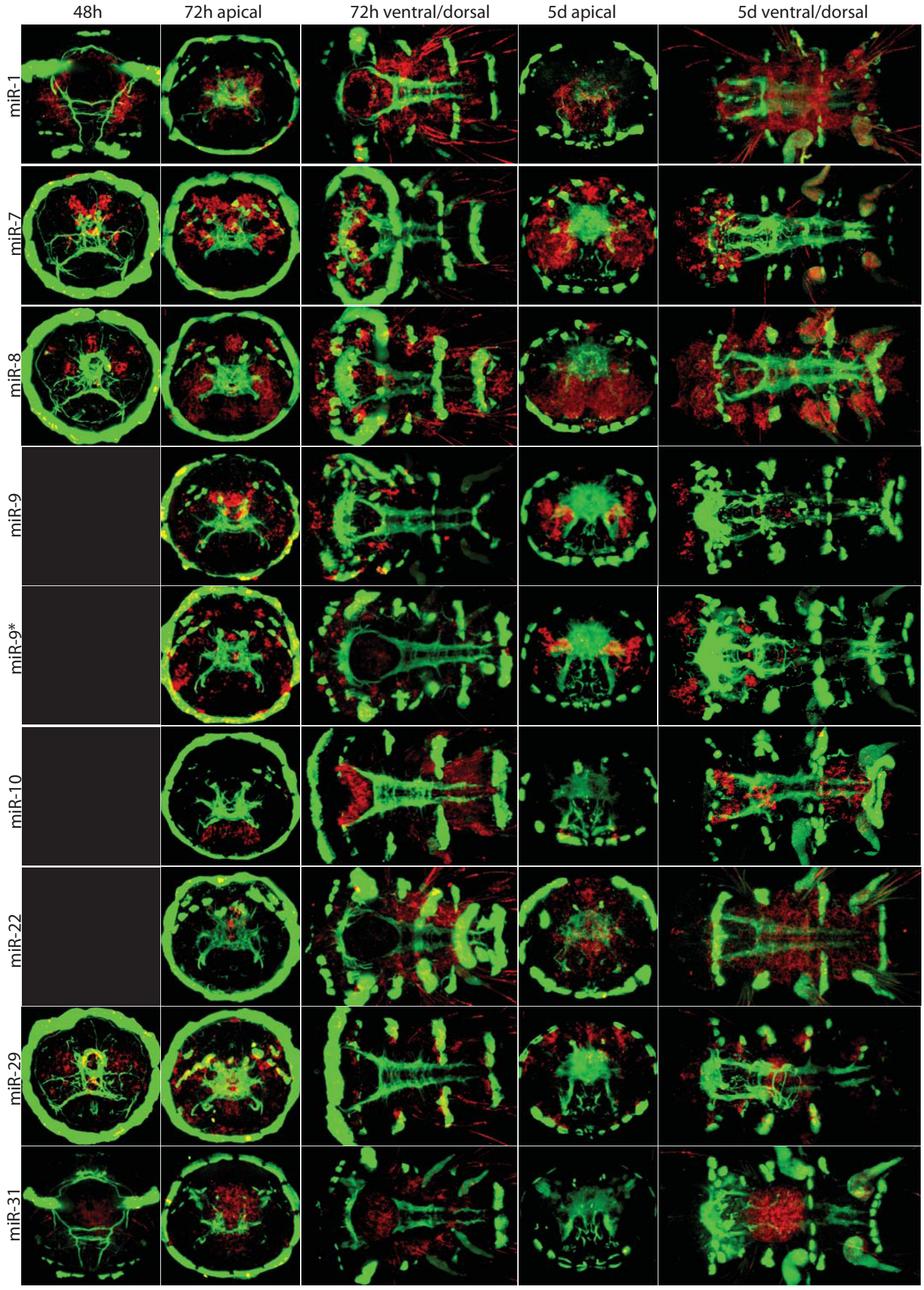
family	members	overall counts	representative sequence	counts	Cc	Hr	Lg
Group L5	9	4241	UGUAAGUUGACAUAGUCCCAGG	3364	1	-	-
Group L4	5	2387	UGGGACUAUGUCAACUUACAAC	2088	1	-	-
Group L2	2	1277	UGUCAAGCAGAUUAAGUAUUGU	1150	1	-	-
Group L6	3	1194	CGGGACUACGUUAACUUCCAGC	1129	1	-	-
Group L40	2	320	CUGGUCUCAAGUGGUGGAUAGA	304	1	-	-
Group L82	2	244	AGUGUUCACUGUGUACGCCUUGGU	233	-	-	1
Group L52	1	143	UGGAAGUUAACGUAGUCCCGGG	143	1	-	-
Group L28	2	126	UAUCCUGGCCUGCAAGUGCACA	116	1	-	-
Group L37	2	122	UUUUGCAAAGUAUCACAGCUUA	95	1	-	-
Group L36	4	142	AUAAGCACCACAAUGUACUAGU	77	1	-	-
Group L1	3	75	AUACUUAUUUUGCUUCUGACAGAU	43	1	-	-
Group L27	1	61	UCGAAUCCUGGUCACGGCACCA	61	1	-	-
Group L18	2	56	UCGACUCCCGGCCGACGCACCA	41	1	-	-
Group L47	1	30	CCAGUCAAUGUUGACACCACCGCU	30	1	-	-
Group L11	2	34	AUGAUGUCACAAUUGACUGCAC	22	1	-	-
Group L73	1	26	GUGUUCACUGUGUACGCCUUGGU	26	-	-	1
Group L79	1	13	UGCUUAGGGUUAGGGUUAGGC	13	-	-	1
Group L57	1	10	UUUGGCACCGUUUUGGUCCACAA	10	1	-	-
Group L80	1	10	CUGCCUGAAAUUUUCGCUCAGACCU	10	-	-	1
miR-1987	6	2578	ACUGCCAGAUGUCAUGUUGUGCA	2259	1	-	-
Group L19	6	1449	AGGUUUUGUGAGAUGUUCAUUGA	1223	1	-	1
miR-2000	2	465	AAAGUCUUCACUACUUUUAGUU	418	1	-	-
Group L30	8	335	UCCCAGGUGGUCUAGUGGG	115	1	-	-
Group L25	8	246	UUAAUCUCAGGGUCGUGGGU	127	3	-	-
Group L20	3	94	UCCUCGUUAGUAUAGUGGG	47	1	-	-
Group L29	4	69	GGUAGCGUGGCCGAGCGGG	25	1	-	-
miR-1995	1	31	GUACAUCUCACAUUGUGACCAU	31	1	-	-

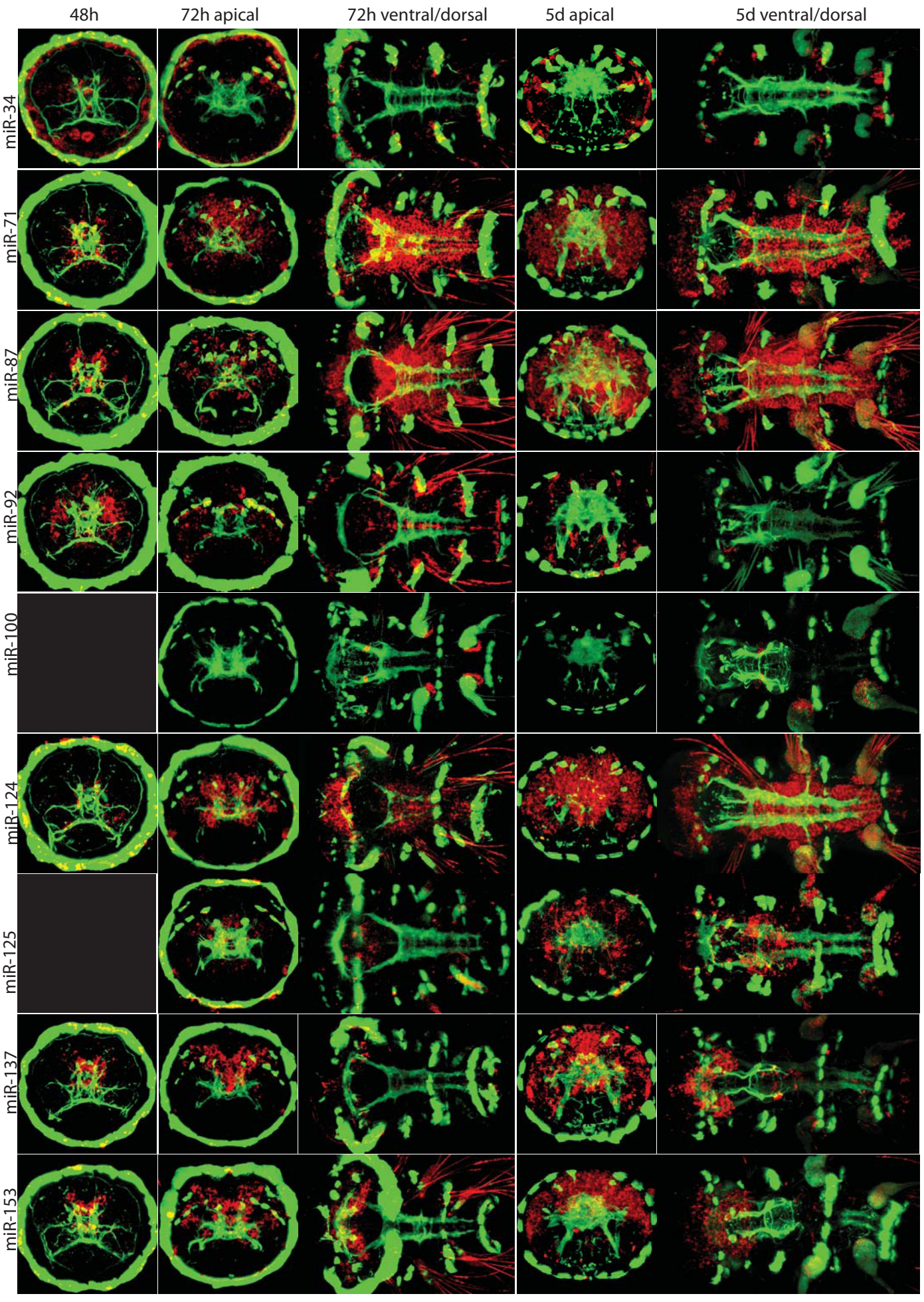
Cc: *Capitella spI*; Hr: *Helobdella robusta*, Lg: *Lottia gigantea*
Numbers indicated how many matching genomic loci were found after blasting the *Platynereis* sequence
Highlighted in red are lophotrochozoan microRNA candidates which were proven to be false predictions (after using the in vivo validation method by Sandmann et al., 2007)
All remaining candidates remain to be tested

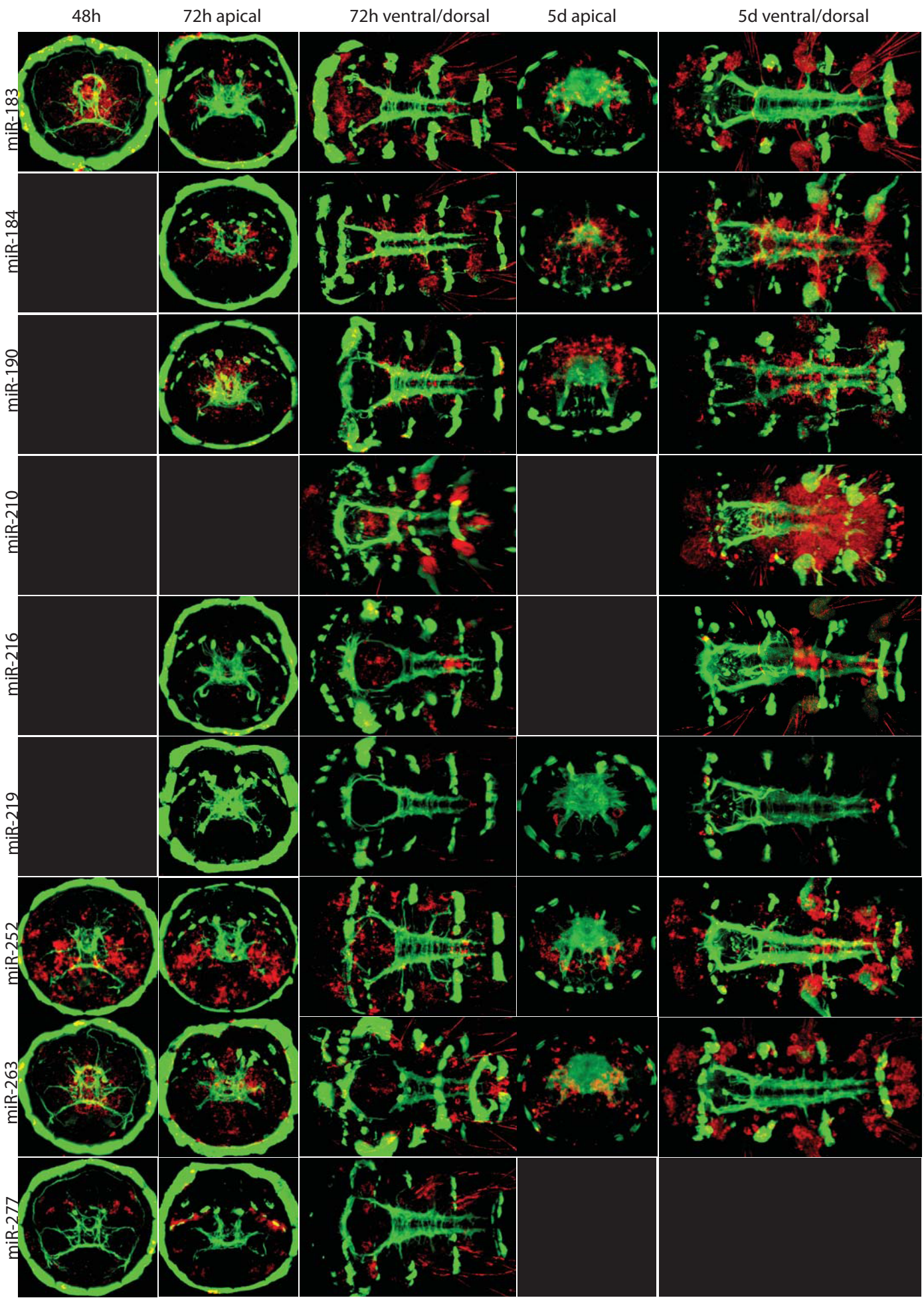
Experimentally confirmed lophotrochozoan microRNA candidates

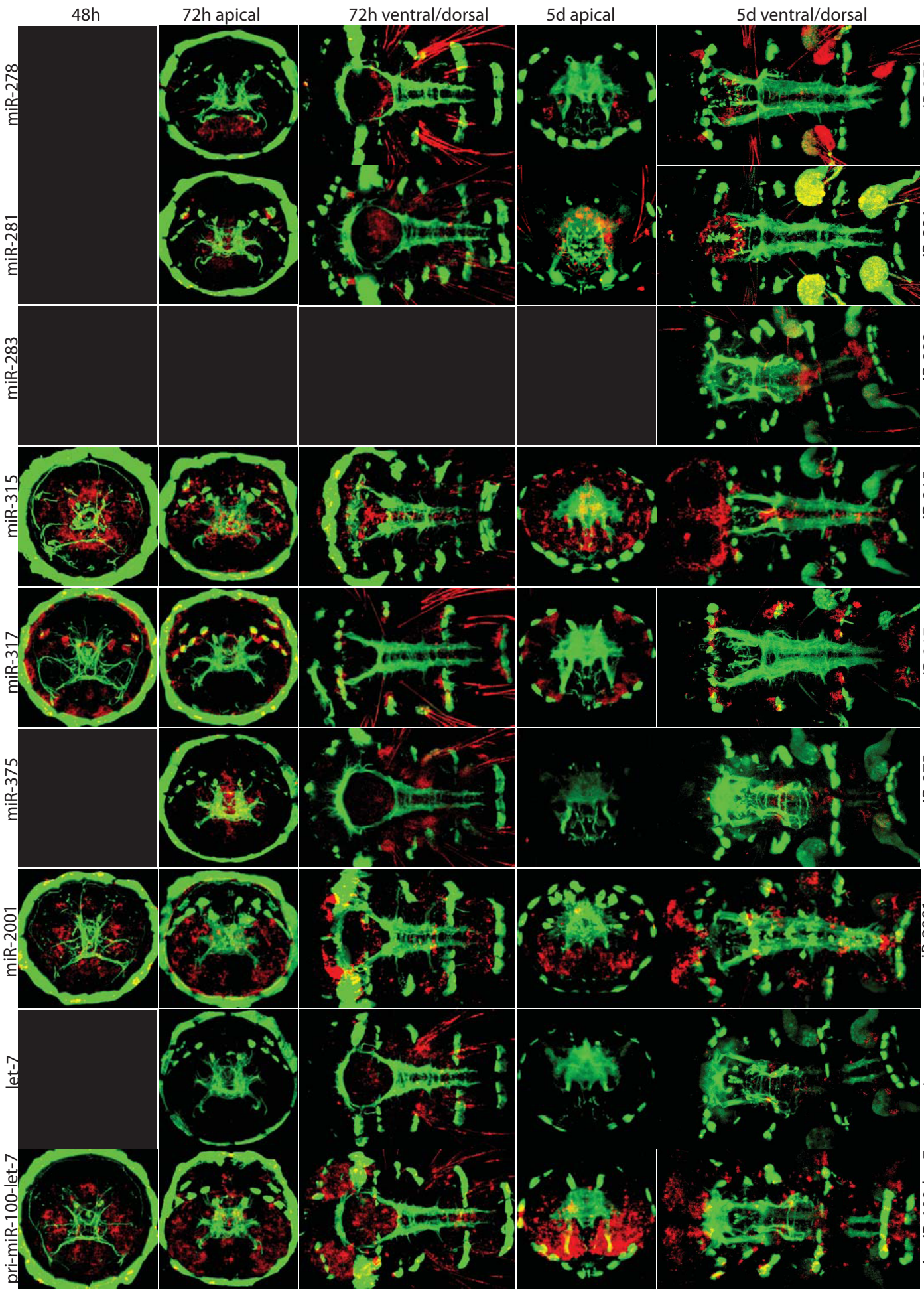


Capitella genomic loci which host candidate microRNAs were over-expressed in S2 cell lines and total RNA was extracted. Northern blot analysis reveals primary transcripts which get processed by the cells' microRNA biogenesis machinery into mature 21- to 23-mers. Precursors are also visible with at ~80nt. Method followed after Sandman et al (2007).


48h
72h apical
72h ventral/dorsal
5d apical
5d ventral/dorsal
miR-1
miR-7
miR-8
miR-9
miR-9*
miR-10
miR-22
miR-29
miR-31
miR-1
miR-7
miR-8
miR-9
miR-9*
miR-10
miR-22
miR-29
miR-31


48h
72h apical
72h ventral/dorsal
5d apical
5d ventral/dorsal
miR-34
miR-71
miR-87
miR-92
miR-100
miR-124
miR-125
miR-137
miR-153


48h
72h apical
72h ventral/dorsal
5d apical
5d ventral/dorsal
miR-183
miR-184
miR-190
miR-210
miR-216
miR-219
miR-252
miR-263
miR-277
miR-183
miR-184
miR-190
miR-210
miR-216
miR-219
miR-252
miR-263
miR-277


48h
72h apical
72h ventral/dorsal
5d apical
5d ventral/dorsal
miR-278
miR-281
miR-283
miR-315
miR-317
miR-375
miR-2001
let-7
pri-miR-100-let-7

Probe	UTR ID	dbEST_Id	Genbank_EST	GenBank_Accn
H2	IB0AAA15CC12FM1_comp_s1	66720341	IB0AAA15CC12FM1	GR911706
H3	IB0AAA15CD03FM1_comp_s420	66720342	IB0AAA15CD03FM1	GR911707
H4	IB0AAA16AE02FM1_comp_s589	66720343	IB0AAA16AE02FM1	GR911708
H8	IB0AAA18DA04FM1_comp_s482	66720344	IB0AAA18DA04FM1	GR911709
H12	IB0AAA19DF09EM1_s1015	66720345	IB0AAA19DF09EM1	GR911710
H13	IB0AAA20CH08FM1_comp_s498	66720346	IB0AAA20CH08FM1	GR911711
H14	IB0AAA20DF03FM1_comp_s1	66720347	IB0AAA20DF03FM1	GR911712
H21	IB0AAA27BE01FM1_comp_s497	66720348	IB0AAA27BE01FM1	GR911713
H22	IB0AAA27DA08FM1_comp_s245	66720349	IB0AAA27DA08FM1	GR911714
H27	IB0AAA28BG11FM1_comp_s1	66720350	IB0AAA28BG11FM1	GR911715
H32	IB0AAA29DB01FM1_comp_s329	66720351	IB0AAA29DB01FM1	GR911716
H34	IB0AAA30CH08FM1_comp_s1	66720352	IB0AAA30CH08FM1	GR911717
H40	IB0AAA33AC08FM1_comp_s1	66720353	IB0AAA33AC08FM1	GR911718
H43	IB0AAA33DC04FM1_comp_s380	66720354	IB0AAA33DC04FM1	GR911719
H44	IB0AAA33DF04FM1_comp_s1	66720355	IB0AAA33DF04FM1	GR911720
H91	IB0AAA32CE06FM1_comp_s1	66720356	IB0AAA32CE06FM1	GR911721
H98	IB0AAA39CG01FM1_comp_s560	66720357	IB0AAA39CG01FM1	GR911722
R7	IB0AAA18BD02FM1_comp_s522	66720358	IB0AAA18BD02FM1	GR911723
R10	IB0AAA19CC08EM1_s1081	66720359	IB0AAA19CC08EM1	GR911724
R13	IB0AAA20AE07FM1_comp_s495	66720360	IB0AAA20AE07FM1	GR911725
R15	IB0AAA20BH02FM1_comp_s117	66720361	IB0AAA20BH02FM1	GR911726
R20	IB0AAA23BH02FM1_comp_s1173	66720362	IB0AAA23BH02FM1	GR911727
R21	IB0AAA24CB04FM1_comp_s120	66720363	IB0AAA24CB04FM1	GR911728
R25	IB0AAA27CD04FM1_comp_s817	66720364	IB0AAA27CD04FM1	GR911729
R29	IB0AAA31DB02EM1_s553	66720365	IB0AAA31DB02EM1	GR911730
R34	IB0AAA35DG01FM1_comp_s650	66720366	IB0AAA35DG01FM1	GR911731
R39	IB0AAA37CC08FM1_comp_s431	66720367	IB0AAA37CC08FM1	GR911732
Y3	IB0AAA43YK07FM1_comp_s640	66720301	IB0AAA43YK07FM1	GR911666
Y7	IB0AAA44YF02FM1_comp_s1	66720302	IB0AAA44YF02FM1	GR911667
Y9	IB0AAA44YH23FM1_comp_s52	66720303	IB0AAA44YH23FM1	GR911668
Y11	IB0AAA44YM16FM1_comp_s260	66720304	IB0AAA44YM16FM1	GR911669
Y13	IB0AAA45YA18FM1_comp_s906	66720305	IB0AAA45YA18FM1	GR911670
Y19	IB0AAA46YD08FM1_comp_s1	66720306	IB0AAA46YD08FM1	GR911671
Y25	IB0AAA47YH22FM1_comp_s267	66720307	IB0AAA47YH22FM1	GR911672
Y34	IB0AAA49YA07FM1_comp_s398	66720308	IB0AAA49YA07FM1	GR911673
Y41	IB0AAA50YB09FM1_comp_s677	66720309	IB0AAA50YB09FM1	GR911674
Y42	IB0AAA50YC17FM1_comp_s278	66720310	IB0AAA50YC17FM1	GR911675
Y43	IB0AAA50YJ12FM1_comp_s924	66720311	IB0AAA50YJ12FM1	GR911676
Y45	IB0AAA51YE03FM1_comp_s333	66720312	IB0AAA51YE03FM1	GR911677
Y49	IB0AAA53YF24FM1_comp_s1	66720313	IB0AAA53YF24FM1	GR911678
Y51	IB0AAA53YJ08FM1_comp_s176	66720314	IB0AAA53YJ08FM1	GR911679
Y62	IB0AAA56YH22FM1_comp_s1	66720315	IB0AAA56YH22FM1	GR911680
Y63	IB0AAA56YM12FM1_comp_s1012	66720316	IB0AAA56YM12FM1	GR911681
Y65	IB0AAA57YB11FM1_comp_s1038	66720317	IB0AAA57YB11FM1	GR911682
Y68	IB0AAA57YH01FM1_comp_s879	66720318	IB0AAA57YH01FM1	GR911683
Y71	IB0AAA58YI08FM1_comp_s297	66720319	IB0AAA58YI08FM1	GR911684
Y74	IB0AAA59YF10FM1_comp_s1	66720320	IB0AAA59YF10FM1	GR911685
Y89	IB0AAA62YH01FM1_comp_s207	66720321	IB0AAA62YH01FM1	GR911686
Y93	IB0AAA63YP10FM1_comp_s1	66720322	IB0AAA63YP10FM1	GR911687
Y98	IB0AAA64YJ11FM1_comp_s972	66720323	IB0AAA64YJ11FM1	GR911688
Y99	IB0AAA64YL06FM1_comp_s1	66720324	IB0AAA64YL06FM1	GR911689
Y102	IB0AAA65YE20FM1_comp_s544	66720325	IB0AAA65YE20FM1	GR911690

Y105	IB0AAA65YM06FM1_comp_s793	66720326	IB0AAA65YM06FM1	GR911691
Y106	IB0AAA65YM20FM1_comp_s1	66720327	IB0AAA65YM20FM1	GR911692
Y108	IB0AAA66YB17FM1_comp_s1	66720328	IB0AAA66YB17FM1	GR911693
Y111	IB0AAA66YK04FM1_comp_s380	66720329	IB0AAA66YK04FM1	GR911694
Y112	IB0AAA66YM10FM1_comp_s1	66720330	IB0AAA66YM10FM1	GR911695
Y115	IB0AAA67YE15FM1_comp_s1104	66720331	IB0AAA67YE15FM1	GR911696
Y118	IB0AAA67YN16FM1_comp_s1	66720332	IB0AAA67YN16FM1	GR911697
Y121	IB0AAA68YI09FM1_comp_s210	66720333	IB0AAA68YI09FM1	GR911698
Y123	IB0AAA68YL05FM1_comp_s909	66720334	IB0AAA68YL05FM1	GR911699
Y124	IB0AAA68YM13FM1_comp_s598	66720335	IB0AAA68YM13FM1	GR911700
Y125	IB0AAA68YN17EM1_s983	66720336	IB0AAA68YN17EM1	GR911701
Y126	IB0AAA69YD14FM1_comp_s1	66720337	IB0AAA69YD14FM1	GR911702
Y132	IB0AAA70YD04FM1_comp_s1	66720338	IB0AAA70YD04FM1	GR911703
Y133	IB0AAA70YE14FM1_comp_s288	66720339	IB0AAA70YE14FM1	GR911704
Y134	IB0AAA70YH24FM1_comp_s748	66720340	IB0AAA70YH24FM1	GR911705

List of abbreviations

3'UTR	3'untranslated region
5'UTR	5'untranslated region
ADARs	Adenosine deaminases acting on RNA
Amphioxus	*Branchiostoma floridae*
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BSA	Bovine serum albumine
Caenorhabditis	*Caenorhabditis elegans*
Capitella	*Capitela species I*
CNS	Central nervous system
C-terminus	Carboxy-terminus of a peptide or protein
DABCO	1,4-diazabicyclo[2.2.2]octane
DAPI	4',6-diamidino-2-phenylindole
Drosophila	*Drosophila* melanogaster
EST	Expressed sequence tag
eIF4E	eukaryotic translation initiation factor 4E
eIF4G	eukaryotic translation initiation factor 4G
GFP	green fluorscent protein
LNA	locked nucleic acid
m_7G	cap 7-methylguanosine cap
mRNA	Messenger RNA
miRNA	microRNAs
miRBase	microRNAs online public database
MZ*dicer*	maternal-zygotic dicer mutant
NBT	Nitro blue tetrazolium
Nematostella	*Nematostella vectensis*
NSW	Natural sea water
N-terminus	Amino-terminus of a peptide or protein
ORF	open reading frame
PBS	Phosphate buffer saline
Pol II	RNA polymerase II
Pdu	*Platynereis dumerilii*
PFA	Paraformaldehyde
RACE	Rapid amplification of complementary ends
RISC	RNA-induced silencing complex
RNAi	RNA interference
RT	Reverse trancriptION
S2 cells	*Drosophila* Schneider cell line
Sea urchin	*Strongylocentrotus purpuratus*
SOP	sensory organ precursor
WMISH	Whole mount *in-situ* hybridization
wt	wild-type