Molecular determinants of miRNA target specificity and tissue-specific studies in *C. elegans*

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Giovanna Brancati

aus Italien

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Mihaela Zavolan

Dr. Helge Grosshans

Prof. Dr. Rene Ketting

Basel, 17 October 2017

The Dean of Faculty
Prof. Dr. Martin Spiess

SUMMARY

MicroRNAs (miRNAs) control gene expression by repressing target messenger RNAs. Target identification is thus key to understand the biological implications of a miRNA in physiological or pathological processes, but it has remained the main challenge in the field.

Traditionally, we refer to "canonical targets" when the 3'UTR of a gene contains a perfect Watson-Crick match to the 5' sequence of the miRNA[1]. However, many non-canonical miRNA binding sites have been identified that display seed mismatches, pairing beyond the seed or both.

In this work, we aimed at understanding the molecular requirements necessary to induce silencing of a transcript by a specific miRNA *in vivo*. Using genome editing and physiological reporters, we focused on miRNA sharing the same seed sequence (miRNA families) as they permit to understand the involvement of both seed and non-seed pairing. For such investigation we studied the *let-7* family of miRNAs because *let-7* is conserved in humans and has been found implicated in several pathologies. We performed our studies in *C. elegans* because this miRNA family has been well characterized in this nematode, and mutant animals have obvious phenotypes that are easy to score.

Our results suggest that target specificity of miRNAs belonging to a family depends on the degree of sequence complementarity between the individual miRNA and the transcript. Particularly, pairing of the 3' sequence of the miRNA is the main determinant to establish preferential binding to a site. In addition, the seed match has a key role in modulating such specificity, as it allows to discriminate between high and low levels of miRNAs. Hence, target specificity of individual miRNAs is not hardwired, but is modulated by the miRNA abundance. We believe that our findings have a broad impact on miRNA target prediction and validation, especially if we want to invest in miRNA therapeutics. Lastly, we show that studying miRNA/target interactions in physiological settings has the power to unequivocally validate targets and expand our knowledge on the miRNA regulatory potential.

In parallel, we succeeded in optimizing a FACS-based protocol to isolate worm cells, which we used to profile cell-type specific small RNAs and tissue-specific transcriptomes at single cell resolution. Given the general lack of methods to obtain primary cells and high quality tissue-specific data in the *C.*

[1] Nucleotides 2 to 8.

elegans community, such results hold the great potential to expand our knowledge about cell-type specific gene expression.

Contents

1 Introduction

1.1 The non-coding RNA era

For a long time, RNA has been considered the intermediary between the information kept in the DNA and protein production, possibly as a consequence of the "one gene-one (protein) enzyme" hypothesis from Beadle and Tatum (Beadle and Tatum, 1941) or the controversial "dogma" proposed by Francis Crick (Crick, 1970). However, a growing amount of evidence suggested unexpected functions for RNA molecules.

When in the late 1950s transfer RNAs and ribosomal RNAs (tRNAs and rRNAs) were discovered, they revealed that RNAs could be involved in protein biosynthesis without being the actual messenger, thus expanding RNA roles to a "platform" for protein production. Later, small nucleolar RNAs with regulatory roles and self-splicing introns with catalytic activity were identified (Cech and Steitz, 2014). Nevertheless, the importance of RNA and the revolutionary effects that newly discovered species of RNA had on gene expression was delayed by the idea of the so-called *junk DNA*. Junk DNA are collectively considered the DNA sequences that do not make polyA+ transcripts and do not encode for proteins (Salditt-Georgieff and Darnell, 1982), whose function was difficult to explain. Introns, for example, have been considered silent pieces of DNA that either jumped in the genomes between contiguous protein-coding elements, or which were vestigial links between them (Gilbert, 1985), until the discovery of alternative splicing revealed that they could be inserted in protein-coding messengers. Even the discovery of the first miRNA, *lin-4*, by the Ambros Lab (Lee et al., 1993) and its regulation of the *lin-14* gene, essential for worm development, did not get the attention that it deserved. miRNAs would have probably remained overlooked without the discovery of RNA interference (Fire et al., 1998) that awarded Fire and Mello the Nobel prize in 2006 and the identification of a second miRNA, *let-7*, again identified in *C. elegans*, which was deeply evolutionarily conserved until humans (Pasquinelli et al., 2000; Reinhart et al., 2000).

Even if the role of introns and the other non-coding RNAs, such as rRNAs, tRNAs, and small RNAs, were eventually recognized, they only explained a small fraction of the junk DNA that lacks protein synthesis potential. In fact, when the human genome was first sequenced, researchers were stunned by the discovery that the ~3 billion bases of DNA contained only ~21,000 genes interspersed in a "sea" of apparently meaningless sequences. Nevertheless, the ENCODE project revealed that 80% of the

human genome, including the junk non-coding DNA, serves some purpose "biochemically speaking" (Pennisi, 2012). It is now established that most of our genome produces RNAs that do not make proteins and that several previously unappreciated classes of RNAs exist, which are characterized by their specific size, biogenesis and functions (Cech and Steitz, 2014).

The relevance of RNA molecules in gene regulation stands out when they are found dysregulated in human diseases. In fact, a role for several types of RNAs, such as microRNA (miRNA), piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs) or long non-coding RNAs, to name a few, has emerged in several pathologies ranging from cancer to neurological or cardiovascular disorders (extensively reviewed in (Esteller, 2011)). Pharmaceutical companies are thus investigating RNA as a new class of *druggable* target, but also as drug. Remarkable examples are the FDA approved antisense oligo *spinraza* that modulates splicing of the gene *smn2* to restore its activity in spinal muscular atrophy patients, and the first siRNA-drug onpattro for hereditary transthyretin-mediated amyloidosis. Although the number of clinical trials involving non-coding RNAs as targets is raising, it is still early days for RNA therapeutics. For example, *miravirsen* (Santaris Pharma), an oligo that binds to miR-122 and affects its biogenesis, was developed to treat chronic hepatitis C (Janssen et al., 2013), but never reached the market. Another example is miR-34, a miRNA with tumor suppressor function that is often dysregulated in cancer, for which several strategies to replenish its pool have been explored (Rupaimoole and Slack, 2017). The company Mirna Therapeutics managed to bring miR-34 mimics (synthetic double-stranded small RNA molecules that match miR-34) to the clinics as a treatment for hepatocellular carcinoma, non-small lung cell carcinomas or pancreatic cancers. However, the study was terminated in phase I for severe adverse effects, although it is unclear if they are due to off-targets of the small RNA or the delivery system utilized (Press Release).

We believe that miRNAs hold great therapeutic potential, especially given their broad involvement in human pathologies. Nevertheless, some aspects of their biology are not fully understood, such as their physiologically relevant targets. Therefore, we studied the determinants for miRNA targeting in the model organism *C. elegans* that allows target validation of conserved miRNAs in a physiological context.

1.2 microRNAs (miRNAs), tiny regulators of gene expression

MiRNAs are endogenous small RNA molecules of about 21 nucleotides that control gene expression by binding through partial complementarity to the 3' untranslated regions (3'UTRs) of messenger RNAs. The latest version of the miRNA database (miRbase v22) lists 2654 mature miRNAs in humans, 469 in *D. melanogaster* and 437 in *C. elegans* (Kozomara and Griffiths-Jones, 2014). Because they are so many and each could potentially have hundreds of targets, it is not surprising that miRNAs are involved in the regulation of almost every biological process and thus found to be misregulated in many human pathologies, ranging from various kinds of cancer to hepatitis and cardiovascular diseases (Rupaimoole and Slack, 2017).

1.2.1 Overview of biogenesis and function

RNA Polymerase II transcribes miRNA genes, which are usually independent genes or can be located in introns, into a precursor molecule called "primary miRNA" (pri-miRNA). This hairpin-containing transcript is processed by Drosha, an RNAse III enzyme, which cleaves it into a ~70nt precursor, the pre-miRNA (Figure 1. 1). The precursor (pre-miRNA) is then exported into the cytoplasm by Exportin 5 where another RNAse III enzyme, Dicer, cleaves the loop of the hairpins releasing a ~20nt duplex. The resulting duplex comprises the 'guide' miRNA, which will be loaded onto the Argonaute (Ago) protein, and its complementary sequence, the passenger strand or miR*, which is usually degraded (miRNA biogenesis is extensively reviewed in (Ha and Kim, 2014)).

The guide miRNA loaded onto Ago forms the core of the miRNA-induced silencing complex (miRISC) that silences the mRNA targets. Through the recruitment of GW182 proteins and other co-factors including the CCR4/NOT complex and the decapping enzymes, miRISC triggers translation inhibition or deadenylation and decay of the target (Jonas and Izaurralde, 2015). In animals, unlike other classes of small RNAs and except for some exceptions (Davis et al., 2005; Yekta et al., 2004), miRNAs usually bind to their targets through imperfect complementarity, while in plants full complementarity is required (Bartel, 2004).

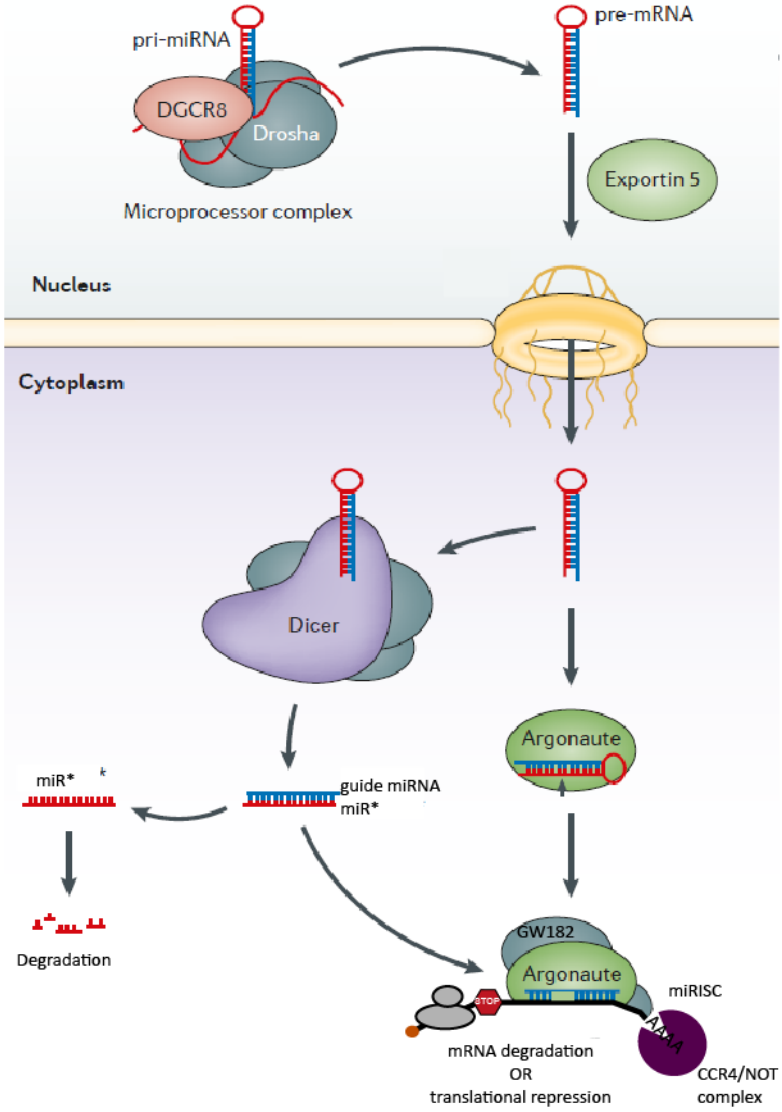


Figure 1. 1 miRNA biogenesis. The primary miRNA (pri-miRNA) is cleaved by the microprocessor complex into the precursor miRNA (pre-miRNA), which is subsequently exported to the cytoplasm through Exportin 5. Dicer then cleaves the pre-miRNA. This cleavage event releases the duplex containing the guide miRNA that will be loaded onto Argonaute protein. The passenger strand, miR*, is usually degraded. Sometimes, Argonaute itself can process the pre-miRNA. The silencing complex miRISC recruits several proteins that will lead to mRNA degradation or translational inhibition of the target. Adapted from (Meister, 2013).

1.2.2 MiRNA site architecture

Generally, the interaction with Ago divides the miRNA into different functional domains, counting from the 5' end: the anchor (nucleotide 1), the seed (2-8), the central (9-12), the 3' supplementary (13-16) and the tail (16-21) (Wee et al., 2012). In this thesis, we will refer to the seed as nucleotide 2-8, and the seed distal sequence as the sequence beyond the seed.

Figure 1.2 Argonaute divides the miRNA into distinct functional domains. Position 2-8, light blue = seed sequence; position 9-21, black = seed distal. Positions 13-16, purple = 3’ supplementary pairing as defined by (Wee et al., 2012)

The seed sequence (nucleotides 2-8) has been shown to be the main determinant for target identification and silencing, being both necessary and sufficient in most cases (Bartel, 2009; Doench and Sharp, 2004). The heptamer that has perfect Watson-Crick (W-C) complementarity to the seed is called “seed match” (Lewis et al., 2005). It was originally identified computationally as the sequence yielding the strongest signal in miRNA target predictions (Lewis et al., 2003). The strong conservation suggests a central role for target specification and in fact many experiments have shown that seed-containing sites are functional in most cases (Brennecke et al., 2005; Broughton et al., 2016; Doench and Sharp, 2004; Grimson et al., 2007; Lai, 2002; Lai, 2004).

The role of the seed/seed match pairing has also been investigated through structural studies (Schirle and MacRae, 2012; Wang et al., 2009) that have been instrumental for understanding how the guide miRNA is loaded onto Argonaute and how they interact with the target mRNAs. The current model is based on the structures of Ago bound to the miRNA, before and after duplex formation (Schirle et al., 2014). This model suggests that the protein exposes to the solvent the seed nucleotides in a pre-arranged helical form that favors initial target recognition and binding, overcoming the entropic penalty for miRNA/target duplex formation (Parker et al., 2009). In particular, it was suggested that nucleotides 2-5 are exposed while 6-8 are not, due to a kink in the small RNA backbone. After pairing of the target to the miRNA nucleotides 2-4, Ago would change conformation and allow access to the remaining seed nucleotides (Chandradoss et al., 2015; Schirle et al., 2014). Following complete (2-8) seed/seed match binding, another conformational change would expose to the solvent the remaining 3’ supplementary domain of the miRNA (nucleotide 13-16) allowing further pairing and stabilization of the miRNA/target duplex (Chandradoss et al., 2015; Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012; Schirle et al., 2014). Additional recent single molecule studies have also highlighted the importance of the seed sequence in mediating the interaction between miRISC and the target *in vitro* (Jo et al., 2015; Salomon et al., 2015).

Sites containing a perfect Watson-Crick paired seed match are considered *canonical* and have been divided in categories according to the number of pairing nucleotides. We distinguish 8mers

(nucleotides 2-8 with an A as the first nucleotide on the target, T1A), 7mer-m8 (nucleotide 2-8), 7mer-1A (nucleotides 2-7 with T1A) and 6mers (nucleotides 2-7), (Figure 1. 3 and (Bartel, 2009)).

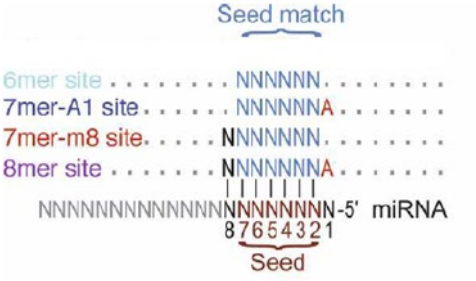


Figure 1. 3 Canonical miRNA sites. The four canonical sites have contiguous Watson-Crick matches to the miRNA 5' region. Adapted from (Friedman et al., 2009)

While sixmers are almost not active (Agarwal et al., 2015; Grimson et al., 2007), sites that feature more complementarity between the miRNA and the target are more effective in mediating silencing in this order 6mer < 7mer-1A < 7mer-m8 < 8mer (Baek et al., 2008; Chandradoss et al., 2015; Grimson et al., 2007).

Lastly, different miRNAs can share the same seed sequence and are then grouped into "families". Considering the importance of the seed sequence in target recognition, miRNAs belonging to the same family are thought to share the same targets and be functionally redundant.

Nucleotide 1 and 9-12 do not play active roles in target identification. The first nucleotide of the miRNA, the anchor, does not seem important for binding to the target, but rather for loading of the small RNA onto Ago (Schirle et al., 2014). The central domain (9-12) is usually unpaired and forms a loop that avoids cleavage of the target, by preventing miRISC from assuming a cleavage-competent conformation (Wee et al., 2012). Interestingly, the supplementary pairing (nucleotides 13-16) can in rare cases enhance target repression (Brennecke et al., 2005; Friedman et al., 2009; Grimson et al., 2007) and mutations in this region can affect miRNA function (Chin et al., 2008; Wee et al., 2012; Zhang et al., 2015).

1.2.3 MiRNA target prediction

Identification of miRNA targets is key to understand miRNA functions, and yet it is the most difficult task. Some of the best-known and validated miRNA targets have been identified through genetic screens that, in model organisms like the worm or flies, led to the discovery of the miRNAs *lin-4*, *let-7, lsy-6, bantam* and their targets (Hipfner et al., 2002; Johnston and Hobert, 2003; Lee et al., 1993; Reinhart et al., 2000). Genetic screens are usually appealing because they are able to detect genes that are physiologically relevant to the miRNA of interest, but because many miRNAs do not induce an obvious phenotype when mutated (Miska et al., 2007), they are often an unviable option to identify miRNA targets, other than the ones that have already been identified. Screens have been traditionally difficult to carry out in mammalian systems, but thanks to the advent of the CRISPR/Cas9 editing technique genome wide screens are becoming possible (Doudna and Charpentier, 2014).

Experimental methods that allow quantification of gene expression can identify the effects of miRNA misregulation on several genes and have the potential to instruct target identification. Small-scale approaches (such as Northern blots and RTqPCR) or genome-wide methods (such as microarrays and RNA sequencing (RNA-seq)) allow estimation of changes in mRNA levels in the presence and absence of a miRNA. Ribosome profiling is often coupled to RNA-seq because it can indirectly test protein production, by assaying ribosome associations and position on the messenger (Ingolia, 2016). Just like RNA-seq, but on the protein side, mass-spectrometry approaches, such as stable isotope labeling by amino acids in cell culture (SILAC), can reveal protein changes following miRNA misregulation (Baek et al., 2008; Selbach et al., 2008). Collectively, these approaches are useful to determine how putative targets are regulated by specific miRNAs, whether they are degraded or if they are stalled in translation (Eichhorn et al., 2014; Guo et al., 2010; Hendrickson et al., 2009). However, such genome-wide approaches rarely allow identification of direct targets, as many genes might indirectly change their expressions because of the induced perturbation, without being direct miRNA targets. Moreover, although some genes are strongly affected by miRNAs, many others do not show a strong reduction in their output even if they are supposedly bona-fide targets and carry a miRNA binding site (Baek et al., 2008). For these reasons, these methods alone are not good predictors of miRNA targets, although instrumental to understand miRNA effects on gene expression.

To identify targets genome-wide, *in silico* prediction tools have been developed. Because the seed match sequence shows strong evolutionary conservation, the algorithm requirement for a conserved perfect Watson-Crick pairing to the seed has greatly improved the reliability of target

identification and exclusion of false positives (Lewis et al., 2003). Some additional features seem to be important to determine a functional and effective miRNA site and are thus taken in considerations by prediction tools such as TargetScan (Agarwal et al., 2015). 1) High AU content near the miRNA site, which would make the site more accessible. 2) positioning 15nt away from the stop codon but away from the center of long 3'UTR, probably to avoid interference of a translating ribosome and because the center of a UTR could be less accessible, respectively. 3) Proximity to another site for a co-expressed miRNA site (no more than 40nt) (Baek et al., 2008; Bartel, 2009; Grimson et al., 2007; Selbach et al., 2008). 4) An adenosine base in the target (T1A), opposite position 1 of the miRNA, which is recognized by a binding pocket in Ago and somehow anchors the target and increases the affinity (Schirle et al., 2014). Different prediction programs such as TargetScan, PicTar, miRanda, RNA hybrid (Agarwal et al., 2015; Betel et al., 2010; Krek et al., 2005; Kruger and Rehmsmeier, 2006; Lall et al., 2006) give results that only partially overlap, as the algorithms are slightly different and reward or penalize different nucleotides and site features. RNAHybrid for example predicts miRNA sites looking for the most favorable hybridization site considering minimum free energy and conservation (Rehmsmeier et al., 2004) and allowing pairing outside of the seed match (Kruger and Rehmsmeier, 2006). Several tools are often used in parallel to ensure prediction of all the possible targets given that predictions can vary significantly (Min and Yoon, 2010). Regardless of the tool used, experimental validation is necessary due to the high false discovery rate of all predictions.

1.2.4 CLIP based methods

In silico predictions are more powerful when coupled to biochemical methods, especially when they can identify endogenous interactions. In early days, pulldown of Ago and associated miRNAs held great promises to identify miRNA-bound transcripts. However, they did not help refining target prediction methods because the false rate discovery was still very high (Beitzinger M, 2007; Easow et al., 2007; Hammell et al., 2008). In fact, it seems that during the lysis and the immunoprecipitation artificial complexes can form due to association of molecules that are not interacting *in vivo* (Mili and Steitz, 2004).

Part of the problem was solved with the advent of CLIP based method (Hafner et al., 2010; Licatalosi et al., 2008; Ule et al., 2003). Relying on the fact that UV-irradiation (254nm) of tissues can induce covalent bonds between a protein and a molecule of RNA that are in close proximity and likely in contact, the cross-linking and immunoprecipitation (CLIP) method allows stringent purification of

RNA-protein complexes and cDNA synthesis of the bound segment of RNA. When coupled to high-throughput sequencing (HITS-CLIP), the method allows genome-wide identification of RNA-protein interaction in an unbiased way (Licatalosi et al., 2008) and relies mostly on the ability to pull-down the protein of interest. Ago HITS-CLIP performed in mouse brains revealed that both the miRNA and the target are cross-linked to Ago, as expected from their close proximity (Schirle et al., 2014) with an "Ago-footprint" on the messenger of about 62 nucleotides (Chi et al., 2009).

It was then observed that, during cDNA synthesis, the reverse transcriptase stops at the cross-linking sites creating truncated fragments or introducing deletions. Sequencing of such molecules allows mapping of the cross-linked site at single nucleotide resolution (individual-nucleotide resolution CLIP (iCLIP) (Konig et al., 2011; Zhang and Darnell, 2011).

A slightly different approach, photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP), which requires incorporation of 4-thiouridine (4SU) into RNAs prior UV cross-linking at 365nm, seems more efficient in capturing miRNA and targets (Hafner et al., 2010). In this technique, the exact cross-linked nucleotide can be identified *in silico* because a cross-linked thymidine is converted to cytidine during cDNA synthesis of the recovered RNA segments. However, the method is not optimal as it relies on 4SU that can potentially induce some cytotoxicity.

The real breakthrough came with CLASH, which stands for cross-linking, immunoprecipitation and sequencing of hybrids (Helwak et al., 2013). In fact, this method allows simultaneous capture of the miRNA and the target in one hybrid molecule, called chimera, which forms through a ligation step added to the CLIP protocol. The method and its variants, which sometimes do not require an independent ligation step to yield chimeras, have successfully been applied to several systems, such as mouse brains, cell cultures and *C. elegans,* and identified many miRNA target interactions that happen *in vivo* (Broughton et al., 2016; Grosswendt et al., 2014; Moore et al., 2015).

However, CLIP-based methods are not perfect. Amino acids have different properties and thus can cross-link differently, influencing the efficiency of cross-linking of different proteins to a site. Additionally, target enrichment depends on the transcript expression and its decay, with low abundant transcripts difficult to be detected. Lastly, transient interactions between Ago and its targets might be missed.

Most of the sites identified with such approaches reside in the 3'UTRs of the messengers, close to the stop codon or the poly(A) site, and are enriched for a canonical seed match, in accordance with

previous knowledge (Bartel, 2009). However, the CLIP-based methods revealed that miRNA sites can be present also in coding regions or ncRNAs, opening up a new research direction. Currently, it is unclear if sites in coding exons are functional. For example, in worms lacking Argonaute (in which miRNA targets should show upregulation) mRNAs with binding sites in their coding regions were not as upregulated as the ones with sites in 3'UTRs (Zisoulis et al., 2010), neither did they appear downregulated in HEK293 (Hafner et al., 2010). However, it has been suggested that such sites might drive translational inhibition (Hausser et al., 2013) and cooperate with sites in 3'UTR to achieve efficient silencing (Fang and Rajewsky, 2011).

Additionally, many of the identified Ago bound fragments (27% in (Chi et al., 2009), 40% (Zisoulis et al., 2010), 6% (Hafner et al., 2010), 60% (Helwak et al., 2013)) had no evident canonical seed match (orphan clusters) and showed a high degree of base paring of the 3' sequence of the miRNA (Broughton et al., 2016; Moore et al., 2015). In particular, clustering of chimeric miRNA/target interactions identified 5-7 different types of miRNA sites (Broughton et al., 2016; Helwak et al., 2013; Moore et al., 2015), comprising "seed-only" sites, "seed match plus different degrees of 3' supplementary pairing" or "non-seed" sites in which target and miRNA pair elsewhere.

Clearly, the simple evidence that Ago is bound to a miRNA and a target does not necessarily mean that this has functional consequences. Agarwal and colleagues have shown that most of the non-canonical sites, highly abundant in chimeras, do not seem to be effective in mediating silencing (Agarwal et al., 2015). Therefore, putative targets still need validation with other methods that measure gene expression. Nevertheless, the high amount of non-canonical sites identified through these biochemical approaches suggests that more flexible pairing rules might exist for different miRNA/mRNA pairs and that endogenous interactions are more complicated than simple seed/seed match pairs. Considering that some non-canonical sites have been previously validated *in vivo* and found to be functional (Ha et al., 1996; Vella et al., 2004a), it is clear that canonical targeting "laws" do not always apply.

1.2.5 Non-canonical miRNA binding sites

Although prediction of perfect matches to the seed sequence has been instrumental to understand miRNA targeting and their function, a growing amount of authors reports occurrence of non-canonical targets *in vivo*, suggesting that miRNA targeting is more flexible than expected.

As previously described, there are several arguments that favor the seed match as the main determinant for targeting, from the general experimental observations that, following a miRNA perturbation, the upregulated or downregulated transcripts sport a seed match, to the fact that seed matches show an astonishing conservation throughout evolution (Bartel, 2009). However, target validation is mostly based on studies in which miRNAs and their targets are ectopically expressed in heterologous context, which could potentially induce artificial interactions (Brennecke et al., 2005; Doench and Sharp, 2004; Krek, 2005; Lewis et al., 2005). When many of the *lsy-6* targets predicted *in silico* based on the seed match occurrence did not show repression in worms, it was suggested that the seed match sequence is not always sufficient for silencing (Didiano and Hobert, 2006). This was an intriguing result at the time and called for more physiological assays for target validation (Didiano and Hobert, 2008). However, *in vivo* studies have lagged behind, possibly because they have a lower throughput compared to cell-based assays, and left the scene to the "canonical seed match" as main determinant for targeting.

Nevertheless, sites with seed 'imperfections', such as G: U wobble pairs or bulges exist. These "imperfect" or "non-canonical" sites have imperfect Watson-Crick seed match pairing. For example, the two sites in the *lin-41* 3'UTR or one of the *lin-4* site in the *lin-14* 3'UTR in *C. elegans* contain a bulge or a G: U wobble in the center of their seed match (Ha et al., 1996; Vella et al., 2004a). The list of non-canonical sites contains also the "pivot" sites, first described in mouse brains (Chi et al., 2012) and characterized by a bulged G between position 5 and 6; the "centered" sites, in which the central sequence of the miRNA pairs with the target (Shin et al., 2010; Wu and Belasco, 2005), and the more exotic "seedless" sites that show no seed match at all (Flamand et al., 2017; Lal et al., 2009). According to structural studies and modeling, such seed imperfections can be accommodated into Ago proteins (Gan and Gunsalus, 2015; Wang Y, 2008) without strong increase of the energy of interaction (Wee et al., 2012), especially if they lie at the "periphery" of the seed (Grosswendt et al., 2014; Moore et al., 2015). However, controversy on the functionality of imperfect sites remains, as shown by the contrasting results that described a G: U wobble pair in the seed as both detrimental (Brennecke et al., 2005; Doench and Sharp, 2004) and benign (Didiano and Hobert, 2006).

Although imperfect sites comprise some of the best experimentally validated miRNA sites, their functionality is generally considered compromised and they have been traditionally disregarded. However, sites with poor seed complementarity can be functional when compensated by additional binding of the 3' sequence of the miRNA (Brennecke et al., 2005).

1.2.6 MiRNAs can bind their targets through their 3'-end sequence

In contrast to the many studies showing the importance of the seed sequence, very few reports focus on the role of the sequences beyond the seed/seed match. Studies in mammals and flies showed that the majority of sites do not have more 3'pairing than expected by chance (Lewis et al., 2005) (Brennecke et al., 2005). However, several sites with 3'pairing can be detected across flies genomes and for those the degree of paired nucleotides rather than their identity is conserved. This suggests that such pairing might have a function and that sequence conservation is not necessarily a good predictor (Brennecke et al., 2005). In addition, further conservation analysis revealed that 3' supplementary pairing centered at nucleotide 13-16 of the miRNA is a conserved feature (Grimson et al., 2007) and its importance in duplex formation is backed up by structural studies of Ago (Schirle et al., 2014). Nevertheless, the importance of the 3' additional pairing is still controversial as in vitro experiments show that RISC has the same binding and dissociation properties with both seed-only sites and seed+3'pairing, suggesting that the distal pairing contributes poorly to target binding (Wee et al., 2012). Others suggest that 3' supplementary sites tend to be slightly more effective than seed-only sites, although they are rather rare and comprise less than 5% of total sites that can be predicted (Friedman et al., 2009; Garcia et al., 2011; Hafner et al., 2010).

Pointing to an underappreciated role for the sequence beyond the seed, deleterious effects for mutations affecting the seed distal pairing have been described. The Fire lab has shown that mutations outside of the seed diminish *let-7* functionality, but they do not affect *lin-4* (Zhang et al., 2015). Similarly, a mutation outside of the seed match in the *let-7* site on the 3'UTR of the gene *kras* affects silencing and is correlated with bad prognosis in non-small cell lung carcinoma (Chin et al., 2008). In addition, many of the binding sites in drosophilid genomes, like the miR-7 site on the *hairy* 3'UTR, contain a perfect seed match and extensive 3' binding (Brennecke et al., 2005; Stark et al., 2003). More recently, while showing that the seed/seed match pairing characterizes most of the miRNA/target interactions *in vivo*, Ago-chimeras revealed the occurrence of many non-canonical sites (Chi et al., 2012; Helwak et al., 2013; Luna et al., 2017) and identified duplexes featuring pairing beyond the seed match in several cases (Broughton et al., 2016; Moore et al., 2015).

1.2.7 Differential targeting of miRNAs belonging to families

Duplexes in which targets and miRNAs pair beyond the seed seem to be common *in vivo*, as they can be widely identified in Ago-chimeras. The Darnell and Pasquinelli labs proposed that 3' pairing can establish specific silencing by individual miRNAs in single cell reporters or *in vivo* in worms (Broughton et al., 2016; Moore et al., 2015). This is particularly relevant for miRNAs sharing the same seed sequence, but having divergent 3' ends. It was observed that different paralogous miRNAs cross-link with their targets at different positions. This might be caused by the different duplexes they form thanks to their unique 3' sequences (Zhang and Darnell, 2011).

Paralogous miRNAs are in most cases considered redundant because they share the same seed and are predicted to target the same messengers. However, the seed sequence alone does not explain why miRNAs belonging to the same family are not redundant and why some targets are preferentially repressed by only one individual paralogue. Non-redundant miRNA families whose individual family members have specific activities, such as the *let-7* family, have been characterized in *C. elegans* (Ambros, 2011; Broughton et al., 2016; Drexel et al., 2016; Tsialikas et al., 2017) and in mammals too, (Cimadamore et al., 2013; Liu et al., 2008; Subasic et al., 2015; Zhao et al., 2013).

The existence of non-redundant miRNA families is usually justified by their expression patterns. Different paralogues with the same seed can be redundant and repress the same messengers when co-expressed; alternatively, they might be non-redundant because their expression pattern differs either in space or in time. However, paralogues miRNAs are often co-expressed and they still show non-redundant activity (Drexel et al., 2016; Roush, 2008). Recently, miR-790 and miR-791 in worms were reported to show non-redundant activity in CO_2 mediated response even if present in the same neurons (Drexel et al., 2016). This is a beautiful example of a unique miRNA function, which is apparent only when studied with a specific assay. It is possible that many miRNAs belonging to a family and whose mutation scored as wild type in previous studies (Alvarez-Saavedra and Horvitz, 2010; Miska et al., 2007) have a phenotype, which will be revealed only once a precise assay to detect it becomes available. Therefore, more miRNAs with specific functions and likely more families that are non-redundant might be waiting to be identified.

Given all these observations, it is difficult to conceive how the seed/seed match pairing alone could explain non-redundant activity of miRNAs sharing the same seed. Because the 3' sequence is unique to each miRNAs (with few miRNAs showing a high degree of similarity), it could indeed be responsible for specific silencing. Accordingly, PAR-CLIP, CLASH and iCLIP experiments suggested that

only a fraction of detected sites has a perfect seed match, whereas the rest contains non-canonical sites and overall an extensive 3' pairing (Broughton et al., 2016; Hafner et al., 2010; Helwak et al., 2013; Moore et al., 2015).

One of the reasons that possibly slowed down discovery and validation of non-canonical sites and additional 3'pairing is that *in silico* tools predict them rarely (except for a few exception, such as miRanda and RNA hybrid, (Betel et al., 2010; Enright, 2003; Kruger and Rehmsmeier, 2006)). Interestingly, different programs largely fail in predicting sites identified through HITS-CLIP, with only 3% of sites predicted simultaneously by different programs (Zisoulis et al., 2010). Generally, they tend to predict different set of targets (Min and Yoon, 2010). This shows that, although a conserved seed match is an important requirement for target identification, it does not explain the interactions that supposedly occur *in vivo,* thus a big gap between *in silico* and biochemical approaches exists. A biophysical method that predicts sites building up from the knowledge obtained through CLIP experiments has been developed recently (MIRZA) (Khorshid et al., 2013). Predictions are based on two main parameters: the *quality of the target*, which considers the affinity of a miRNA for a CLIP fragment (calculating all the possible structures that they can form) and the *frequency of a target*, which is the amount of miRNA-loaded RISC bound to a messenger. This method shows that non-canonical sites can be confidently predicted and suggests that they are more abundant for miRNAs whose level is very high. In other words, lowly expressed miRNAs might efficiently bind (high affinity) canonical sites, while highly abundant miRNAs bind both canonical and (low affinity) non-canonical sites (Khorshid et al., 2013).

1.2.8 MiRNA levels

It is somehow uncommon to consider the cellular level of miRNAs and the way it affects their repressive activity, although the so-called "rheostat model" had already been proposed by Bartel in 2004 (Bartel and Chen, 2004). In this model, every cell expresses a specific set of miRNAs and the combination of their identity and levels differentiates cell types. Thus, a target will be silenced if it contains sites that are either highly complementary to a miRNA, regardless of its levels, or by another miRNA that is only partially complementary, i.e. it has a perfect canonical seed match, or it is highly abundant (Bartel and Chen, 2004).

This model fits well with some previous studies that showed in HeLa cells that a transgene carrying *let-7* seed matches was not silenced by the endogenous miRNA and was repressed only when additional *let-7* was introduced via transfection (Doench and Sharp, 2004). Additional results obtained *in vivo* with *C. elegans* suggest that extremely high levels of *let-7* family members can rescue the *let-7* mutation, which is usually lethal (Hayes et al., 2006; Li et al., 2005). However, in this case, it is not known if the over-expressed miRNAs repress the same *let-7* targets or other genes that can anyway allow survival in the absence of *let-7*.

In conclusion, the growing amount of data based on CLIP methods showing the existence of several kinds of non-canonical sites suggests that individual miRNAs might be unique regulators. Each of them might sample the targets in ways that cannot be explained by a mere seed match. However, identification of Ago signature on targets via CLIP does not imply functionality of the site. In fact, binding is not always a sign of regulation, as seen for transcription factors. Even though as many as thousands of binding sites can be identified via ChIP (Chromatin and Immuno-Precipitation) experiments, only a tenth of them seems to be actually regulated (Slattery et al., 2014). Further studies are thus needed to clarify how a miRNA finds its targets.

1.3 The miRNA repertoire of *Caenorabditis elegans*

Historically, *Caenorabditis elegans* has been "the" model organism to study miRNA biology. In fact, the very first two miRNAs: *lineage-4* (*lin-4*) and *lethal-7* (*let-7*) were discovered in worms (Lee et al., 1993; Reinhart et al., 2000). Apart from historical reasons, the nematode is often the model of choice for its simple handling and the relative ease of performing genetic screens. Notably, at least for some miRNAs, mutant phenotypes are obvious and therefore easy to work with. Because most miRNA and their interactors are conserved to higher eukaryotes, the worm is an ideal model to study how miRNAs shape development under physiological conditions.

1.3.1 *C. elegans* and the heterochronic pathway

C. elegans is a one-millimeter long nematode that normally lives in soil and eats bacteria. It can grow in Petri dishes or liquid medium in the lab (Jorgensen and Mango, 2002). The worms can be maintained as hermaphrodites, making it easy to expand a clonal population. Males can spontaneously arise or be induced and are essential to transfer alleles among strains. Furthermore, the worm is transparent and ideal for microscopy analysis, features that allowed John Sulston to follow and map the fate of each single cell (Sulston and Horvitz, 1977; Sulston et al., 1983). An adult hermaphrodite has 959 somatic nuclei and has all the basic tissues: neurons, muscles, skin, gonads and gut. The worm reproduces with a short life cycle of about 3 days under optimal conditions. Its development comprises an embryonic phase and four larval stages (L1- L4), which last about 8 hours, and are separated by a lethargus phase, in which the animal stops feeding and molts, shedding the old cuticle that will be replaced by a new one. By the end of the L4 stage, the animal matures the sexual organs, transitions to the adult stage and stops molting. Additionally, in the absence of food or in stressful conditions the worm enters an alternative L3 developmental stage, the long-lived *dauer*.

From hatching to adulthood, the worm undergoes a series of finely controlled cell divisions that are invariant and that could be mapped (Sulston and Horvitz, 1977). Thanks to the constant cell lineage, Ambros and Horvitz were able to identify some of the first mutant strains in which the timing of cell division was altered and caused visible changes in morphology (Ambros and Horvitz, 1984; Chalfie et al., 1981). These *heterochronic* mutants are characterized by developmental events that either occur "earlier" (precocious events) or "later" (retarded events) relative to other invariant and normally timed developmental events (Ambros and Horvitz, 1984).

Temporally altered events occur in different tissues, such as the neurons or the vulva, but best studied are the ones in the epidermis. This tissue, traditionally referred to as the "hypodermis", even if of ectodermal origin, covers almost the whole body of the worm and comprises mainly two cell types, the hypodermis and specialized epithelial cells, which include the seam cells. The hypodermis consists of the hyp7 syncytium and some other cells in the head and tail. Seam cells are stem-like cells that, together with hypodermis, are involved in formation of stage-specific cuticle. Seam cells alone can generate the alae, which are cuticular ridges typical of L1 or adult stages only. At hatching there are 10 seam cells per side (H0-H2, V1-V6, and T), but with every molt all the cells (except one, H0) divide asymmetrically. The cells of the V lineage undergo one round of symmetrical division in L1 that doubles their number (except for V5). Then, once at every stage from L2 to L4, they asymmetrically divide into another seam cell and an anterior daughter cell that fuses with hyp7. When the worm becomes sexually competent, at the larval-to-adult transition (L/A switch), there are 16 seam cells per side that stop dividing, exit cells cycle and differentiate, fuse with each other in a syncytium and secrete adult alae.

Mutations in heterochronic genes can affect seam cell terminal differentiation; in particular, they can either go through premature or delayed development. Loss-of-function mutants of the *lin-4* gene for example are "retarded": they maintain a juvenile state and reiterate early fate developmental programs late in development with seam cells that never undergo asymmetrical cell division typical of L2 stage or do not produce adult alae at the L4-adult transition (Chalfie et al., 1981; Lee et al., 1993). On the contrary, *lin-28* loss-of-function mutants are "precocious": their seam cells skip the L2 division and produce adult-alae when the worm is still not sexually competent in L3 stage (Ambros and Horvitz, 1984; Moss et al., 1997).

1.3.2 The *let-7* family of miRNAs

Some of the core components of the heterochronic pathway are miRNAs, with *lin-4* controlling L1-to-L2 transition, miR-48/miR-241/miR-84 regulating the L2-to-L3 and *let-7* being responsible for development from L4 to adulthood (Rougvie, 2001).

Interestingly, the four heterochronic miRNAs *let-7*, miR-48, miR-241, miR-84 share the same seed sequence and are thus grouped in the so-called *let-7* family, which is only partially redundant. In fact, only *let-7* is essential and even if an animal triple mutant for miR-48/miR-241/miR-84 (miR-

48/241/84 (-)) is sick and poorly viable, single mutation in those miRNA genes have low penetrant developmental defects (Abbott et al., 2005).

The *let-7* null allele is lethal and the animals die bursting through the vulva at the L4-to-adult transition (Reinhart et al., 2000). However, another allele *let-7(n2853)* (henceforth called *let-7ts*) carries a point mutation (G-A) in the fifth nucleotide of the seed sequence and is a temperature sensitive allele, which allows survival at the permissive temperature 15°C, but behaves like a null at the restrictive 25 °C (Reinhart et al., 2000). Close analysis of worms carrying this loss-of-function allele revealed that they have supernumerary seam cells and absent alae due to an extra division at the L4 stage. Animals undergo a fifth larval stage and when they reach adulthood, seam cells finally differentiate and produce alae (Reinhart et al., 2000; Rougvie, 2001). Conversely, overexpression of *let-7* resulted in precocious seam cell differentiation at the L3-L4 transition. However, the effect of such experiments can be difficult to interpret because different transgenic animals over-express the miRNA at different levels, due to the intrinsic stochasticity of the methods used to achieve such over-expression that involves multi-copy arrays, sometimes randomly integrated in the genome, whose copy number is unknown.

The four *let-7* family members are differentially expressed: *let-7* accumulates late in development, with particularly high abundance in larval stage 4 (L4) and adult animals, whereas the three sisters accumulate already in L2 (Abbott et al., 2005; Esquela-Kerscher et al., 2005; Reinhart et al., 2000; Vadla et al., 2012) and remain high throughout development. This different temporal pattern of expression can partially explain why the *let-7* sisters are needed to specify entry into the L3 stage. MiR-48/241/84 triple mutant animals show a retarded phenotype characterized by supernumerary seam cells caused by reiteration of L2-to-L3 division (Abbott et al., 2005). Furthermore, miR-48/241/84 (-) animals also have incomplete alae and burst through the vulva or poorly survive due to molting defects. Because only the miR-48/241/84 (-) has such a strong phenotype that is reminiscent of a *let-7* loss-of-function mutant, the miR-48/241/84 are probably redundant with each other, but only partially redundant with *let-7*.

Besides the different temporal expression pattern of these four miRNAs, the exact reason why the four miRNAs are not completely redundant is not fully understood. In fact, given that they have the same seed sequence, they should repress the same mRNA targets. However, only *let-7* is lethal and necessary for the *lin-41* mRNA repression (Ecsedi et al., 2015; Reinhart et al., 2000).

Lin-41 is an essential gene for fertility that produces an RNA binding protein involved in the heterochronic pathway at the time of the larval-to-adult transition (Slack et al., 2000). It has a key role in switching off developmental programs associated with adult development, and it achieves it by translationally repressing the transcription factor LIN-29 (Ambros and Horvitz, 1984; Rougvie and Ambros, 1995), and by inducing the mRNA degradation of its cofactor MAB-10 (Aeschimann et al., 2017). Accordingly, *lin-41* loss-of-function mutants exhibit precocious phenotype, characterized by premature onset of adulthood with early seam cell differentiation and alae secretion, though with incomplete penetrance. These mutants also show precocious LIN-29 accumulation in the skin, further confirming a precocious onset of adult developmental programs. Conversely, *lin-41* overexpression causes a phenotype similar to *let-7* loss of function, with vulval bursting and extra seam cell division (Slack et al., 2000).

To achieve timely *lin-41* downregulation by *let-7* only, which is necessary to direct proper vulva development and transition to adulthood (Ecsedi et al., 2015), there must be a mechanism in place that avoids early *lin-41* repression by the *let-7* sisters that accumulate a stage earlier.

Let-7 binds the *lin-41* 3'UTR at two sites, LCS1 and LCS2 (*let-7* Complementary Site) (Vella et al., 2004a; Vella et al., 2004b). Different from canonical miRNA sites, which are usually characterized by perfect seed/seed match pairing (Bartel, 2009), the two sites have two imperfect seed matches (harboring an A bulge and a G: U wobble pair, respectively) and extensive 3' complementarity to *let-7* only. Therefore, it was speculated that those sites harbor an imperfect seed match to ensure specific *let-7* binding and avoidance of the *let-7* sisters. In fact, binding of the sisters only through the seed/seed match would result in a weak and unstable duplex that would not engage in silencing, while *let-7* compensates for the imperfect seed match with additional distal pairing (Bartel, 2009; Brennecke et al., 2005; Doench and Sharp, 2004). Therefore, it has been postulated that the imperfect seed pairing in the *lin-41* 3'UTR avoids redundant activity of the *let-7* sisters (miR-48, miR-84, miR-241) which would lead to precocious *lin-41* downregulation that would in turn trigger premature adult development. However, this has never been experimentally validated.

Surprisingly, two of the *let-7* paralogues have been found to rescue a *let-7* lesion. In both cases the paralogue was exogenously supplemented through multi-copy arrays that possibly over-expressed the miRNA at levels far from physiological (Hayes et al., 2006; Li et al., 2005). Even though this suggests that some miRNAs can compensate for each other, it is not clear whether this happens through silencing of *lin-41* or other targets. In fact, other genes that induce precocious phenotype, like *hbl-1*,

can rescue the *let-7* mutation (Abrahante et al., 2003; Lin et al., 2003). Furthermore, the artificially high level of the miRNA questions the relevance of such results.

1.3.3 Transcriptional and post-transcriptional control of the *let-7* family

The *let-7* family undergoes transcriptional and post-transcriptional regulation that ends up in a very complicated network of factors that influence the expression of these miRNAs.

HBL-1 is a transcription factor homolog of the *Drosophila* hunchback (hb) (Fay et al., 1999). Because its downregulation induces precocious seam cell differentiation, it is a heterochronic gene. Phenotypic observations are based on hypomorphic alleles or post-embryonically provided RNAi, as the null is embryonic lethal (Fay et al., 1999). Surprisingly, even if some seam cells differentiate already in L3, they do not fully exit the cell cycle because their nuclei divide again during adulthood resulting in a syncytium with more than 16 nuclei (Abrahante et al., 2003; Lin et al., 2003). Furthermore, *hbl-1* loss of function alleles induce protruding vulva phenotype, caused by precocious vulva cell development in L2 (Lin et al., 2003).

Hbl-1 was shown to be target of the *let-7* sisters in the hypodermis and of *let-7* in the ventral nerve cord, but the exact binding sites are still unknown (Abbott et al., 2005). In turn, HBL-1 transcriptionally represses *let-7* in a tissue specific manner (Roush and Slack, 2009). Thus, *let-7* represses *hbl-1* in the ventral nerve cord, but *hbl-1* represses *let-7* in the seam cells to confine its expression to later stages; meanwhile the sisters repress *hbl-1* in the hyp7 (Roush and Slack, 2009). However, *hbl-1* seems to be epistatic to *let-7,* although it was shown to disappear from the epidermis before the L3 stage and to be important for the L2-to-L3 transition in the seam cells (Abrahante et al., 2003; Lin et al., 2003). Additionally, *hbl-1* function seems to be semi-redundant to that of *lin-41* in determining the L/A switch through LIN-29 as suggested by the fact that *lin-29* is epistatic to both genes (Abrahante et al., 2003; Lin et al., 2003). Currently, the exact expression pattern, function and regulation of this transcription factor are mysterious.

More transcription factors seem involved in the *let-7* family regulation, such as ELT-1, LIN-42, BLMP-1 and DAF-12. ELT-1 was shown to stimulate expression in seam cells (Cohen et al., 2015); LIN-42 might be involved in repression of miRNA transcription, as suggested by increased pri-*let-7* levels in *lin-4*2 mutants (McCulloch and Rougvie, 2014; Perales et al., 2014; Van Wynsberghe et al., 2014). BLMP-1 is important for terminal differentiation of hypodermis, as *blmp-1* null mutants fail to secrete

alae (Horn et al., 2014) and DAF-12, which seems to downregulate the *let-7* sisters, rescues *let-7* mutation, while also being a target of the *let-7* family (Bethke et al., 2009; Grosshans et al., 2005; Hammell et al., 2009).

Additionally, it has been shown that *let-7* can stimulate its own processing through Argonaute (Grishok et al., 2001). Indeed, in both *C. elegans* and human cells, *let-7*-loaded-Argonaute binds the pri-*let-7* transcript in a conserved element downstream of the miRNA sequence and induces processing into the mature form (Zisoulis et al., 2012). Furthermore, the RNA binding protein LIN-28 has been found to prevent maturation of *let-7* through direct binding of its precursors (Heo et al., 2009; Newman et al., 2008; Viswanathan et al., 2008). In *C. elegans* this interaction is expected to avoid precocious accumulation of *let-7* (Stefani et al., 2015; Van Wynsberghe et al., 2011), the misregulation of which would eventually result in premature adult development, as seen in *lin-28* null animals (Ambros and Horvitz, 1984). In mammals, where LIN-28 binding to *let-7* pre-miRNA was first identified, LIN-28 was proposed to promote reprogramming of somatic cells to an embryonic state (Balzeau et al., 2017; Viswanathan et al., 2008).

1.3.4 Other miRNA families in *C. elegans*

The *let-7* family is not the only miRNA family in *C. elegans* and in fact about 60% of worm miRNAs belong to a family and a third of those is conserved across species (Lim et al., 2003a; Lim et al., 2003b; Miska et al., 2007). With very few exceptions, mutations in single nematode's miRNA genes do not result in altered development or viability. Conversely, mutations affecting more than one single member of a miRNA family result in synthetic phenotypes (Abbott et al., 2005; Alvarez-Saavedra and Horvitz, 2010; Miska et al., 2007; Sherrard et al., 2017).

The *lin-4* miRNA family comprises *lin-4* and *miR-237. lin-4* mutations lead to strong heterochronic phenotypes (Chalfie et al., 1981; Lee et al., 1993), it is required to down-regulate *lin-14* to induce the larval stage 1 to larval stage 2 (L1-to-L2) transition and *lin-28* to induce the L2-to-L3; accordingly, worms lacking this miRNA reiterate L1 specific cell division patterns (Resnick et al., 2010). On the contrary, mir-237 seems to enhance or suppress heterochronic phenotypes and does not show obvious defects when mutated singularly (Miska et al., 2007; Tsialikas et al., 2017).

Another family with clear role in worm development is the miR-35 family. As opposed to most miRNAs, the eight members of this family are maternally provided and simultaneous deletion of all of

them leads to embryonic lethality (Alvarez-Saavedra and Horvitz, 2010). One of its function, recently characterized, seems to prevent activation of male developmental programs (McJunkin and Ambros, 2017). Interestingly, miR-35 family has another function for which it cooperates with another miRNA family, the miR-58, to avoid precocious apoptosis of cells during development (Sherrard et al., 2017). The miR-58 family comprises six members and it is related to the *bantam* miRNA in *Drosophila*, which is involved in cell proliferation and survival (Brennecke et al., 2003; Hipfner et al., 2002). This group of miRNAs functions redundantly and shows a cooperative behavior, as suggested by stronger upregulation of their targets when an increasing number of miRNAs is deleted (Subasic et al., 2015).

For many other families a function has not yet been described (Miska et al., 2007), but this could be due to our inability to detect subtle phenotypes. In fact, some mutant worms that look superficially wild type revealed their essential role only when using dedicated assays, as reported for lsy-6, which controls the expression of chemosensory receptors that establishes the left/right asymmetry of the gustatory ASE neurons, and miR-791, which is involved in CO_2 sensing (Drexel et al., 2016; Johnston and Hobert, 2003). Thus, it is plausible to imagine that most miRNA mutants are superficially wild type, and thus considered not-essential, only because we have not yet identified their specific function. This could be the case for the mir-72 family, whose mutations do not induce any obvious phenotype (Miska et al., 2007); although its conservation hints to a function, (this family is related to miR-73 in *C. briggsae*, which separated from *C. elegans* 80-100 million years Ago, and the miR-31 family in humans).

1.3.5 The *let-7* miRNA in higher eukaryotes

Understanding miRNA family biology, targets and regulation of each member is not only helpful to understand miRNA targeting in general, but might have therapeutic potential, too. In fact, at least a third of *C. elegans* miRNA families is conserved across species (Miska et al., 2007). For example, the *let-7* family is conserved in humans and often downregulated in human cancers (Roush, 2008). Except for the fruit fly that has only one *let-7*, usually several *let-7* isoforms exist: in worms there are seven members including *let-7* itself (although only miR-48, miR-241 and miR-84 are well characterized and discussed in this thesis), and eleven in zebrafish. In mammals, there are ten miRNAs sharing the exact *let-7* seed sequence and they are usually identified by a letter. *Let-7a* is 100% conserved to the worm and drosophila's *let-7* proper, while the other isoforms have slightly divergent sequences.

The evolutionary conservation of each of the 22 nucleotides of *let-7* is difficult to explain. One explanation could be that because *let-7* has to bind several targets, its sequence cannot vary, as this would change the target repertoire. Hence, paired regions would remain conserved to maintain invariant the miRNA repressive activity. However, in worms, although several targets can be predicted *in silico* and might be important for yet to discover functions, *lin-41* seems to be the only *let-7* target necessary for animal viability (Ecsedi et al., 2015). If paired regions are conserved to ensure broad target recognition, regions that are unpaired and form loops or bulges would be conserved as they might function as recognition pads for RNA binding proteins (RBPs) or other molecules that modulate silencing (Ruvkun, 2008). Currently, none of these has been shown.

Similarly to *let-7* involvement in stem cell differentiation (i.e. seam cells) in *C. elegans*, *let-7* roles in flies neuromuscular junction formation or mammalian limb development, together with more general role in cell proliferation and development have been described (Roush, 2008). Generally, *let-7* levels increase as cells differentiate. Lower levels of *let-7* isoforms have been described in cancerous cells. In contrast, their expression was reported during differentiation of breast stem cell progenitor (Roush, 2008; Rupaimoole and Slack, 2017). *Let-7* miRNAs can target oncogenes, such as *kras* that is often mutated in cancers (Johnson et al., 2005; Ratner et al., 2010). In addition to the sequences, the genomic location and post-transcriptional regulation of the *let-7* family is similar in different animals (Roush, 2008). For example the LIN-28/*let-7* axis, originally found in mammalian cells and later in worms (Newman et al., 2008; Van Wynsberghe et al., 2011; Viswanathan et al., 2008), stands out in showing the conserved role of such miRNA in differentiation and development. Because LIN-28 post-transcriptionally represses *let-7* and its pro-differentiation activity, *lin-28* mutant worms prematurely become adults (Ambros and Horvitz, 1984; Vadla et al., 2012). Accordingly, LIN-28 is upregulated in cancer cells where *let-7* levels are low (Roush, 2008) and *lin-28* has been described as one of the genes that has to be upregulated to reprogram cells to an undifferentiated state, possibly through its repressive activity on *let-7* (Balzeau et al., 2017).

1.4 Tissue-specific miRNA profiling in *C. elegans*

Even if transcriptional reporters give us an idea about temporal and spatial distribution of miRNAs, it should be kept in mind that this is only a proxy. First, cloning of promoter sequences carries the inherent risk that some regulatory elements might be missed, thus resulting in imprecise expression pattern detection. Second, miRNAs, and in particular the *let-7* family, undergo a tightly

regulated biogenesis and we must take into account the effect of post-transcriptional control on mature miRNA levels. For instance, it has been recently shown that *let-7* biogenesis is controlled by LIN-28 (Stefani et al., 2015; Van Wynsberghe et al., 2011) and *let-7* itself (Zisoulis et al., 2012).

Molecular biology techniques such as northern blots, RT-qPCRs or small RNA sequencing have the potential to complement transcriptional reporters and enrich our understanding of miRNA temporal expression patterns. They are usually performed on whole worm lysates (Bracht et al., 2004; Esquela-Kerscher et al., 2005; Li et al., 2005; Vadla et al., 2012), as primary cells isolation is extremely limited in *C. elegans* and not widely applicable (Zhang, 2013). Efforts in optimizing *in situ* hybridization for miRNAs in worms confirmed the expression of some miRNAs, including the four *let-7* members, as inferred through transcriptional reporters. However, the method lacks the resolution needed to reveal cell specific expression (Andachi and Kohara, 2016).

Due to the fact that the cuticle makes dissociation of the worm and collection of intact cells challenging, tissue-specific data have been mostly limited to analysis of transcriptional reporters (Hope lab databases, GFPfusion database and (Martinez et al., 2008)) or obtained through computational analysis of several microarrays performed in different conditions (Kim et al., 2001). Some diverse protocols to obtain tissue-specific data bypassing the inability to obtain viable cell suspensions have been described. One of them is based on the culturing of embryonic cells that differentiate in neurons, which are then successfully FACS sorted (Fox et al., 2005). Embryonic cells are easier to obtain than larval cells, but it is currently unknown how to differentiate them in other tissues, thus limiting applicability of this technique (Tintori et al., 2016). Alternatively, microarray or sequencing analysis of nuclear material was reported for FACS sorted intestinal nuclei (Haenni et al., 2012), while nuclei originating from the muscles have been obtained by the Henikoff lab by applying to adult worms the INTACT method. INTATC involves tissue specific biotinylation of nuclei that are then purified through streptavidin beads (Steiner and Henikoff, 2015; Steiner et al., 2012). However, these techniques are limited to the recovery of only nuclear mRNA species and are not suitable for miRNAs.

Pull-down of a tissue-specifically expressed poly-A binding protein has been reported for several tissues and is based on the assumption that tissue-specific transcript co-precipitate with the protein because they are bound *in vivo* in the tissue of interest. This technique, which was originally referred to as “mRNA tagging”, was initially coupled to microarray and used to obtain gene expression data for muscles, seam cells, gut and neurons (Gorrepati et al., 2013; Pauli et al., 2006; Roy et al., 2002; Takayama et al., 2010). Later it has been coupled to high-throughput RNA sequencing leading to

analysis of gene expression profiles from several tissues in different stages of development (Blazie, 2015; Blazie et al., 2017; Spencer et al., 2011). However, the mRNA tagging technique tends to be characterized by a high background signal and can suffer from artifacts due to the lysis procedure. In addition, like the nuclei-based approaches, it does not allow recovery of miRNAs. Some FACS-based techniques based on (Zhang, 2013) have been described, but only with limited application, as shown by the profiling of specific larval neurons (Spencer et al., 2014). Only recently, a single-cell RNA sequencing experiment that profiled 27 individual worms tissue was described (Cao et al., 2017).

In conclusion, although some protocols to obtain tissue-specific data have been developed, the field is still at its infancy. Some cell-type specific transcriptomes have been described, but the methods used are not well suited for miRNA profiling. Therefore, a protocol that allows investigation of the actual level of the mature and active form of miRNAs in specific tissues is still missing.

2 Results

Against a "seed-centric" view for miRNA targeting, previous data obtained in the lab showed that the miRNAs of the *let-7* family can repress different targets in different tissues (Ecsedi, 2015). In particular, a GFP reporter with *lin-41* 3'UTR, although generally specific to *let-7* (Ecsedi, 2015; Ecsedi et al., 2015) revealed that *lin-41* is silenced by different members in different tissues. While the *let-7* sisters had barely any role in repressing it in the vulva, they were required in the hypodermis, as the *lin-41* reporter was partially de-silenced in *miR-48/241/84 (-)* animals. Distinct results were found for other reporters containing the *hbl-1* 3'UTR, or to a lesser extent for *daf-12* 3'UTR, which seemed specific to the *let-7* sisters. The data thus suggested that in the same tissues *let-7* and the sisters can have different targets and show different extent of repression.

To understand the mechanisms that in different tissues determine the target specificity observed for the *let-7* family, we aimed to test what we originally named the *quality or quantity hypothesis,* i.e. whether specificity is an intrinsic property (based on sequence complementarity between a miRNA and a target) or if it is due to the context. We hypothesized that different levels of mature miRNAs in different tissues might modulate specificity, likely with the involvement of unknown tissue-specific factors. Therefore, we envisioned a scenario in which complementarity between a miRNA and a target is important to establish silencing, but in which the cellular context can nevertheless modulate the interactions that take place.

The results on the "sequence determinants of specificity" are described in the form of a manuscript (Brancati and Grosshans, 2018) in the next chapter 2. 1, whereas the results on the "tissue-specificity" are described in chapter 2.2.

In addition, we optimized a worm dissociation and cell sorting approach, which we named FACSeq (FACS sorting and sequencing of *C. elegans* cells). The analysis of the transcriptomes of different cell types, the miRNA profile of seam cells and a single-cell sequencing experiment are described in the chapter 2.3.

2.1 Publication: An interplay of miRNA abundance and target site architecture determines miRNA activity and specificity

Brancati, G. and Grosshans, H. (2018).
An interplay of miRNA abundance and target site architecture determines miRNA activity and specificity
Nucleic Acids Res 46, 3259-3269
https://www.ncbi.nlm.nih.gov/pubmed/29897601

(NAR Breakthrough Article;
Recommended Article on F1000Prime: doi:10.3410/f.733028808.793544522 and
doi:10.3410/f.733028808.793546212)

Published online 21 March 2018

Nucleic Acids Research, 2018, Vol. 46, No. 7 ***3259–3269***
doi: 10.1093/nar/gky201

NAR Breakthrough Article

An interplay of miRNA abundance and target site architecture determines miRNA activity and specificity

Giovanna Brancati[1,2] and Helge Großhans[1,2,*]

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland and [2]University of Basel, Basel, Switzerland

Received January 15, 2018; Revised March 02, 2018; Editorial Decision March 06, 2018; Accepted March 13, 2018

ABSTRACT

MicroRNAs often occur in families whose members share an identical 5′ terminal 'seed' sequence. The seed is a major determinant of miRNA activity, and family members are thought to act redundantly on target mRNAs with perfect seed matches, i.e. sequences complementary to the seed. However, recently sequences outside the seed were reported to promote silencing by individual miRNA family members. Here, we examine this concept and the importance of miRNA specificity for the robustness of developmental gene control. Using the *let-7* miRNA family in *Caenorhabditis elegans*, we find that seed match imperfections can increase specificity by requiring extensive pairing outside the miRNA seed region for efficient silencing and that such specificity is needed for faithful worm development. In addition, for some target site architectures, elevated miRNA levels can compensate for a lack of complementarity outside the seed. Thus, some target sites require higher miRNA concentration for silencing than others, contrasting with a traditional binary distinction between functional and non-functional sites. We conclude that changing miRNA concentrations can alter cellular miRNA target repertoires. This diversifies possible biological outcomes of miRNA-mediated gene regulation and stresses the importance of target validation under physiological conditions to understand miRNA functions *in vivo*.

INTRODUCTION

MicroRNAs (miRNAs) are small RNAs of about 22 nucleotides that silence target messenger RNAs by binding to partially complementary sequences in their 3′ untranslated regions (3′UTRs). miRNAs are loaded onto an Argonaute (Ago) protein to form the core of the miRNA-induced silencing complex (miRISC), which induces decay or translational repression of the targets (1). Conceptually, miRNAs can be separated into two parts: the 'seed', comprising nucleotides two through eight, and the 'seed-distal' 3′ end (Figure 1A). The seed sequence has emerged as the main determinant for target identification (2). Usually, functional miRNA targets contain 'seed matches', heptamers that base pair with perfect Watson-Crick complementarity to the miRNA seed. These were found to be necessary and sufficient for silencing in studies using ectopic miRNA expression (3–5). Structural and biochemical analyses of miRISC have provided an explanation for these results: the seed of a miRNA bound by Ago exists in a pre-arranged conformation, thus reducing the entropic cost of binding and favoring duplex formation with a target (6–8).

miRNAs frequently occur in families that share the seed sequence but differ in the seed-distal part. Given the reliance of target silencing on seed matches, it is assumed that miRNA family members can function redundantly, and most computational approaches that predict miRNA targets make predictions for miRNA families rather than for individual miRNAs (2). Consequently, it was hypothesized that in order to attain specificity among family members, miRNAs require imperfect seed matches. In this scenario, an imperfect seed match impairs binding and activity of most family members, but extensive seed-distal base pairing would enable a specific family member to compensate for the unfavorable seed binding (3).

However, high-throughput biochemical capture of Ago-bound miRNA/target duplexes revealed numerous instances of interactions that frequently extended beyond the seed, to involve the seed-distal parts of the miRNA (9–12). In cell culture and *in vivo* assays, some of the targets that could base pair through their seed-distal parts were silenced preferentially by specific family members (9,12). Because such specificity also occurred for target sites with perfect seed matches, these findings argued that seed match im-

*To whom correspondence should be addressed. Tel: +41 61 6976675; Fax: +41 61 6973976; Email: helge.grosshans@fmi.ch

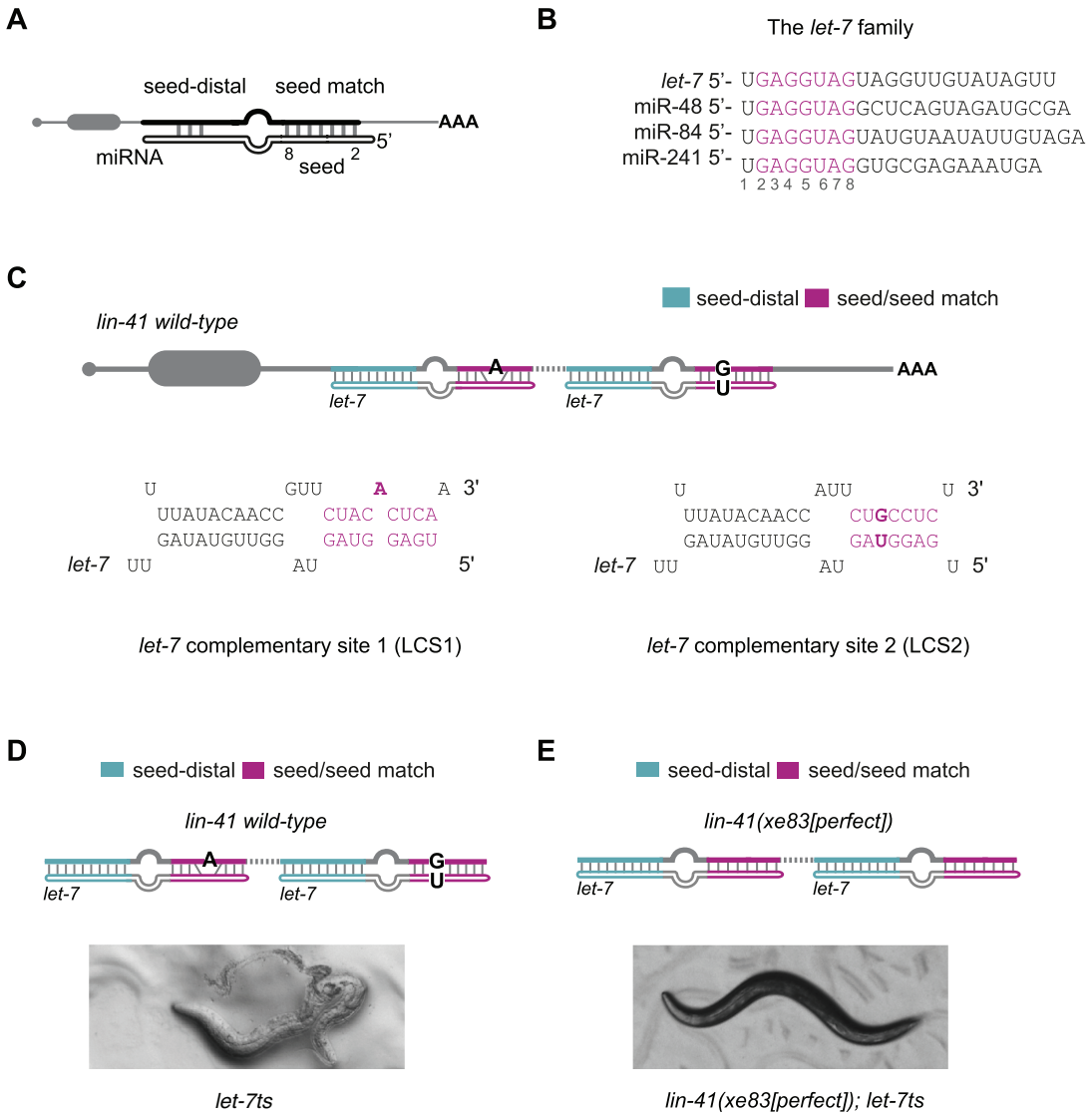


Figure 1. *let-7* becomes dispensable for viability when the *lin-41* 3′UTR contains perfect seed match sites. (**A**) Schematic drawing of a miRNA/target duplex with seed (nucleotides 2–8)/seed match and limited seed-distal pairing indicated. Top mRNA, bottom miRNA. (**B**) The *let-7* family with the seed sequence (nucleotides 2–8) highlighted in magenta. (**C**) The two *let-7* complementary sites (LCS1 and LCS2) in the *lin-41* 3′UTR of *C. elegans.* Each site contains an imperfect seed match (a bulged A and a G: U wobble, respectively, in bold) to the *let-7* family and an extensive seed-distal pairing to *let-7* only. The sites are separated by 27 nt of intervening sequence (dashed line). (**D**, **E**) Representative images of animals carrying the *let-7ts* mutation and (D) wild-type *lin-41* or (E) the *lin-41(xe83[perfect])* allele with perfect seed match to the *let-7* family and unchanged seed-distal region. Animals were grown at 25°C. *let-7ts: let-7(n2853) X,* temperature sensitive lesion. miRNA site legend: magenta = seed/seed match; cyan = *let-7* seed-distal binding.

perfections might not be a requirement for miRNA family member specificity.

By contrast, specificity of miRNA silencing through seed mismatches would explain why members of the *let-7* family of *Caenorhabditis elegans* have partially non-redundant functions. Indeed, among four members with overlapping expression patterns (13), *let-7*, miR-48, miR-84 and miR-241 (Figure 1B), only *let-7* is essential for viability (14,15). *let-7* ensures proper development of *C. elegans* by repressing one crucial target, *lin-41* (16,17), whose 3′UTR contains two functional *let-7* binding sites (*let-7* complementary sites, LCSs) (18). Both LCSs contain imperfect seed-matches, which yield a bulged-out nucleotide and a G:U wobble base-pair respectively (Figure 1C). Moreover, both LCSs exhibit extensive complementarity to the seed-distal sequence of *let-7* but to none of its sisters. Here, we test if this miRNA site architecture ensures specific silencing by *let-7* and explore miRNA site architectures as a mechanism for the selectivity of different family members towards distinct targets.

We show that extensive seed-distal pairing favors miRNA silencing by an individual miRNA family member even when the seed match is perfect, but that an imperfect seed match greatly enhances this family member specificity. Thus, we find that perturbing *let-7*-specific regulation of *lin-41,* by introducing a perfect seed match, impairs normal *C. elegans* development through allowing the *let-7* family sisters miR-48, miR-241 and miR-84 to prematurely silence

lin-41. Moreover, specificity of targets with perfect or nearly perfect seed matches can be overcome through elevated levels of a miRNA that is incapable of seed-distal pairing. Hence, although sequence-instructed, specificity is not fully hard-wired and can be altered by changes in miRNA expression levels.

Our observations are consistent with a model where *let-7* family miRNAs act as rheostats (19), such that the interplay of target site architecture and miRNA abundance determine the extent of target silencing. This flexible targeting mechanism expands the regulatory potential of miRNA families and indicates that miRNA activity may differ on *bona fide* targets at a given miRNA concentration. Conversely, alterations in miRNA concentrations may then change the miRNA target repertoire, expanding the range of possible biological outcomes, and revealing a need for target validation under physiological conditions to understand miRNA function *in vivo*.

MATERIALS AND METHODS

Worm handling and strains

Worms were grown using standard methods at 25°C. The transgenic *unc-54 + miRNA sites* reporter strains were obtained by single-copy integration into the *ttTi5605* locus on chromosome II (20). Injected plasmids were cloned using the MultiSite Gateway Technology (Thermo Fisher Scientific) and the destination vector pCFJ150 (21) or Gibson assembly (22). All strains are listed in Supplementary Table S1.

unc-54 + miRNA sites reporters

All *unc-54 + miRNA sites* reporters were constructed using the MultiSite Gateway Technology (Thermo Fisher Scientific) and the destination vector pCFJ150 (21) or Gibson assembly (22). First, the pGB0 vector was obtained via site-directed mutagenesis (23) of the pDONR P2R-P3_p37 vector to insert the AscI restriction site. Then, the pGB01 plasmid was obtained via LR reaction (Gateway LR Clonase II Enzyme mix, Thermo Fisher Scientific; 11791020) of the three entry vectors pdpy-30 x pGFP::H2B x pGB0 and the pCFJ150 backbone.

All the plasmids listed in Supplementary Table S2 were obtained via Gibson assembly of the digested pGB01 plasmid and gBlocks® Gene Fragments (Integrated DNA Technologies) listed below. All plasmids were verified by sequencing. Transgenic worms were obtained by single-copy integration into the *ttTi5605* locus on chromosome II, following the published protocol for injection with low DNA concentration (20). We optimized our previous mCherry reference transgene (16) by replacing the artificial 3′UTR with an endogenous *unc-54* 3′ UTR, to achieve more physiologic and brighter expression. The resulting *Pdpy-30::mCherry::H2B::unc-54* transgene was integrated on chromosome I to yield strain HW1454.

Genome editing

Mutations in the endogenous *lin-41* 3′UTR sequence were obtained by CRISPR-Cas9 to generate the *lin-41(xe83[perfect])*, *lin-41(xe76[ap427_W-C])*, and *lin-41(xe99[48-ized])* alleles. Wild-type worms were injected as described in (24) with a mix containing 50 ng/μl pIK155, 100 ng/μl of each pGB48 and plin-41sgRNA, 20 ng/μl repair oligo (see Supplementary Table S4), *dpy-10* cocrispr mix containing 100 ng/ml pIK208 (Addgene plasmid #65630) and 20 ng/ml AF-ZF-827 oligo PAGE purified (IDT). Single F1 roller progeny of injected wild-type worms were picked to individual plates and the F2 progeny screened for the mutated allele using PCR assays and sequencing (Supplementary Table S3). The alleles were outcrossed three times to the wild-type strain.

let-7 over-expression. A *let-7(++)* strain (HW 1909 [*xeSi287*, V]) was obtained by injection of the plasmid pGB26, obtained via Gibson assembly of the PCR amplified minimal rescue fragment from (15) and the pIK37 plasmid. Transgenic worms were obtained by single-copy integration into the *oxTi365* locus on chromosome V (universal MosSCI strain #EG8082 (25).

Reporter quantification

For confocal assays, worms were grown at 25°C. *Let-7ts* worms were maintained at 15°C and adults were transferred to 25°C for 48 h before imaging. Z-stacks of 0.313 μm μm thickness were acquired in green, red and transmitted light channels at 40× magnification on a Zeiss LSM700 confocal microscope coupled to Zeiss Zen 2010 software equipped with a multi-position tile scan macro. The z-stacks were stitched together and compiled into a single image using scripts in Matlab and Fiji (26).

For data analysis, late L4 worms were selected based on visual inspection of gonad length and vulva morphology (27). Ten to fourteen vulva cells were selected in the 'cell counter' macro in Fiji. Images around these seed points were de-noised using a Richardson-Lucy algorithm and segmented using an Otsu global threshold. Remaining holes were filled using a morphological filter. Signal intensity in the green channel was divided by the red signal intensity for each cell; relative signal intensities were then averaged for each worm. 10–12 vulva cells in 5–10 worms per genotype were quantified, mean signal intensity and SD were calculated and graphed using GraphPad Prism software.

Confocal analysis of LIN-29 precocious accumulation

Synchronized arrested L1 larvae of animals carrying endogenously tagged LIN-29, *lin-29(xe61[lin-29::gfp::3xflag])* (28), in wild-type or *lin-41(xe83[perfect])* background, were plated on food and incubated at 25°C on 2% NGM agar plates with *Escherichia coli* OP50 bacteria and imaged at the L3 stage (20–22 h after plating). Images were acquired in green and transmitted light channels (with Differential Interference Contrast, DIC) with 40×/1.3 oil immersion objective on a Zeiss LSM700 confocal microscope coupled to Zeiss Zen 2010 software. Further image processing was performed with Fiji (26).

RESULTS

Perfect seed matches in the *lin-41* 3′ UTR make *let-7* miRNA dispensable for animal viability

Specific regulation of *lin-41* by *let-7* and not by its sisters was previously speculated (3) to derive from the imperfect seed-matches in the two *let-7* miRNA Complementary Sites (LCS1 and LCS2) in the *lin-41* 3′UTR (Figure 1C (18,29)). When bound by *let-7* family miRNAs, the seed match sequences of LCS1 and LCS2 generate an A-bulge and a G:U wobble pair. Both sites contain seed-distal complementarity to *let-7*, but not to its sisters. However, Broughton and colleagues recently identified a target site in the 3′UTR of *dot-1.1* that appeared specific to the *let-7* family member miR-48 in the absence of seed match imperfections (9). Given this unexpected finding, we tested the possibility that seed mismatches in LCS1 and LCS2 were similarly dispensable for specific recognition by *let-7*. To this end, we generated a *lin-41* allele, *lin-41(xe83[perfect])*, which differs from the wild-type allele in two nucleotides: We eliminated the A bulge in LCS1 and converted the G:U wobble pair of LCS2 into a standard Watson-Crick base pair.

Strikingly, these two nucleotide changes rescued the larval lethality caused by loss of *let-7*, both in the *let-7(mn112)* null mutant strain and the *let-7(n2853)* temperature-sensitive strain (henceforth *let-7ts*), which recapitulates the *let-7* null phenotype at the restrictive temperature, 25°C ((15), Figure 1D and E). Thus, ≥98% ($N = 3$, each with $n \geq 200$ animals) of *lin-41(xe83[perfect])*; *let-7ts* double mutant animals survived into adulthood, as did 100% ($N = 2$, $n \geq 98$ animals) of *lin-41(xe83[perfect])*; *let-7(mn112)* double mutant animals, of which 6% subsequently died as adults. These findings suggest that seed mismatches are required to restrict silencing of *lin-41* to *let-7*, because other *let-7* family members confer silencing in their absence.

A perfect seed match allows redundant activity of the *let-7* sisters

To confirm that the perfect seed matches of the *lin-41(xe83[perfect])* allele allow redundant binding of the *let-7* family, we monitored the activity of the four miRNAs through a GFP reporter modified from (16) (Materials and Methods). In our assay, each animal contains a red mCherry reporter, which is used as reference during image analysis, and a GFP reporter, which is the miRNA activity sensor (Figure 2A). Both reporters are driven by the ubiquitous and constitutively active *dpy-30* promoter and contain the *unc-54* 3′UTR, generally thought to be devoid of regulatory elements. Finally, each reporter is integrated by Mos1-mediated single copy integration into a distinct genomic location (20).

To monitor *let-7* activity, we generated the reporter '*unc-54 + let-7 sites*' in which only a stretch of 111 nucleotides of the *lin-41* 3′UTR, comprising LCS1 and LCS2, was transplanted into the *unc-54* 3′UTR (Figure 2A). Silencing of this minimal target reporter by *let-7* was comparable to that of a reporter containing the full-lenght *lin-41* 3′UTR (Figure 2B, C and Supplementary Figure S1A), confirming functionality. We focused our analysis on the vulva because *lin-41* repression by *let-7* in this organ is required and likely sufficient to prevent vulval rupturing (16).

As expected, the '*unc-54 + let-7* sites' reporter was expressed in young L1 or L2 animals (Supplementary Figure S1B), when the *let-7* family levels are low (30). Moreover, it was robustly silenced in older, L4-stage larvae, when *let-7* family levels are high (Figure 2B and C). Finally, it was de-silenced in *let-7ts* animals, but not in animals lacking the three *let-7* sisters (*[mir-48/mir-241(ndf51)V, mir-84(n4037)X],* henceforth *mir-48/241/84(−)*) (Figure 2C). Therefore, the stretch of 111 nucleotides suffices for efficient and specific *let-7*-dependent silencing.

Next, we generated an '*unc-54 + let-7 sites_perfect seed match*' reporter, modified to contain LCSs with perfect seed matches, as in the endogenous *lin-41(xe83[perfect])* mutation (Figure 1E). Like the '*unc-54 + let-7 sites*' reporter, the new reporter was expressed in young L1 or L2 animals (Supplementary Figure S1B), but robustly silenced in L4-stage larvae (Figure 2D). However, unlike the '*unc-54 + let-7 sites*' reporter, the new reporter was only marginally de-repressed in L4-stage larvae lacking *let-7* (*let-7ts*) or the three *let-7* sisters (*mir-48/241/84(−)*) (Figure 2D).

A seed-distal match establishes specificity to one miRNA in the presence of an imperfect seed match

Taken together, the genetic interaction and the reporter assay data presented thus far validate the hypothesis that the seed mismatches in the *let-7* complementarity sites of *lin-41* are necessary for specific regulation of *lin-41* by *let-7*, to the apparent exclusion of the other family members. However, this conclusion appears at odds with the results of biochemical miRNA–mRNA duplex identification, which indicate preferential target binding by individual family members even in the presence of perfect seed matches (9,12). Thus, to challenge our finding, we sought to reprogram the LCSs to another *let-7* family member, miR-48, and test the effect of seed match imperfections. We chose miR-48 because its expression levels and spatial expression patterns appear very similar to those of *let-7* (13,14,31).

Because structural data suggest that base pairing between nucleotides 13–16 of the miRNA and a target may be favored (8), we started out by generating a reporter with seed-distal base pairing to only these nucleotides. However, this reporter failed to be silenced even in wild-type conditions, i.e. with both *let-7* and miR-48 present (Figure 3A and Supplementary Figure S2A). Hence, it appears that more extensive seed distal complementarity is required for functionality of targets with a sub-optimal seed match. Indeed, an '*unc-54 + miR-48 sites*' reporter that emulated the LCS architecture by carrying a central bulge in the seed sequence and an extensive seed distal match to miR-48 (Supplementary Figure S2B), was silenced in L4 stage animals. Moreover, and in agreement with our predictions, the '*unc-54 + miR-48 sites*' reporter was repressed at the L4 stage in both the presence and absence of *let-7* miRNA, but became de-repressed when miR-48 was absent (Figure 3B).

Consistent with our results for the *let-7* reporters, the specificity of the '*unc-54 + miR-48 sites*' reporter was largely lost when we modified it to contain perfect seed matches: the resulting '*unc-54 + miR-48 sites_perfect seed match*' reporter

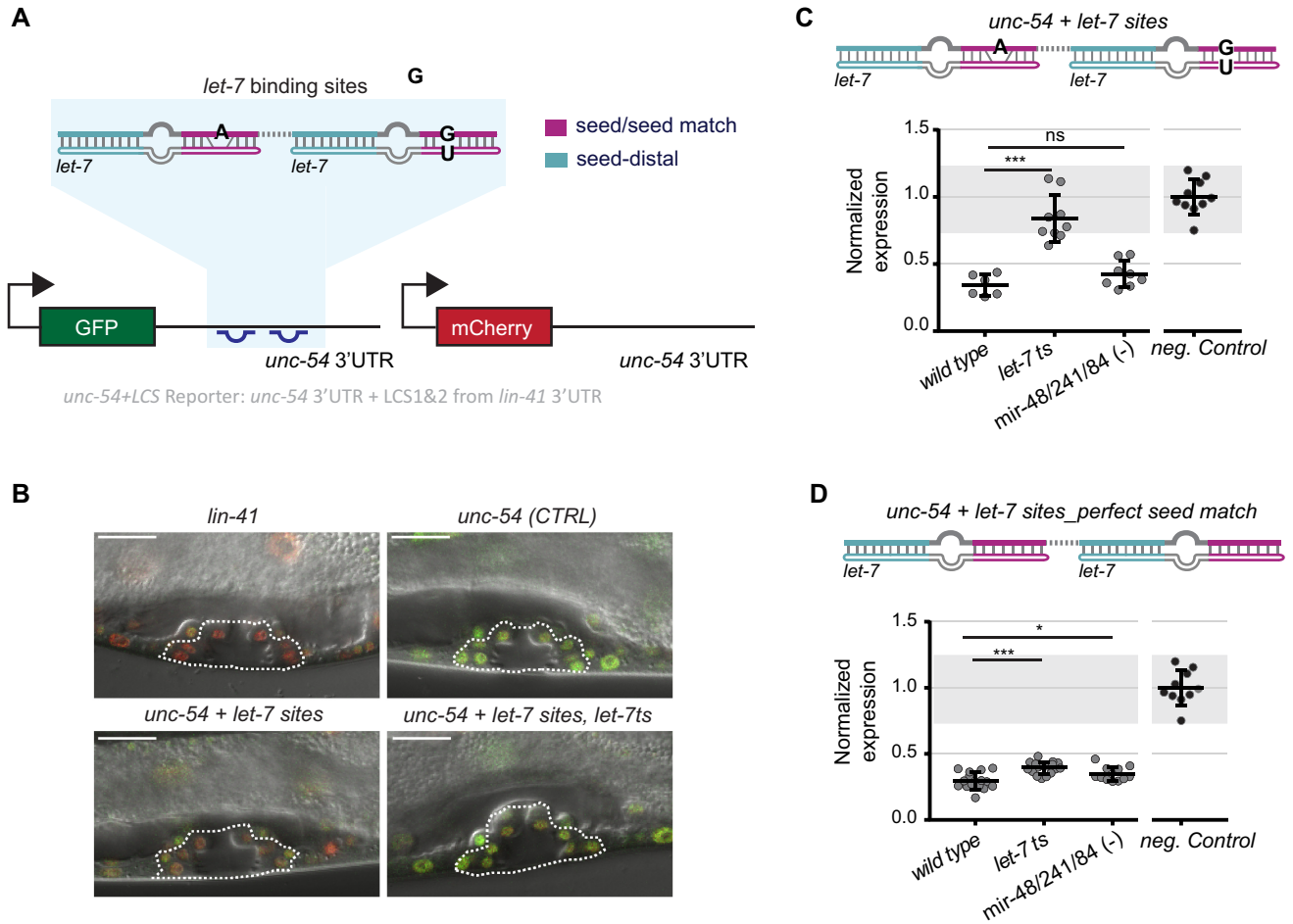


Figure 2. Redundant activity of the *let-7* family in the presence of a perfect seed match. (**A**) Schematic of the reporters used to monitor miRNA activity *in vivo.* The depicted GFP transgene *unc-54 + let-7 sites* reporter contains 111 nucleotides of the *lin-41* 3′UTR (shaded in blue), which harbor the two *let-7* binding sites and the 27 nt-long intervening sequence, grafted into the heterologous, unregulated *unc-54* 3′UTR. Worms also contain a red mCherry reporter for normalization. Transcription of the single-copy integrated reporters from the ubiquitously active *dpy-30* promoter is constitutive. miRNA site legend: magenta = seed/seed match; cyan = *let-7* seed-distal binding. (**B**) Representative confocal images of the vulvae of animals carrying the red mCherry reporter (for normalization) and GFP reporters with the indicated 3′UTRs. These are '*lin-41* 3′UTR full-length', '*unc-54'* (CTRL, unregulated) and '*unc-54 + let-7 sites*' in wild-type and '*unc-54 + let-7 sites*' in the *let-7ts* background. Images are merged GFP, mCherry and DIC channels. Red color indicates a greater, and green color a lesser degree of reporter repression. Dashed lines outline the vulvae of the animals, which confirm appropriate late Larval stage 4 (L4). Scale bars 15 μm. (C, D) Quantification of (**C**) '*unc-54 + let-7 sites'* reporter, (**D**) '*unc-54 + let-7 sites_perfect seed match'* reporter. Each dot represents the average of the GFP signal intensity, obtained by confocal imaging, divided by the mCherry intensity for a single animal per condition. 10–12 vulva cells were quantified per worm. Mean values are normalized to the average value of the GFP/mCherry ratio of the negative control *unc-54* 3′UTR reporter, which is not silenced. Horizontal line and error bars indicate mean values per condition ± SD. $*P < 0.05$ and $***P < 0.001$, two-tailed unpaired t-test. For reference, data obtained for the *unc-54, Neg.Control* reporter are replotted in panel D; gray shading is bounded by the min-max values of this control.

continued to be silenced extensively in both *let7ts* and *mir-48/241/84(−)* animals (Figure 3C). However, silencing appeared marginally impaired in the absence of the *let-7* sisters (Figure 3C), mirroring an analogous result for the '*unc-54 + let-7sites_perfect seed match*' reporter in *let-7ts* animals (Figure 2D). We conclude that the imperfect seed match and the extensive 3′ pairing are both important determinants for the robust target specificity of the *lin-41* sites.

A G:U wobble base-pair in a peripheral seed match location promotes miRNA specificity

The duplexes formed between *let-7* and *lin-41* contain a bulge between nucleotides 4–5 in LCS1 and a G:U wobble base-pair at position 6 in LCS2 (Figure 1C). We wondered if such centrally located 'imperfections' were required for specificity. We turned to the miRNA binding site in the *dot-1.1* 3′UTR, which had been shown to be specific to miR-48 (9). Broughton *et al.* found that substitution of the *let-7* complementary sites in the endogenous *lin-41* 3′UTR by two copies of the *dot-1.1* site rendered animals insensitive to loss of *let-7* (9), but made them depend on the presence of miR-48. This finding was attributed to the fact that the site features an extensive seed-distal match to miR-48 (Figure 3D and Supplementary Figure S2C). However, we noticed that the *let-7* family/*dot-1.1* predicted duplexes exhibited not only perfect Watson–Crick pairing from nucleotides 2–7, but also a G:U wobble pair at position 8 (Supplementary Figure S2C). Although hexameric seed match

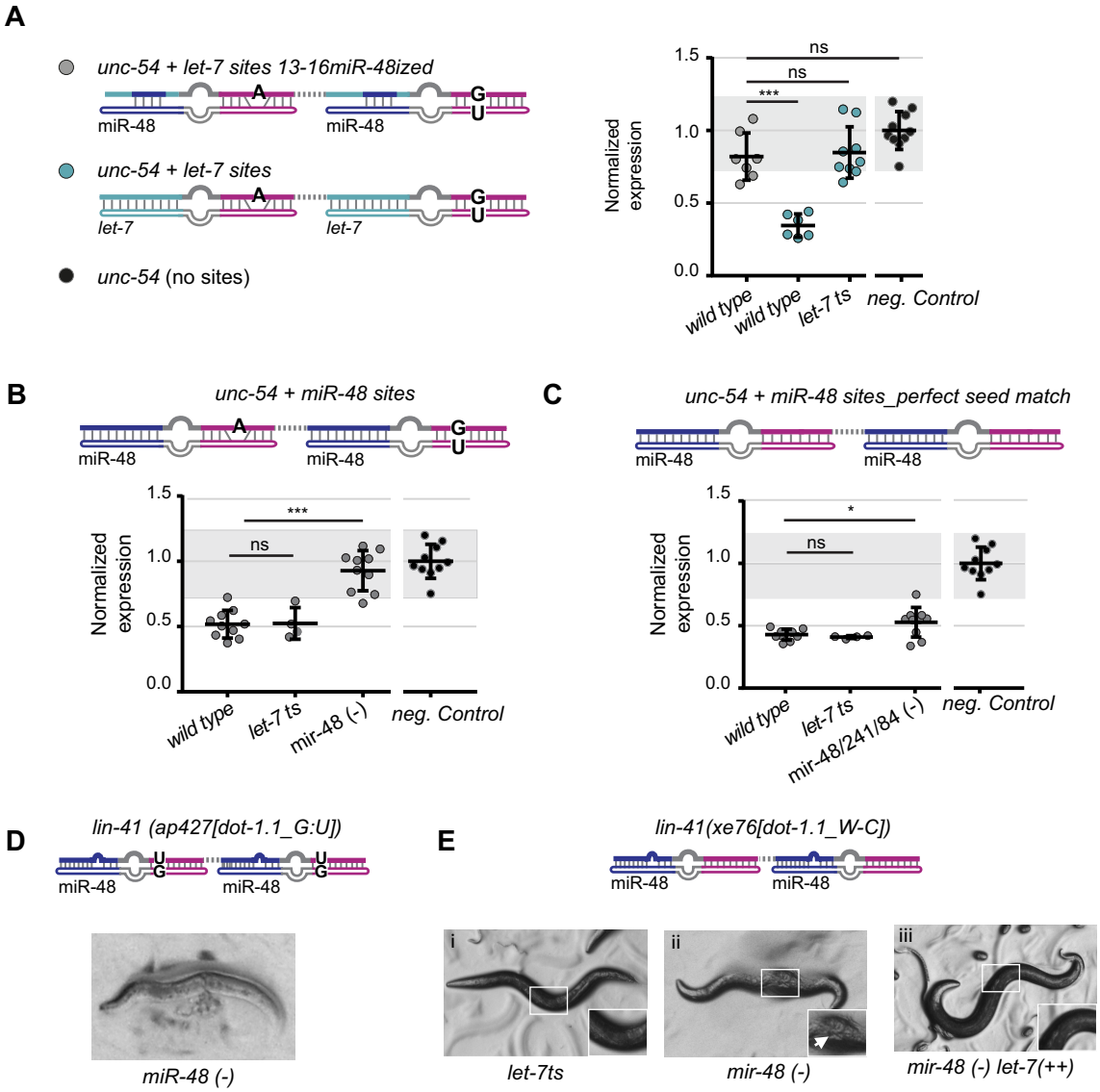


Figure 3. Imperfect seed matches and extensive 3′ pairing confer target specificity. (A–C) Reporter quantification as in Figure 2, from which the negative control data (black dots) are also replotted for reference; gray shading is bounded by the min-max values of the negative control. (**A**) The *'unc-54 + let-7 sites 13–16miR-48-ized'* reporter contains *let-7* complementary sites modified to pair miR-48 at position 13–16 but not other seed-distal nucleotides (*gray dots*). Results from the unmodified *'unc-54 + let-7 sites'* reporter in wild-type and *let-7ts* mutant background are from Figure 2C and included for reference (*cyan dots*). (**B**) The *'unc-54 + miR-48 sites'* reporter combines extensive seed-distal complementarity to miR-48 with seed match imperfections whereas (**C**) the '*unc-54 + miR-48 sites_perfect seed match'* reporter contains extensive seed-distal complementarity to miR-48 and perfect seed matches. Horizontal line and error bars indicate mean values per condition ± SD. $*P < 0.05$ and $***P < 0.001$, two-tailed unpaired *t*-test. (**D**) Animals carrying the *lin-41(ap427[dot-1.1_G: U]* (9)) allele die in the absence of miR-48. (**E**) Survival of strain *lin-41(xe76[dot-1.1_W-C]* upon manipulation of *let-7* and miR-48 activity. In this strain, a U at position 8 in the two target sites of the *lin-41(ap427[dot-1.1_G:U])* allele has been converted to a C, to permit Watson-Crick instead of G:U wobble base-pairing with the *let-7* family seed sequence (Supplementary Figure S2C and D). This allele was crossed into a (i) *let-7ts*, (ii) *mir-48(–)* or (iii) *mir-48(–) let-7(++)* background, where *let-7(++)* denotes *let-7* overexpression from a single copy integrated transgene. Insets magnify the central part of the animal body to reveal egg retention (arrow), i.e. and egg-laying defective (Egl) phenotype. *let-7ts: let-7(n2853) X, temperature-sensitive* lesion, grown at the restrictive temperature 25°C*; mir-48(–): mir-48(n4097) V; mir-48/241/84(–): mir-48/mir-241(ndf51) V, mir-84(n4037) X.*

sites, with complementarity to nucleotides 2–7, are considered canonical and functional (2), genome-wide studies also suggested that they are less functional than heptameric sites that match nucleotides 2–8 (6,32,33). Since G:U wobble base pairs elsewhere in seed-seed match duplexes appear detrimental to silencing (3,4,34–36), we wondered if this 'peripheral G:U' in seed match position 8 might affect silencing and specificity.

To test this hypothesis, we modified the endogenous target sites in *lin-41* to those of *dot-1.1*, but with the G:U wobbles at positions 8 converted to Watson-Crick G:C pairs, yielding allele *lin-41(xe76[dot-1.1_W-C])* (Supplementary Figure S2D). We then compared the reliance of this and the *lin-41(ap427[dot-1.1_G:U])* strain, which carried the unmodified G:U-wobble-containing *dot-1.1* sites, on *let-7* and miR-48 for survival. Whereas both strains

were insensitive to loss of *let-7* (Figure 3E(i) and (9)), *lin-41(ap427[dot-1.1_G:U])* but not *lin-41(xe76[dot-1.1_W-C])* required miR-48 for survival into adulthood (Figure 3D and E(ii)). We conclude that the G:U wobble at position eight repels binding by all *let-7* family members such that only miR-48 can exert repression by compensating through extensive complementarity of its 3′ seed-distal sequence. Collectively, our data thus reveal that bulges or wobbles in different positions of a seed match can serve to avoid redundancy of the *let-7* family and confer strong target specificity.

miRNA abundance affects silencing *in vivo*

Although our experiments provided strong evidence that seed mismatches are required for robust specificity among *let-7* family members, we consistently observed evidence of residual specificity even for targets that contained a perfect seed match. In target reporters containing perfect seed matches, we observed modest but reproducible de-silencing specifically when the family member with seed-distal match was lost (Figures 2D and 3C), and phenotype (Figure 3E(ii)). In fact, although *lin-41(xe76[dot-1.1_W-C]); mir-48(–)* animals survived into adulthood, they exhibited an egg-laying (Egl) defect (Figure 3E (ii), 93%, $n = 132$), i.e. a partial vulval dysfunction that is consistent with incomplete repression of *lin-41* (16).

We wondered if this partial specificity could be overridden by increased levels of another miRNA family member. Since we were unable to overexpress *mir-48*, we tested this possibility by overexpressing *let-7*. *Mos1*-mediated single copy integration (25) of a genomic fragment, known to rescue *let-7* lethality (15), to a locus on chromosome V that is ∼5 cM apart from *mir-48*, yielded a ∼2-fold increase in expression levels (data not shown). Consistent with our hypothesis, *lin-41(xe76[dot-1.1_W-C])* animals that overexpressed *let-7* were no longer Egl in the absence of miR-48 (Figure 3E(iii), compare to E(ii)). We conclude that, *in vivo,* increased miRNA levels can override the specificity imparted by seed-distal pairing.

Seed match imperfections maintain specificity upon miRNA overexpression

Since the modest preferential silencing imposed by the seed-distal pairing to miR-48 could be overcome by increasing the levels of *let-7* in the presence of a perfect seed match (Figure 3E (ii) and (iii)), we wondered about the effect of *let-7* over-expression on sites with more extensive target specificity. Hence, we examined two reporters specific to miR-48 that harbored imperfect seed matches: the previous '*unc-54 + miR-48 sites*' (Figures 3B and 4A) and the new '*unc-54 + dot-1.1 sites*' reporter, obtained by inserting two copies of the binding sites from the *dot-1.1* 3′UTR (Figure 4B). Consistent with the *in vivo* data ((9) and Figure 3E), silencing of both reporters was dependent on miR-48 but not *let-7* (Figures 3B, 4A, B and Supplementary Figure S2E). However, the response of the two reporters differed when we overexpressed *let-7* in the absence of miR-48. The '*unc-54 + miR-48 sites*' reporter, with central seed mismatches, was insensitive to a doubling of *let-7* expression (Figure 4A). By contrast, silencing of the '*unc-54 + dot-1.1 sites*' reporter, with peripheral seed mismatches, was restored to almost wild-type level in the same conditions (Figure 4B). This suggests that for miR-48 targets with extensive seed-distal pairing, sensitivity to *let-7* levels depends on seed match quality.

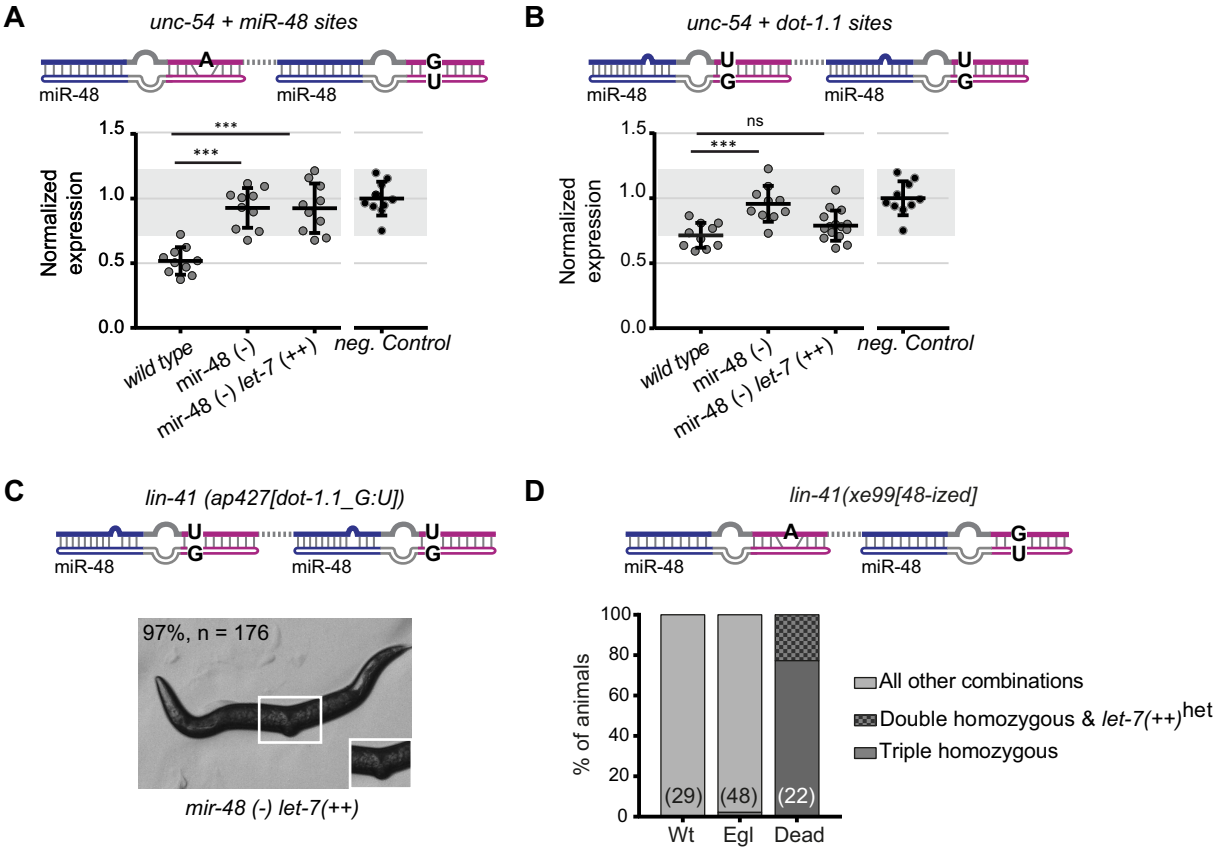
To confirm this result on a functional level, we tested whether *let-7* overexpression could suppress the dependence on miR-48 of animals carrying *lin-41* alleles analogous to those in the miR-48-specific reporters, namely the *lin-41(ap427[dot-1.1_G:U])* allele and the newly generated *lin-41(xe99[48-ized])* allele (Figure 4C and D, respectively). As predicted by the reporter assay, overexpression of *let-7* rendered *lin-41(ap427[dot-1.1_G:U])*; *mir-48(–)* double mutant animals viable, although Egl (Figure 4C). By contrast, we were unable to obtain viable animals of the *lin-41(xe99[48-ized]) I; mir-48(–) let-7 (++) V* genotype (Figure 4D). Instead, we readily observed dead animals, which had burst through the vulva. Genotyping revealed that such animals were homozygous for the three alleles of interest, *lin-41(xe99[48-ized]), mir-48(–)*, and *let-7(++)* (Figure 4D). [Note that *mir-48(–)* and *let-7(++)* are closely linked loci on chromosome V, explaining why we did not find dead animals that were *lin-41(xe99[48-ized]); mir-48(–)* double mutant but lacked the *let-7* over-expression transgene.] In contrast, randomly selected wild-type animals were never doubly homozygous for *lin-41(xe99[48-ized])* and *mir-48(–)*, irrespective of *let-7* transgene status, and only one Egl animal was found to be *lin-41(xe99[48-ized]); mir-48(–) let-7(++)* mutant. Hence, although an increase in *let-7* levels can overcome the specificity to miR-48 imposed by seed-distal matches in combination with a perfect seed (Figure 3E) or in the presence of peripheral seed mismatches (Figure 4C), it cannot do so with a central seed bulge or wobble (Figure 4D), at least within the physiological ranges of the expression levels that we tested.

We conclude that specificity arises through seed-distal pairing of a miRNA, but that it is enhanced in extent and robustness by appropriate seed match architecture (Figure 4E).

Loss of miRNA specificity impairs robust development

Our results suggest that sites with central seed match imperfections, such as LCS1 and LCS2 in the *lin-41* 3′UTR, are extremely specific to one miRNA, even when a paralogue is highly expressed. We suspected that such robust specificity would be physiologically relevant in the case of *lin-41*. This is because the *let-7* sisters are all expressed prior to *let-7*, in the L2 stage (30). Given their overlapping spatial expression patterns, lack of mechanisms to prevent *let-7* sisters' action on *lin-41* might cause inappropriately early repression of *lin-41*, as speculated previously (2,3). Consistent with this notion, we found that the '*unc-54 + let-7 sites_perfect seed match*' reporter was precociously repressed during the L3 stage, whereas the '*unc-54 + let-7 sites*' reporter was still expressed at the same stage (Figure 5A).

To test whether this precocious repression of *lin-41* had physiological consequences, we examined the accumulation of LIN-29A, a target of LIN-41. In wild-type animals, LIN-41 translationally represses LIN-29A until the L4 stage, when repression is released following *let-7* accumulation and consequent LIN-41 downregulation (28). Premature



E

Seed match	*lin-41* allele	*mir-48(-)*	*mir-48(-) let-7 (++)*	Specificity
central imperfection	*lin-41(xe99[48-ized])*	Dead	Dead	
peripheral imperfection (position 8)	*lin-41(ap427[dot-1.1_G:U])*	Dead	Egl	
perfect	*lin-41(ap427[dot-1.1_W-C])*	Egl	Wt	

Figure 4. Robust miRNA specificity relies on imperfect seed matches. (A, B) Reporter quantification as in Figure 2, from which the negative control data are also replotted for reference. (**A**) '*unc-54 + miR-48 sites*' reporter and (**B**) '*unc-54 + dot-1.1 sites*' reporter are assayed in worms of the indicated genotypes. Horizontal line and error bars indicate mean values per condition ± SD, $*P < 0.05$ and $***P < 0.001$, two-tailed unpaired *t*-test. (**C**) Representative image of a viable *lin-41(ap427[dot-1.1_G:U]), mir-48(-) let-7(++)* animal. (**D**) Progeny ($n = 99$) derived from a cross of *lin-41(xe99[48-ized])* with *mir-48(–) let-7(++)* animals were categorized by phenotype and genotyped to determine the viability of *lin-41(xe99[48-ized]); mir-48(-) let-7(++)* 'triple homozygous' mutant animals. (**E**) Summary of the effect that different site architectures and miRNA abundance have on silencing *lin-41* alleles 'recoded' towards miR-48. *mir-48(–): mir-48(n4097)V; unc-54(CTRL):* wild-type *unc-54 3′ UTR; let-7(++): let-7* over-expression allele (MosSCI, V).

loss of LIN-41 activity causes inappropriately early activation of LIN-29A and thereby precocious execution of the so-called larval-to-adult transition, which includes fusion of hypodermal seam cells into a syncytium and secretion of an adult cuticular structure termed alae (17). We observed LIN-29A levels through use of a *lin-29(xe61[lin-29::gfp::3xflag])* strain, in which the endogenous *lin-29* locus has been edited to produce GFP-tagged LIN-29A and B isoforms, and in which loss of *lin-41* activity yields a specific upregulation of only LIN-29A (28). At mid-L3 larval stage, wild-type animals have LIN-29::GFP signal only in their seam cells (Figure 5B). By contrast, animals carrying the *lin-41(xe83[perfect])* allele show additional GFP expression in the major hypodermal syncytium, hyp7, at the same developmental stage (Figure 5B). Therefore, precocious downregulation of *lin-41(xe83[perfect])* is responsible for premature LIN-29 translation and accumulation

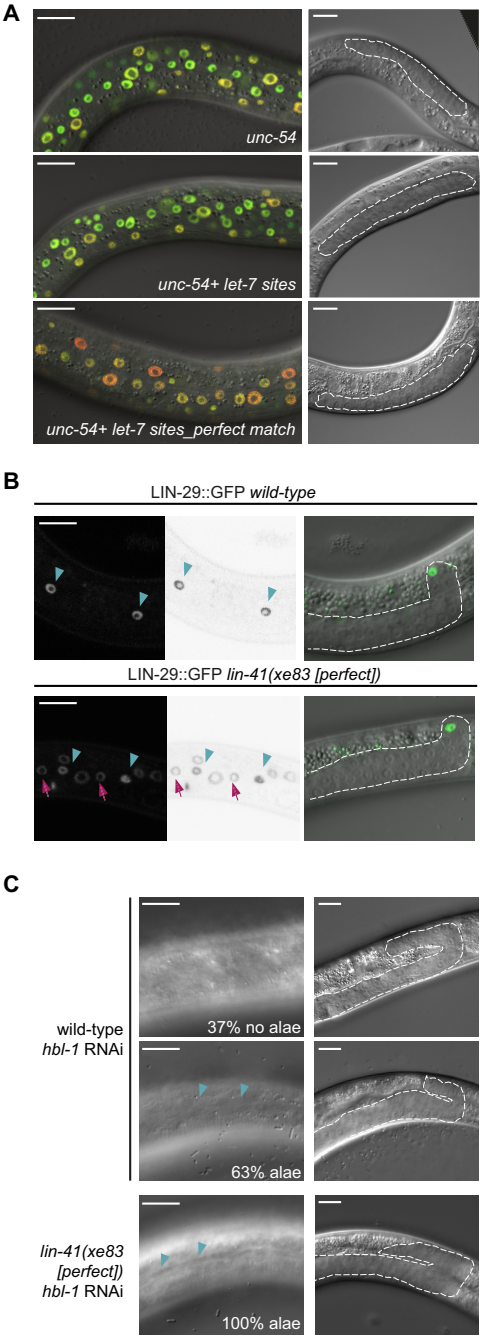


Figure 5. Developmental robustness requires an imperfect *let-7* seed match in *lin-41*. (**A**) Representative confocal images of skin cells of animals carrying an *unc-54* 3′UTR reporter (top), an *'unc-54 + let-7 sites'* (center), or an *'unc-54 + let-7 sites_perfect seed match'* reporter (bottom). At the L3 stage, levels of miR-48 but not *let-7* are already high (39). Scale bars 15 μm. (**B**) Microscopy images of the skin of late L3 worms expressing endogenously tagged LIN-29::GFP (*xe61*) (28) in wild-type and *lin-41(xe83[perfect])* background. Cyan arrowheads point to LIN-29 signal in seam cells, magenta arrows to LIN-29 accumulation in hyp7 cells. Images in the middle are inverted to increase clarity. Worms are staged according to the position of the distal tip cell (green) and gonad length. Scale bars 15μm. (**C**) Representative images of wild-type ($n = 27$) or *lin-41(xe83[perfect])* ($n = 36$) animals treated with *hbl-1* RNAi. Percentages of animals with the indicated alae status at the L3/L4 transition are indicated. Gonads are outlined to confirm appropriate staging. The strains used, SX346 and HW2144, additionally contain the *mjIs15* and *wIs51* transgenes. Scale bars 15μm.

in the hypodermis, as described for other *lin-41* loss-of-function alleles (17).

The *lin-41(xe83[perfect])* animals looked superficially wild-type, but the premature upregulation of LIN-29 was sufficient to promote precocious larval-to-adult transition in a sensitized background. Specifically, the transcription factor HBL-1 inhibits larval-to-adult transition, possibly in parallel to LIN-41 (37,38), and its RNAi-mediated depletion causes partially penetrant and partially expressive precocious alae formation (Figure 5C). This phenotype was enhanced when we depleted HBL-1 in *lin-41(xe83[perfect])* mutant animals, resulting in fully penetrant precocious secretion of alae (although weak or patched in some cases) (Figure 5C). We conclude that loss of specificity of repression by *let-7* alone in the *lin-41(xe83[perfect])* background impairs the robustness of temporal patterning through premature LIN-29 accumulation.

DISCUSSION

It has been an open question to what extent and by which mechanisms miRNA family members can function non-redundantly despite a shared seed sequence. Previously, it was proposed that redundancy was the rule (2). Rare occasions of non-redundant function were hypothesized to require targets with both an imperfect seed match and extensive seed-distal pairing to only one specific family member (3). According to this view, the seed match imperfection impairs silencing by all family members but extensive seed-distal pairing can compensate to facilitate silencing by an individual miRNA. However, this hypothesis has remained untested, and recent observations have challenged it by providing evidence that non-redundant target binding appears wide-spread and that seed-distal pairing may suffice to achieve specificity (9,12).

Our systematic study through gene editing and fluorescent reporter analysis with cell-type resolution resolves the discrepant views on specificity-promoting features for the *let-7* family: We demonstrate that extensive seed-distal pairing to a specific family member suffices to generate a weak but consistent preference for silencing by this family member. However, more robust discrimination requires an imperfect seed match and depends on the quality of such imperfections: a central bulge or G:U wobble base pair, as in the *lin-41* 3′UTR, confers the strongest specificity, while a peripheral G:U wobble base pair, as in the *dot-1.1* 3′UTR, gives an intermediate level. The physiological importance of extensive, seed-mismatch-dependent specificity is evident from the decreased developmental robustness that results when perfect *let-7* seed matches permit promiscuous silencing of *lin-41* by the whole *let-7* family.

Perfect seed matches can still be compatible with selective targeting by individual miRNAs, but the effect depends on miRNA abundance: A moderate increase in *let-7* levels (~2-fold) could overcome the specificity of a binding site that was silenced by miR-48 and had a perfect seed match. However, it only partially did so when the seed match contained a peripheral G:U wobble, and it was insufficient to override sequence-determined specificity when a site contained a central seed match imperfection. This suggests that *in vivo*, miRNA binding site architecture, particularly seed

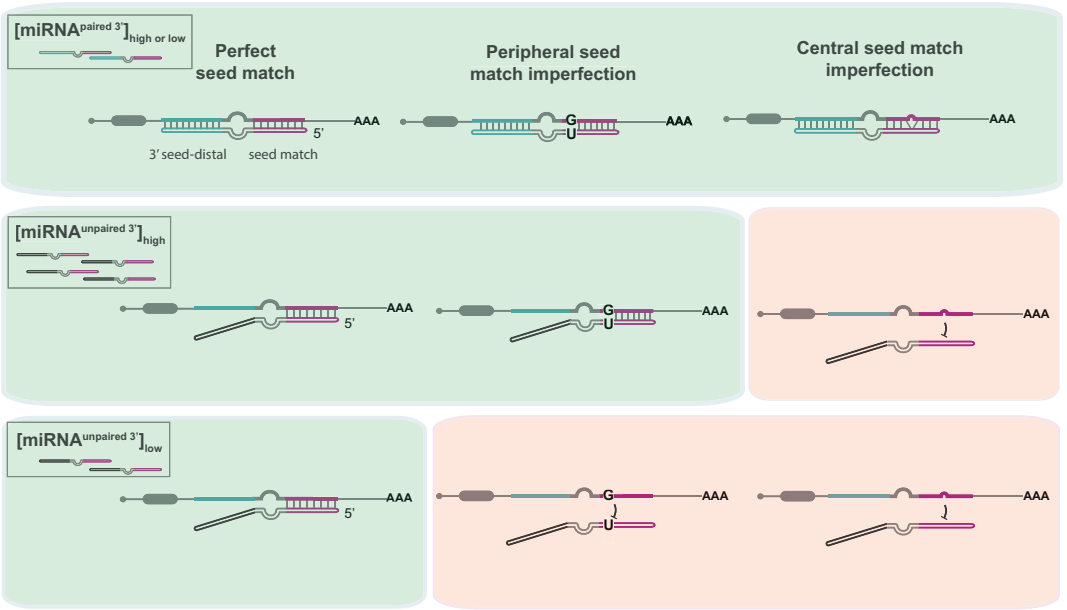


Figure 6. miRNA abundance and architecture of the target site determine mRNA silencing. (top) Extensive complementarity (paired 3′) between a miRNA and a target site allows for efficient and specific silencing, independently of the miRNA level and the presence of imperfections in the seed match. (middle) Abundant miRNAs can silence targets carrying a perfect seed match or a nearly-perfect seed match (e.g. a peripheral G:U wobble), even in the absence of complementarity to the sequence outside the seed. A 'central mismatch' repels poorly complementary miRNAs. (bottom) Lowly abundant and poorly complementary miRNAs can silence targets carrying a perfect seed match, but not the ones carrying a seed match imperfection (e.g. peripheral G:U or central bulge). Green shading: functional site; pink shading: nonfunctional site. Magenta: seed/seed match

match quality, and miRNAs abundance act together to determine miRNA activity towards individual targets (Figure 6).

The finding that miRNA activity is determined at the level of individual targets has implications beyond the issue of miRNA family member specificity. It contrasts with a view where a miRNA is globally either 'on' or 'off' in a cell, silencing all of its targets at sufficiently high concentrations and none at low ones. Variable, target site-dependent activity was already entertained in the early days of the miRNA field when miRNAs were likened to rheostats, whose activity is adjusted by two features, namely the extent of target site complementarity to the miRNA and miRNA abundance (19). A lack of explicit experimental testing of such context-dependent function (4) and the rising popularity of the 'seed-match only' model caused this hypothesis to fade from view. We propose that it is time to revisit the idea of miRNAs functioning as rheostats and subject it to further testing.

We note that target validation experiments that rely, as often done, on ectopic miRNA expression appear to make the implicit assumption that miRNAs are uniformly active on their targets. However, if the goal of target validation is to provide insights into pathway biology, physiology and/or pathology, our results and those of others (34) strongly suggest that it must be conducted in a relevant physiological context, avoiding ectopic expression or overexpression of miRNAs. Ideally, validation will also involve functional studies such as those offered by direct manipulation of individual miRNA/target interaction through genome editing. We predict that such efforts will reveal a more nuanced picture of dynamic, context-dependent miRNA target repertoires, and thereby improve our understanding of the diversity of biological outcomes that miRNA-mediated gene regulation can achieve *in vivo*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Kathrin Kunzer and Lan Xu for their help with *C. elegans* strain generation. We are grateful to Matyas Ecsedi for initial observations on target specificity of *let-7* family members. We thank Florian Aeschimann for reagents and helpful discussions and Iskra Katic for worm injections and reagents. We thank Laurent Gelman and Steven Bourke for help with confocal imaging; Roland Nitschke (Life Imaging Center, University of Freiburg, Germany) and Carl Zeiss (Jena, Germany) for sharing the macro for Multiple Position/Tile Imaging acquisitions; Raphael Thierry, Jan Eglinger and Moritz Kirschmann (University of Zurich) for help with image analysis; and Amy Pasquinelli for *C. elegans* strains. Some strains were provided by the Caenorhabditis Genetics Center (CGC), which is funded by the National Institutes of Health Office of Research Infrastructure Programs (P40 OD010440). We thank Matyas Ecsedi, Sarah Carl, Benjamin Towbin, Iskra Katic and Witold Filipowicz for a critical reading of the manuscript.

FUNDING

NCCR RNA & Disease funded by the Swiss National Science Foundation; Novartis Research Foundation through

the FMI (to H.G.); Boehringer Ingelheim Fonds PhD Fellowship (to G.B.). Funding for open access charge: Internal Funds.

Conflict of interest statement. None declared.

REFERENCES

1. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
2. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
3. Brennecke,J., Stark,A., Russell,R.B. and Cohen,S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
4. Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
5. Lai,E.C. (2002) Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.*, **30**, 363–364.
6. Chandradoss,S.D., Schirle,N.T., Szczepaniak,M., MacRae,I.J. and Joo,C. (2015) A Dynamic Search Process Underlies MicroRNA Targeting. *Cell*, **162**, 96–107.
7. Parker,J.S., Parizotto,E.A., Wang,M., Roe,S.M. and Barford,D. (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell*, **33**, 204–214.
8. Schirle,N.T., Sheu-Gruttadauria,J. and MacRae,I.J. (2014) Structural basis for microRNA targeting. *Science*, **346**, 608–613.
9. Broughton,J.P., Lovci,M.T., Huang,J.L., Yeo,G.W. and Pasquinelli,A.E. (2016) Pairing beyond the seed supports microRNA targeting specificity. *Mol. Cell*, **64**, 320–333.
10. Grosswendt,S., Filipchyk,A., Manzano,M., Klironomos,F., Schilling,M., Herzog,M., Gottwein,E. and Rajewsky,N. (2014) Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol. Cell*, **54**, 1042–1054.
11. Helwak,A., Kudla,G., Dudnakova,T. and Tollervey,D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, **153**, 654–665.
12. Moore,M.J., Scheel,T.K., Luna,J.M., Park,C.Y., Fak,J.J., Nishiuchi,E., Rice,C.M. and Darnell,R.B. (2015) miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.*, **6**, 8864.
13. Roush,S. and Slack,F.J. (2008) The let-7 family of microRNAs. *Trends Cell Biol.*, **18**, 505–516.
14. Abbott,A.L., Alvarez-Saavedra,E., Miska,E.A., Lau,N.C., Bartel,D.P., Horvitz,H.R. and Ambros,V. (2005) The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. *Dev. Cell*, **9**, 403–414.
15. Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvie,A.E., Horvitz,H.R. and Ruvkun,G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, **403**, 901–906.
16. Ecsedi,M., Rausch,M. and Großhans,H. (2015) The let-7 microRNA directs vulval development through a single target. *Dev. Cell*, **32**, 335–344.
17. Slack,F.J., Basson,M., Liu,Z., Ambros,V., Horvitz,H.R. and Ruvkun,G. (2000) The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol. Cell*, **5**, 659–669.
18. Vella,M.C., Choi,E.Y., Lin,S.Y., Reinert,K. and Slack,F.J. (2004) The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev.*, **18**, 132–137.
19. Bartel,D.P. and Chen,C.Z. (2004) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.*, **5**, 396–400.
20. Frokjaer-Jensen,C., Davis,M.W., Ailion,M. and Jorgensen,E.M. (2012) Improved Mos1-mediated transgenesis in C. elegans. *Nat. Methods*, **9**, 117–118.
21. Frokjaer-Jensen,C., Davis,M.W., Hopkins,C.E., Newman,B.J., Thummel,J.M., Olesen,S.P., Grunnet,M. and Jorgensen,E.M. (2008) Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat. Genet.*, **40**, 1375–1383.
22. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. 3rd and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
23. Zheng,L., Baumann,U. and Reymond,J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
24. Katic,I., Xu,L. and Ciosk,R. (2015) CRISPR/Cas9 genome editing in Caenorhabditis elegans: evaluation of templates for homology-mediated repair and knock-ins by homology-independent DNA repair. *G3 (Bethesda)*, **5**, 1649–1656.
25. Frokjaer-Jensen,C., Davis,M.W., Sarov,M., Taylor,J., Flibotte,S., LaBella,M., Pozniakovsky,A., Moerman,D.G. and Jorgensen,E.M. (2014) Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. *Nat. Methods*, **11**, 529–534.
26. Schindelin,J., Arganda-Carreras,I., Frise,E., Kaynig,V., Longair,M., Pietzsch,T., Preibisch,S., Rueden,C., Saalfeld,S., Schmid,B. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682.
27. Mok,D.Z., Sternberg,P.W. and Inoue,T. (2015) Morphologically defined sub-stages of C. elegans vulval development in the fourth larval stage. *BMC Dev. Biol.*, **15**, 26.
28. Aeschimann,F., Kumari,P., Bartake,H., Gaidatzis,D., Xu,L., Ciosk,R. and Großhans,H. (2017) LIN41 post-transcriptionally silences mRNAs by two distinct and position-dependent mechanisms. *Mol. Cell*, **65**, 476–489.
29. Vella,M.C., Reinert,K. and Slack,F.J. (2004) Architecture of a validated microRNA::target interaction. *Chem. Biol.*, **11**, 1619–1623.
30. Vadla,B., Kemper,K., Alaimo,J., Heine,C. and Moss,E.G. (2012) lin-28 controls the succession of cell fate choices via two distinct activities. *PLoS Genet.*, **8**, e1002588.
31. Martinez,N.J., Ow,M.C., Reece-Hoyes,J.S., Barrasa,M.I., Ambros,V.R. and Walhout,A.J. (2008) Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. *Genome Res.*, **18**, 2005–2015.
32. Baek,D., Villen,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
33. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
34. Didiano,D. and Hobert,O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.*, **13**, 849–851.
35. Lai,E.C., Tam,B. and Rubin,G.M. (2005) Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev.*, **19**, 1067–1080.
36. Wolter,J.M., Le,H.H., Linse,A., Godlove,V.A., Nguyen,T.D., Kotagama,K., Lynch,A., Rawls,A. and Mangone,M. (2017) Evolutionary patterns of metazoan microRNAs reveal targeting principles in the let-7 and miR-10 families. *Genome Res.*, **27**, 53–63.
37. Abrahante,J.E., Daul,A.L., Li,M., Volk,M.L., Tennessen,J.M., Miller,E.A. and Rougvie,A.E. (2003) The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev. Cell*, **4**, 625–637.
38. Lin,S.Y., Johnson,S.M., Abraham,M., Vella,M.C., Pasquinelli,A., Gamberi,C., Gottlieb,E. and Slack,F.J. (2003) The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev. Cell*, **4**, 639–650.
39. Esquela-Kerscher,A., Johnson,S.M., Bai,L., Saito,K., Partridge,J., Reinert,K.L. and Slack,F.J. (2005) Post-embryonic expression of C. elegans microRNAs belonging to the lin-4 and let-7 families in the hypodermis and the reproductive system. *Dev. Dyn.*, **234**, 868–877.

2.2: Tissue Specificity of the *let-7* family

Giovanna Brancati, Matyas Ecsedi & Helge Grosshans conceived the project; ME generated the strains HW1191, 1223, 1127. GB performed and analyzed all the experiments with technical help from Kathrin Kunzer for strain generations and under HG's supervision.

Previous data of the lab revealed that a GFP reporter containing the *lin-41* 3'UTR was silenced differently among tissues (Ecsedi, 2015). In particular, it was proposed that *lin-41* silencing was specific to *let-7* only in the vulva, while all the *let-7* family members were required for silencing in the hypodermis.

One plausible explanation is that the four miRNAs of the *let-7* family have different expression patterns. For example, the sisters could fail to repress *lin-41* in the vulva because they are not present in this tissue. Alternatively, because the *let-7* family regulation is intertwined, deletion of one gene can result in altered expression of the others, an effect that is difficult to study in a tissue-specific manner.

To clarify this issue, we aimed to investigate in detail the expression patterns of the *let-7* family studying both the activity of their promoters, through transcriptional reporters, and their mature levels in different tissues, by establishing a method that allows FACS sorting of specific cell types.

2.2.1 Expression pattern of the *let-7* family (1)

Although the expression pattern of the *let-7* family has been previously described (Roush, 2008), detailed data at the L4 stage, the time when *let-7* silences *lin-41* are missing. Moreover, the transgenes used, which carried a cytoplasmic GFP fused to the miRNA promoters, usually were multi-copy arrays of unknown copy number randomly integrated in the genome or extra-chromosomal (Evans, 2006). This led to inconsistent results, even after analysis of several independent strains carrying the same reporter, as described for miR-241 (Esquela-Kerscher et al., 2005).

Therefore, to confirm the expression patterns and the presence of the *let-7* family in the hypodermis and vulva cells of late L4 worms (the time when *let-7* silences *lin-41*), we generated new transcriptional reporters. These are single-copy integrated in the genome and contain the putative (~2

kb upstream the miRNA sequence) promoter of the miRNA, a nuclear destabilized GFP (GFP/PEST::H2B) and the unregulated *unc-54* or *let-858* 3'UTR (Figure 2.1).

Expression patterns of our single-copy integrated reporters resembled what was described before (Esquela-Kerscher et al., 2005; Hayes et al., 2006; Li et al., 2005; Martinez et al., 2008) with few exceptions. While the *let-7* promoter is active in both seam cells and hyp7, as shown by GFP expression, P*mir-48* driven GFP was restricted to the seam cells. For miR-84 and miR-241 reporters GFP was detected only in the hyp7 and at lower levels than in other tissues (Figure 2.1A). We do not know if the previously reported expression in the seam cells was an artifact due to the overexpressed multi-copy arrays used before (Johnson et al., 2005; Martinez et al., 2008). In fact, some other authors report that the hypodermal expression of miR-241 was not observed consistently in the transgenic lines that they generated (Li et al., 2005).

In the vulva, GFP was detected in all cells in *let-7*, miR-48 and miR-84 reporters, as expected from previously reported expression in the vulval precursor cells (Esquela-Kerscher et al., 2005; Roush, 2008). However, we could not detect vulval GFP expression in the P*miR-241* reporter strain (~1.3kb, HW2279); even if we could confirm the expression in body wall muscles and ventral nerve cord. The same result was true for another strain containing a longer promoter sequence (~2kb, HW2278). P*mir-241* was both reported for being active in the vulva (Li et al., 2005; Roush, 2008) and only present in the vulval musculature (Martinez et al., 2008). We then looked at the VT1189 strain, which was used by Martinez and colleagues and is publicly available. The strain carries an integrated array and shows no GFP in the vulva (Figure 2.1B). This strain showed a strong and diffused GFP signal around this organ (Figure 2.1B, white box) that induced us to think that the previously reported vulval expression was rather the cytoplasmic GFP signal in the muscle cells around it.

However, it is unclear if miR-241 expression can be recapitulated by such transcriptional reporters at all. In fact, the miR-241 gene is upstream of miR-48 in a cluster on chromosome V and it is only 1.7kb apart from the paralog (Figure 2.1C), therefore we cannot exclude that regulatory elements controlling miR-241 expression are present in the intervening sequence between the two miRNAs that is excluded from the reporters. For these reasons, miR-241 expression remains elusive.

In conclusion, our single-copy integrated transcriptional reporters show a more precise expression pattern for the four miRNAs than what was previously described with arrays. In addition, we extended previous published results by assaying the expression in late L4 staged animals: While

the *let-7* family expression in the vulva cells is overlapping (apart from miR-241), in the hypodermis, it is slightly distinct.

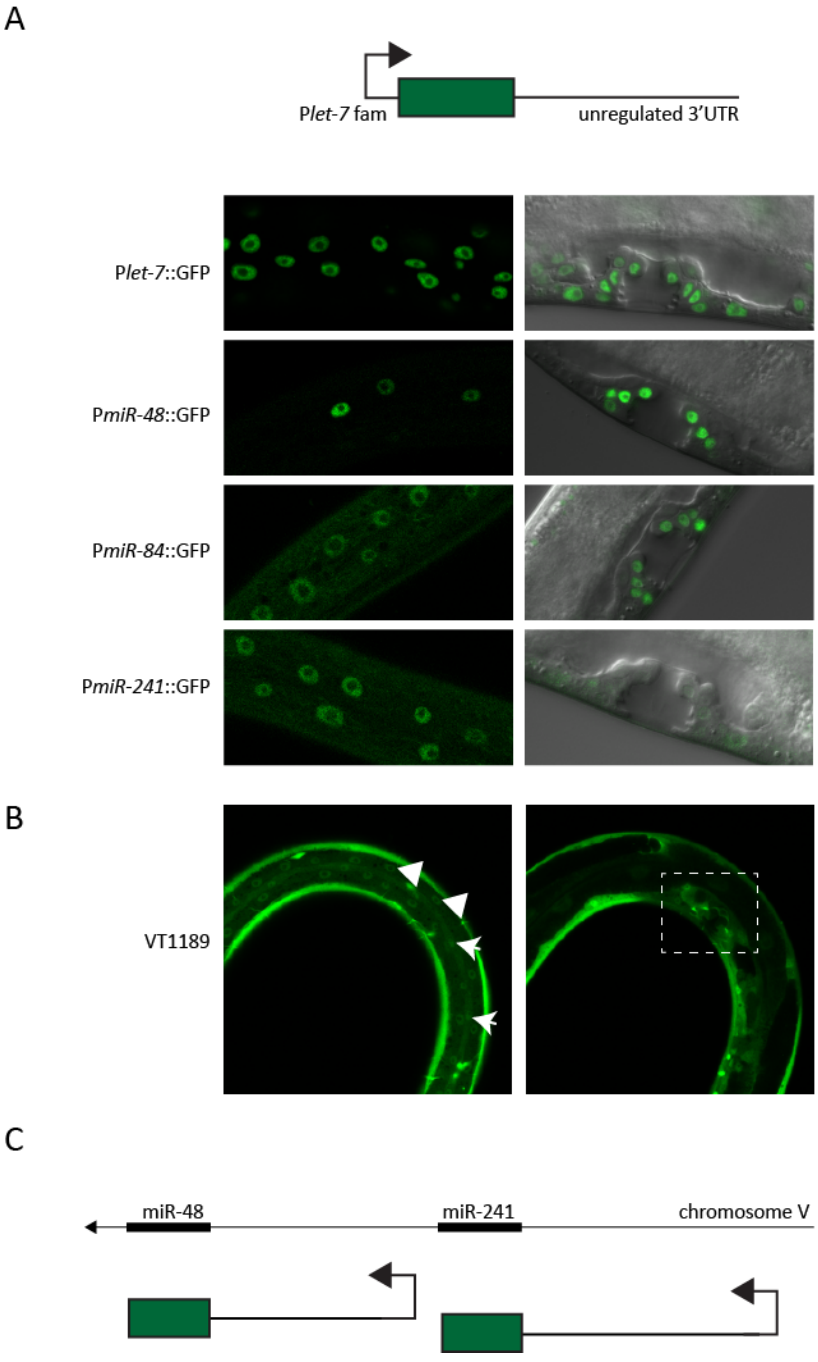


Figure 2.1. *Let-7* family expression pattern in late L4 worms (A) Confocal images of late L4 worms expressing nuclear-localized GFP (PEST)::H2B reporters, driven from the putative promoters of the four *let-7* family members (P*let-7*fam). Shown the hypodermis (left) and the vulva (right). **(B)** Confocal images of late L4 worms carrying an integrated array containing a putative Pmir-241 promoter and cytoplasmic GFP (strain VT1189). Left panel shows the hypodermis, white arrows = seam cells, arrowheads = hyp7 cells. Right panel shows the vulva,

highlighted by the dashed box. **(C)** Cartoon of the *miR-48/miR-241* cluster on chromosome V and of the two GFP reporters used in (A). The two miRNAs are 1.7kb apart.

2.2.2 Expression pattern of the *let-7* family (2): mature miRNA levels in different tissues

To confirm that the *let-7* levels are altered in specific tissues of *miR-48/241/84 (-)* animals, quantification of the miRNA in each tissue is needed. In fact, transcriptional reporters give us an idea about temporal and spatial distribution, but do not consider post-transcriptional regulation (Van Wynsberghe et al., 2011; Zisoulis et al., 2012). Methods that allow tissue-specific studies are missing in the *C. elegans* community (paragraph 1.4) thus molecular biology techniques that allow detection of the mature form of miRNAs have only been performed on whole worm lysates (Esquela-Kerscher et al., 2005; Li et al., 2005; Vadla et al., 2012). Similarly, *in situ* hybridization protocols for miRNAs detection have been described, but they have poor resolution, cannot reveal cell-specific expression and are not quantitative (Andachi and Kohara, 2016). Therefore, evidence for the actual level of the mature and active form of miRNAs in specific tissues is missing.

To fill this gap and obtain quantitative data on the expression of the mature forms of the *let-7* family, we optimized a recently published method that allows dissection of the worm and cell recovery (Materials and Methods and (Zhang, 2013)). With the precious support of Hubertus Koehler (FMI, FACS facility), we were able to sort several cell types at different developmental stages. Because activity of the *let-7* family has been mostly described in seam and vulva cells of L4 animals, we first tried to FACS-sort these cell types to quantify the four miRNAs. However, those tissues in L4 animals were difficult to sort and although we could profile their transcriptome (chapter 2.3), we could not obtain small RNA data due to the amount of cells and RNA extracted, which was low for the protocols available at that time.

In conclusion, we obtained primary *C. elegans* cells and profiled their transcriptome (as described in next chapter 2.3). However, quantification of the *let-7* miRNAs in the two relevant tissues in which they function to test how their abundance in different tissues affects their repressive activity towards *lin-41* or other targets was not possible.

2.2.3 In the hypodermal tissue *lin-41* is probably silenced only by *let-7*

As mentioned, some tissue-specific activity of the *let-7* family towards the *lin-41* reporter has been described (Ecsedi, 2015). Although transcriptional reporters reveal their expression pattern, we

could not quantify the levels of the *let-7* family in hypodermis and vulva and relate their abundance to *lin-41* silencing (paragraph 2.2.2), thus we addressed the problem from another angle.

It was suggested that *lin-41* could be silenced by the whole *let-7* family in the hypodermis, while only by *let-7* in the vulva (Ecsedi, 2015). To confirm this result, we repeated the same analysis in those two tissues with a slightly different reporter, the *unc-54+let-7site*, in wild type, *let-7*ts and *miR-48/241/84 (-)* background. First, we confirmed that this reporter was specific to *let-7* in the vulva of L4 worms and mimicked *lin-41* full-length regulation (Brancati and Grosshans, 2018). Then, we reproduced the results previously described, namely that the *unc-54*+*let-7*sites reporter was silenced in the hypodermis of wild type animals, shown by absence of GFP, but failed to be repressed in both *let-7* and *miR-48/241/84 (-)* background ((Ecsedi, 2015) and Figure 2.2A). This result suggests that all the *let-7* family members can effectively silence *lin-41* in the hypodermis in contrast to the vulva cells, where only *let-7* represses *lin-41*.

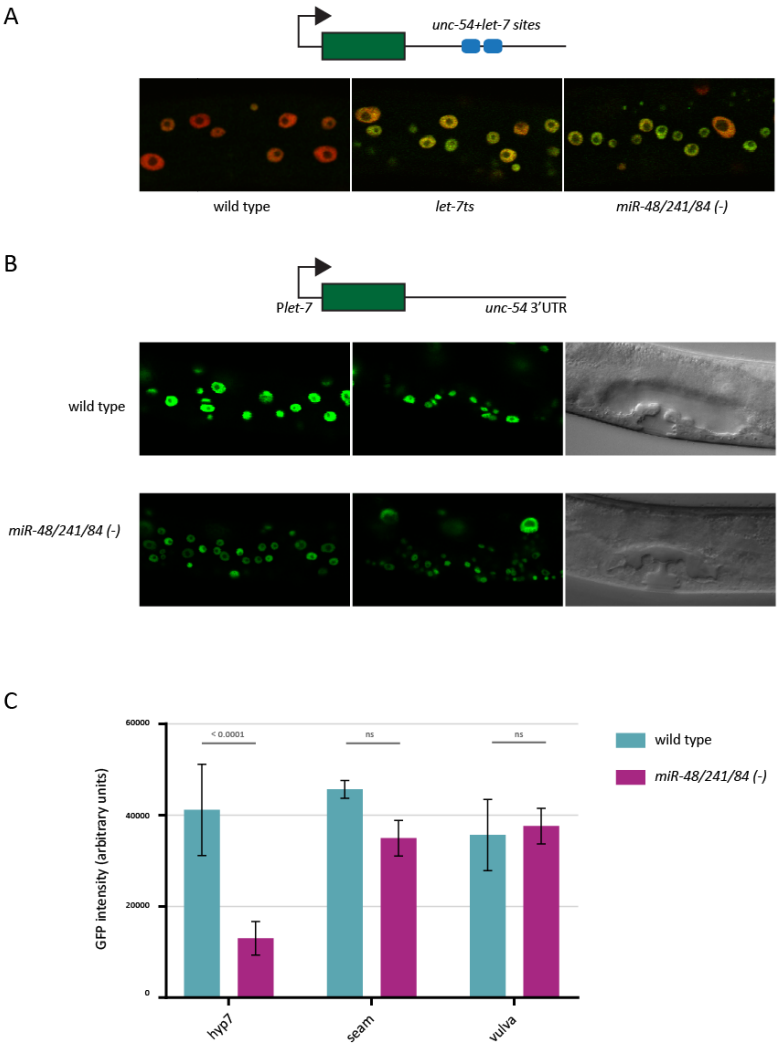


Figure 2.2 *let-7* is transcriptionally repressed in the hypodermis of *miR-48/241/84 (-)* animals
(A) Confocal images of *the unc-54+let-7sites* reporter in the hypodermis of wild type worms, or animals lacking *let-7* (*let-7ts*) or the three sisters miR-48/miR-241/miR-84 (*miR-48/241/84 (-)*). The worms carry a ubiquitous *GFP::unc-54+let-7 sites* reporter (described in chapter 2.1, Figure 3, and carrying *let-7 s*ites) and a ubiquitous *mCherry::unc-54* reporter. Shown GFP and mCherry overlay. **(B)** Confocal images of late L4 animals carrying a nuclear GFP (GFP/PEST::H2B) driven by *let-7* promoter and followed by the unregulated *unc-54* 3'UTR. Shown are hypodermis (left) and vulva cells (right) in wild type and *miR-48/241/84 (-)* animals. **(C)** Quantification of GFP intensity in hyp7, seam and vulva cells of animals carrying P*let-7*::gfp reporters in wt or *miR-48/241/84 (-)* background. Analysis performed with Fiji on 12hyp7 cells, 4 seam cells and 8 vulva cells of four randomly chosen animals per condition after visual inspection of 13 wt and 15 *miR-48/241/84 (-)* animals. ± SD 2way ANOVA, Sidak's multiple comparisons test, performed and graphed through GraphPad PRISM.

We excluded the differential expression of the *let-7* family members as the reason for these different effects in hypodermis and vulva, given that we can infer that the sisters are expressed in the vulval cells from analysis of transcriptional reporters, which describe temporal and spatial activation of their promoters, even with the limitation previously described (Figure 2.1). Furthermore, we can

confidently assume that they are active in this tissue thanks to reporters containing binding sites to the *let-7* family that can be repressed by any *let-7* paralogue (Figure 2 and 3 in (Brancati and Grosshans, 2018) or *hbl-1* 3'UTR reporter (Ecsedi, 2015)). Therefore, we wondered, what if the apparent requirement of the sisters for *lin-41* silencing in the hypodermis was caused by a reduction in *let-7* levels in the *miR-48/241/84 (-)* strain? In other words, in an animal lacking the three *let-7* sisters, *let-7* expression could be affected only in the hypodermis resulting in failure to silence *lin-41* in this tissue. RT-qPCR experiments performed on whole worm lysates of the *miR-48/241/84 (-)* background did not show any significant alteration of *let-7* levels (ME personal communication). However, such experiments are not able to detect subtle changes in gene expression, e.g. caused by a tissue-specific effect.

To test if *let-7* may be altered only in some tissues, we tested the expression of the P*let-7*::GFP that recapitulates *let-7* transcription (introduced in Figure 2.1A). Figure 2.2B shows representative confocal images and the quantification of GFP signal intensity (Figure 2.2C) of worms carrying such reporter in wild type or *miR-48/241/84 (-)* background. As shown by the lower GFP signal, animals lacking the three *let-7* sisters show a reduced activity of the *let-7* promoter in the hypodermal hyp7 cells (Figure 2.2B and C), while seam or vulval cells are unaffected.

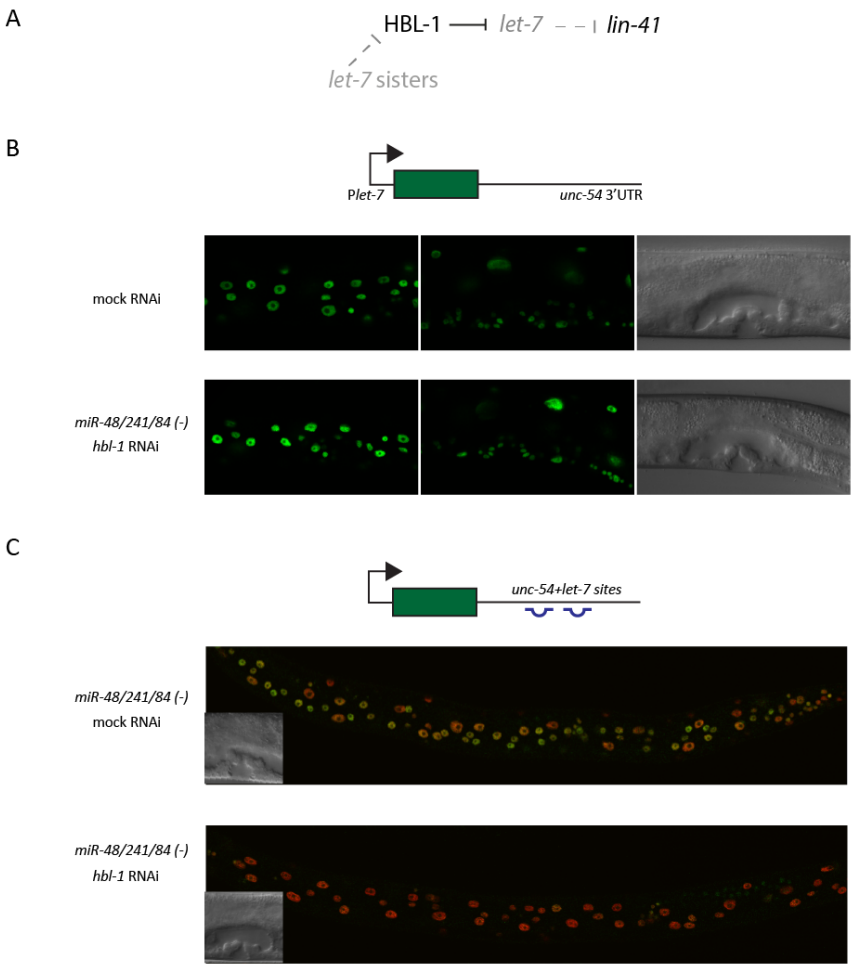


Figure 2.3 *hbl-1* knock down restores *lin-41* silencing in hypodermis of *miR-48/241/84 (-)* animals
(A) Model of part of the heterochronic pathway centered on the *let-7* regulation in the hypodermis of young worms. When the *let-7* sisters are absent, *HBL-1* accumulates and transcriptionally represses *let-7*, thus allowing *lin-41* expression. In wild type animals, the sisters repress *hbl-1* thus allowing *let-7* accumulation (not shown). Color scheme: gray = deleted or repressed; black = expressed. **(B)** Confocal images of *miR-48/241/84 (-)* animals carrying the P*let-7*::gfp reporter subjected to mock or *hbl-1* RNAi (n = 8) show increased GFP signal when *hbl-1* is knocked down. **(C)** Confocal images of late L4 *miR-48/241/84 (-)* animals treated with mock or *hbl-1* RNAi. The worms carry the GFP::*unc-54*+*let-7* sites and the mCherry::*unc-54* reporters as in 2.2A. Shown GFP and mCherry overlay. Vulva is shown at the left side corner as a reference for the worm age (n= 7). *let-7*ts*: let-7(n2853) X; miR-48/241/84 (-): miR-48/miR-241(ndf51) V; miR-84(n4037) X.*

It has been described that *hbl-1* transcriptionally represses *let-7* expression in a tissue-specific fashion, in particular in the hypodermis (Roush and Slack, 2009). It is also known that the three *let-7* sisters downregulate *hbl-1* expression post-transcriptionally (Abbott et al., 2005). Therefore, we hypothesized that in mutant animals lacking the three *let-7* sisters, HBL-1 accumulates and represses *let-7*, thus preventing *lin-41* downregulation specifically in the hypodermis (as outlined in Figure 2.3A).

To prove that transcriptional repression of *let-7* and thus of our P*let-7*::GFP reporter was dependent on *hbl-1*, as suggested earlier, we treated *miR-48/241/84 (-)* worms with *hbl-1* RNAi to reduce *hbl-1* levels and release *let-7* repression. In fact, in *miR-48/241/84 (-)* animals, because post-transcriptional regulation by the *let-7* sisters is missing, the misregulated HBL-1 might repress *let-7* more than in wild type worms. As hypothesized, knockdown of *hbl-1* resulted in enhanced expression of P*let-7*::GFP, and according to the tissue-specific origin of this regulation, this was true in the hypodermis, but not in the vulva nor the seam cells of the animals (Figure 2.3B).

Finally, we wanted to confirm that the tissue-specific effect originally described in (Ecsedi, 2015), namely that *lin-41* was repressed by both *let-7* and the sisters in the hypodermis, was an artifact due to *hbl-1*-induced *let-7* repression in the hypodermis of worms lacking the three sisters. Therefore, we imaged the *unc-54*+*let-7* sites reporter in *miR-48/241/84 (-)* background in *hbl-1* RNAi conditions (Figure 2.3C). While the reporter shows no sign of downregulation in the hypodermis of *miR-48/241/84 (-)* animals treated with mock RNAi, knockdown of *hbl-1* restores complete silencing in the hypodermis (Figure 2.3C, compare also to Figure 2.2A). We concluded that *let-7* levels in hypodermis of animals lacking the three *let-7*sisters are affected by *hbl-1* overexpression.

Chapter Discussion

It was suggested that the specific repression of *lin-41* by *let-7* takes place only in some tissues, such as the vulva (Ecsedi, 2015). In fact, Ecsedi observed that, in the hypodermis, ablation of the three *let-7* sisters partially abolished repression of a GFP reporter carrying the *lin-41* 3'UTR. The author thus concluded that all the members of the *let-7* family are required for *lin-41* silencing in the hypodermis, and that only *let-7* is essential in the vulva.

Even though we cannot measure the exact levels of the mature miRNAs, we can at least exclude the trivial explanation that *lin-41* specificity to *let-7* in the vulva is due to the lack of the sisters in this tissue. In fact, the sisters are likely expressed in vulva cells (Figure 2.1) and our results show that the observed specificity depends on the *lin-41/let-7* site architecture (Brancati and Grosshans, 2018).

Regarding the activity in the hypodermis, our data suggest that the absence of *lin-41* repression in worms lacking the three *let-7* sisters is not evidence for the involvement of the whole *let-7* family in *lin-41* silencing, as suggested in (Ecsedi, 2015). Rather, we hypothesize that it is a

consequence of reduced *let-7* levels in *miR-48/241/84 (-)* animals and of the complicated feedback loops that characterize the *let-7* family and the heterochronic genes (Figure 2.3).

Deletion of the *let-7* sisters and consequent lack of post-transcriptional regulation likely results in higher levels of HBL-1 protein in the hypodermis, thus leading to transcriptional repression and reduced *let-7* levels, which leads to inefficient *lin-41* silencing. Accordingly, in this tissue, reduction of *hbl-1* levels via RNAi in worms lacking the three *let-7* sisters restores complete hypodermal silencing of *lin-41* and of the reporter that we have shown to be strictly *let-7* specific in the (Brancati and Grosshans, 2018). This suggests that *lin-41* silencing is likely specific to *let-7* in the hypodermis, too. Analysis of the HBL-1 protein through GFP reporter might confirm the expression changes of such protein in the two genetic backgrounds. However, a validated single-copy integrated reporter for *hbl-1* is not currently available, thus we could not investigate this further.

In conclusion, we show that *lin-41* is repressed only by *let-7* in each worm tissue. Given what we have learned about specificity (Brancati and Grosshans, 2018), even if the sisters were present in each tissue where *let-7* functions, they would not be able to engage in effective silencing. The duplex that the sisters would form with the *lin-41* sites has imperfect seed matches that do not allow engagement in effective silencing. Moreover, given that the central bulge and wobble in the seed match avoid silencing by miRNAs that do not have extensive 3'pairing, even when they are overexpressed, we suggest that *lin-41* would maintain strict specificity to *let-7* in every tissue. Whether the sisters have different targets, like *hbl-1* or *daf-12*, if *let-7* can contribute to their silencing in some tissues or if more instances of target specificity exist for this miRNA family still needs further investigation.

2.3 The worm, transcripts and miRNAs, one cell at a time

GB & HG conceived and designed the project and analyzed the data with help from Dimos Gaidatzis. GB optimized the dissociation and FACS protocol. Experiments were performed with technical help from Kathrin Kunzer, Hubertus Kohler (FMI, FACS) and Birgitte Lucas (FMI, Functional Genomics), unless otherwise stated. DG or SC performed the bioinformatics analysis.

As discussed in the previous chapter, we aimed at obtaining primary *C. elegans* cells to profile miRNAs, and in particular the *let-7* family, to address the interplay between miRNA abundance and target specificity.

We succeeded in optimizing a FACS-based protocol to isolate specific cell types. Given the general lack of methods to obtain primary cells and high quality tissue-specific data in the *C. elegans* community (paragraph 1.4), such a result holds the great potential to expand our knowledge about cell-type specific gene expression.

Here, we describe isolation of cells from several tissues at the L3 and L4 stage, which we subjected to mRNA sequencing to reveal their unique profiles. Furthermore, we show the strength of our approach applying it to more challenging high-resolution experiments. A single-cell sequencing experiment, the long-sought miRNA profiling and an application of the tissue-specific data to understanding *C. elegans* rhythmic gene expression are described below.

FACSeq: dissociation, sorting and sequencing of C. elegans cells

To isolate different cell types, we selected worm strains carrying tissue-specific fluorescent reporters: *scm::gfp* for seam cells, *egl-17::mCherry* for vulva cells, *myo-3::RFP* for muscles, *dpy-5::GFP* for hyp-7 and *rab-3::GFP* for neurons (see the *Worm strains* table for details). All the markers are strongly specific for the tissue that is listed, except for the *scm* promoter and *egl-17*. *Scm* is exclusively expressed in seam cells in L4 and adult worms, but for younger larvae, we cannot exclude that a small percentage of GFP positive cells are hyp7. Microscopy analysis of the strain carrying the *scm::gfp* reporter revealed that when a seam cell undergoes asymmetric cell divisions, both daughter cells retain GFP until one of them fuses to the hyp7 syncytium and loses it. However, for simplicity, cells isolated from the *scm::gfp* strain at the L3 stage will be referred to as “L3 seam cells”. In L4 worms

there are 32 seam cells and at the larval-to-adult transition, they fuse into a syncytium, a structural feature that possibly makes the extraction more challenging than in younger animals. We refer to such cells as “L4 seam cells”. Lastly, although we generally refer to “L4 vulva cells”, we specifically sorted only two vulval cell types (in a worm, there are 22 vulva cells of seven types (vulA, vulB1-B2, vulC-F, (Schindler and Sherwood, 2013)). Since most of the genes known to be expressed in the vulva are also expressed in other tissues, we chose the promoter of the gene *egl-17* to drive the sort, as it is strictly expressed in vulC and vulD (Inoue et al., 2002).

To obtain tissue-specific transcriptomes, we generated primary cultures from L3 or L4 stage worms carrying one of the fluorescent reporters described above. The dissociation protocol, adapted from (Zhang, 2013) is described in the Materials and Methods section. Briefly, larvae were subjected to mechanical dissociation and enzymatic treatment. The resulting cell suspensions were incubated with propidium Iodide or DAPI to avoid collection of dead cells. 10,000 cells isolated through a BD FACS aria with a 70um nozzle were collected. After RNA isolation, sequencing libraries were prepared with the SmartSeq2 protocol with the help of Birgitte Lucas of the functional genomic facility at FMI.

For each experiment a set of three samples was collected and sequenced: the fluorescently-labeled cells; the *input*, which is the cell suspension that runs through the FACS and that contains all the cells of the worms selected for viability (propidium iodide (PI) or DAPI negative); and the *non-fluorescent* cells (GFP or mCherry/RFP (-)). Figure 2.4 shows representative plots of one L3 seam cell sorting. After analysis of several replicates, we acknowledge that our *input* sample does not correlate 100% with a whole worm, as some genes seem to be lost during the worm dissociation. We hypothesize that some cell types are more sensitive than others are to our dissection protocol and probably die even before sorting. Neurons, for example, are usually depleted in our preparations.

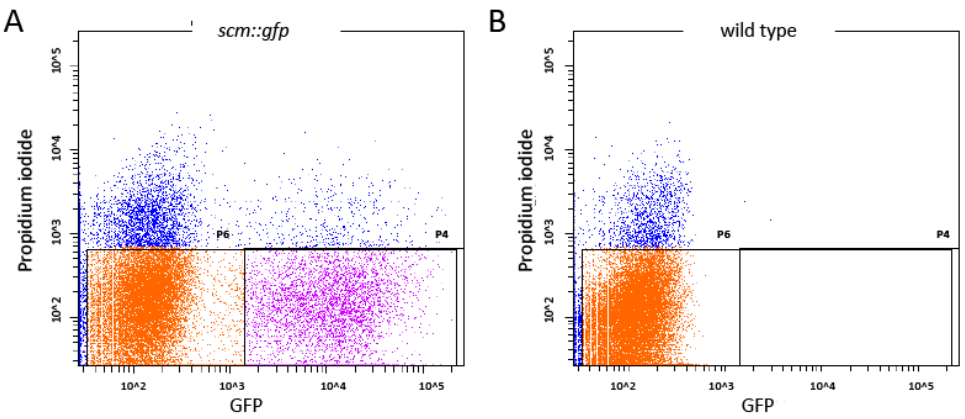


Figure 2.4 Representative L3 seam cell sorting. (A) Seam cells (P4, purple) are GFP positive and propidium iodide (PI) negative. Non-fluorescent cells (P6, orange) are GFP and PI negative. The *input* cointains both GFP positive and negative cells, PI negative (P4+P6). **(B)** Control cells derived from wild type non-fluorescent worms, treated with PI.

2.3.1 Cell-type specific gene expression

Because of our interest in the hypodermal tissue and the vulva of the animals as the tissues where the *let-7* family is functional and where heterochronic phenotypes can easily be observed, we focused our analysis on those tissues, even if we collected data for muscles and neurons, too.

The hypodermal seam cell-enriched genes

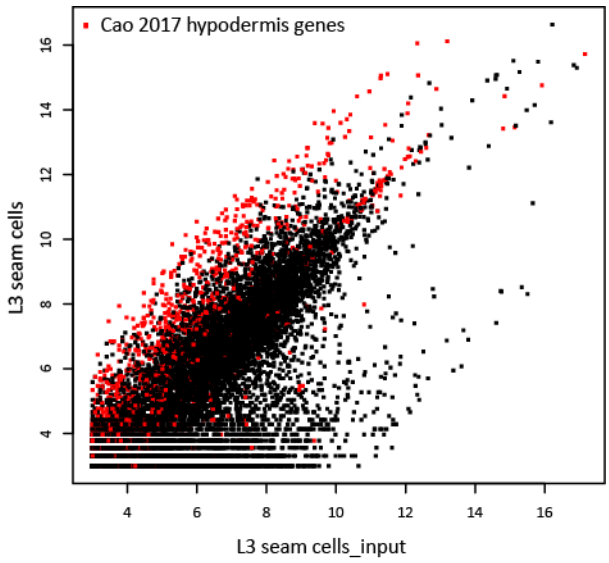


Figure 2.5 Comparison of the L3 seam cells sample to its input identifies the seam cell-enriched genes. Shown are read counts (log2) for every gene. Highlighted in red the "hypodermal genes" by (Cao et al., 2017).

We thus performed gene expression profiling for seam and hyp7 cells at L3 and seam and vulva cells at L4 stage. For every tissue, we collected and profiled three samples: input, fluorescently labeled and non-fluorescently labeled cells. We determined tissue specific expression by comparing fluorescently labeled fraction to the respective input control. Figure 2.5 depicts such comparison for the case of seam cells. To validate that our approach is indeed able to detect tissue specific expression, we compared our data to a recently published single-cell expression atlas, where hypodermal seam cell specific genes have been defined (Cao et al., 2017). When we highlighted those genes in Figure 2.4, we could detect a clear enrichment in the fluorescently labeled fraction as opposed to the input.

Comparative analysis of seam, hyp7 and vulval cells

While comparisons of the fluorescently labeled fraction to the input, one tissue at a time, allows for the detection of genes with higher expression in one tissue than in the whole worm, finding the more subtle differences between two similar tissues is more challenging. In this dataset, we were specifically interested in three such subtle comparisons, namely "L4 seam cells to L4 vulva", "L3 seam to L3 hyp7" and "L3 seam to L4 seam". This inevitably involved comparing experiments that were performed on different days.

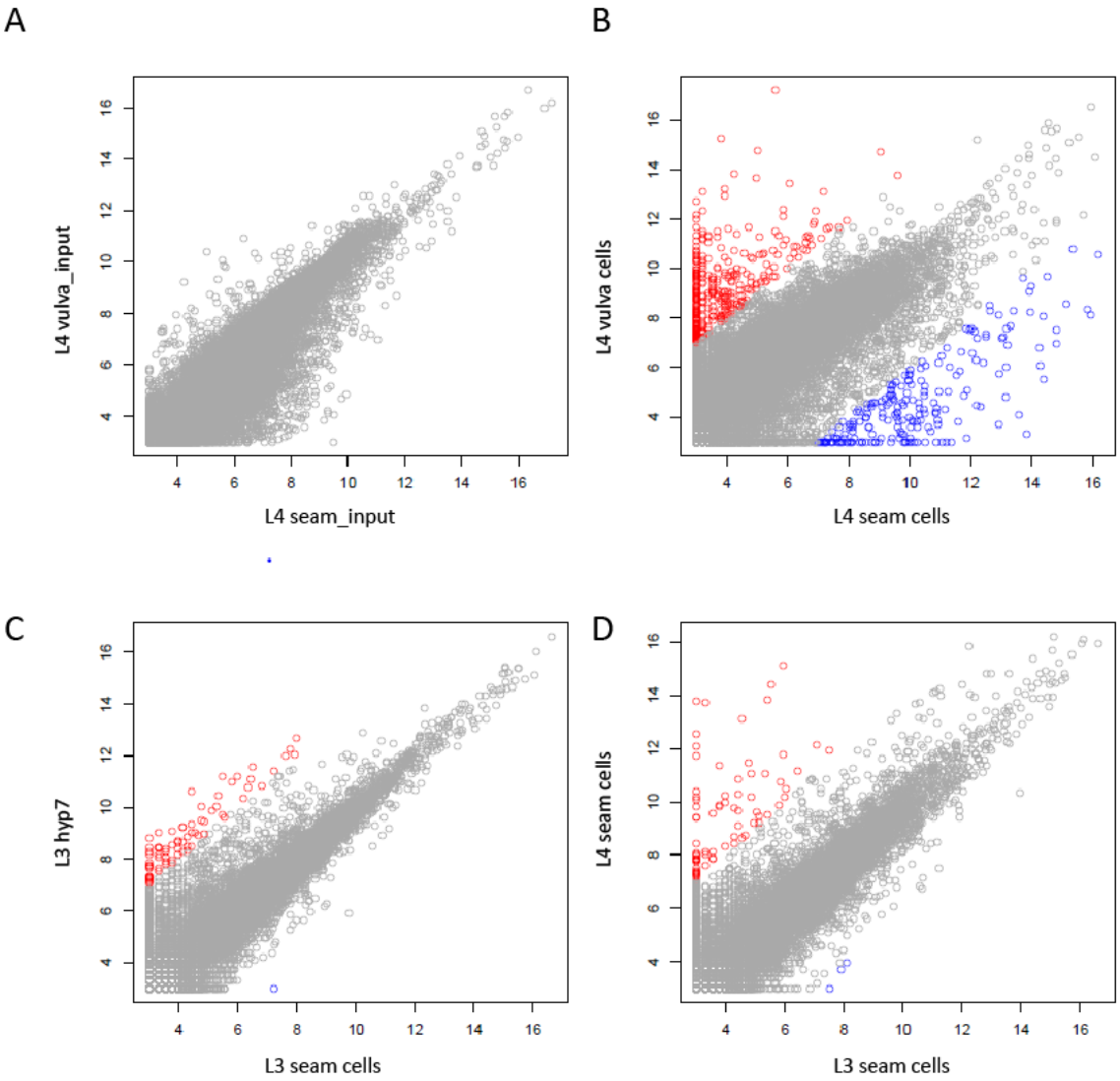


Figure 2.6 Comparative analysis of seam, vulva and hyp7 cells (A) Comparison of the input of the "L4 vulva" and "L4 seam cells" samples reveals variability between sorts performed on different days. Differential analysis for "L4 vulva" vs. "L4 seam" **(B)**, "L3 hyp7" vs. "L3 seam" **(C)** and "L4 seam" vs. "L3 seam" **(D).**

To obtain information about day-to-day variability, we compared individual input samples from the different experiments. Figure 2.6A shows a representative comparison for L4 seam to L4 vulva. The correlation coefficient between the samples was 0.93. This is lower than obtained for typical bulk RNA-seq samples (R > 0.98); demonstrating that creating a single cell suspension introduces substantial variability from experiment to experiment. To account for this variability we set a stringent cutoff for differential expression to 16 fold, as most genes change less than that in the input. Comparing L4 seam cells to L4 vulva showed a high number of differentially expressed genes. We detected 338 genes specific to vulva and 205 genes specific to seam cells (Figure 2.6B). Comparing L3

seam cells to L3 hyp7 cells showed much less differences (Figure 2.6C) which is not surprising given that seam cells and hyp7 cells originate from the same precursors. We detected 74 genes slightly enriched in hyp7 and only one gene slightly enriched in the seam cells. When comparing L4 seam cells to L3 seam cells, we detected 73 upregulated in L4 and 3 slightly downregulated (Figure 2.6D).

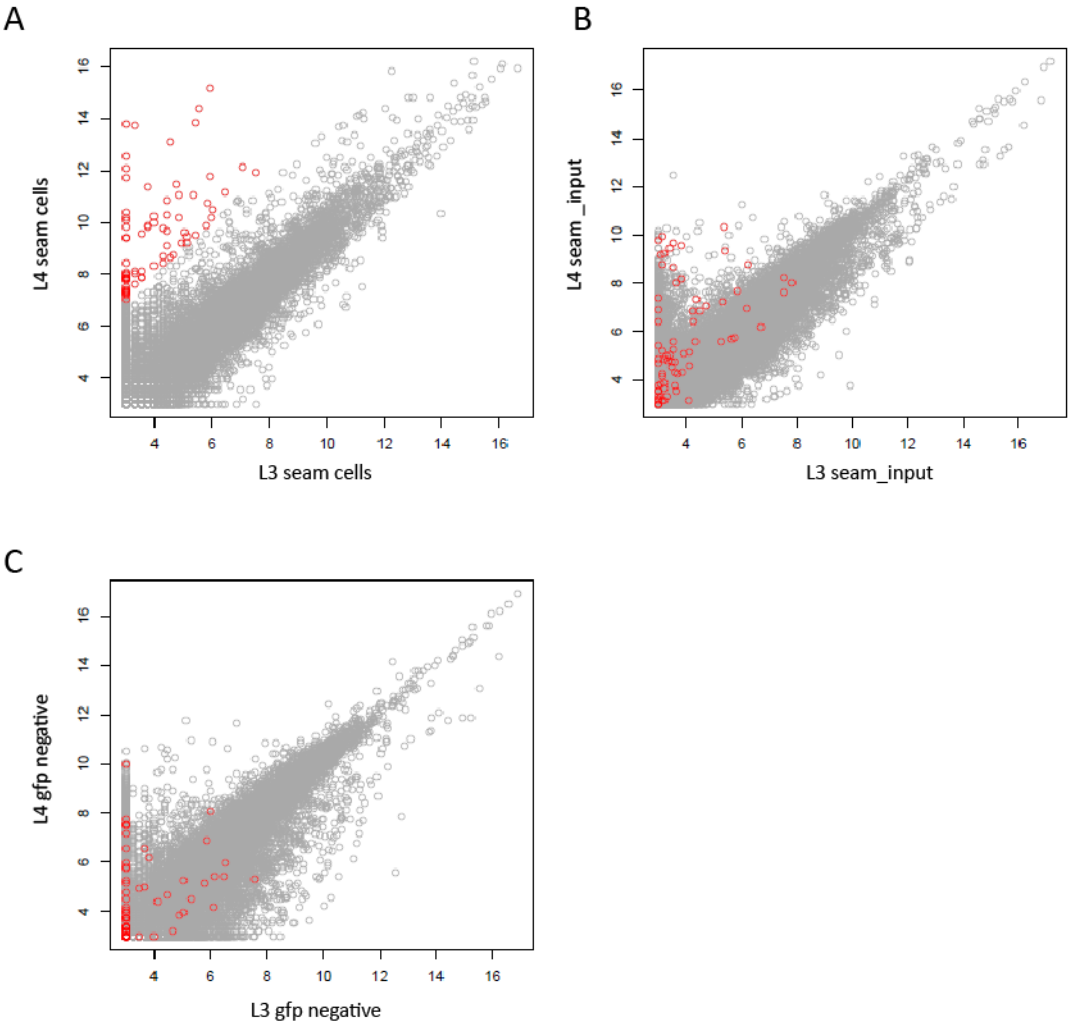


Figure 2.7 Seam cell-enriched genes in L3 and L4 staged larvae. Comparison of "L4 seam cells" vs. "L3 seam cells" **(A)**, of their input **(B)** and of the GFP negative samples **(C)**. Highlighted in red the 73 genes upregulated in "L4 seam cells". (log2) read counts.

In the case of this last comparison (L3 vs. L4 seam cells, Figure 2.6D and 2.7A), estimating the noise from the respective input samples is not possible as these differ in terms of their developmental timing (expression changes are to be expected). We argued that if the 73 genes upregulated in L4 are real, we should see them also being slightly upregulated in the input samples (since seam cells are part

of the input), but less so in the non-fluorescently labeled fraction (since these should be depleted in seam cells). Figure 2.7B shows a comparison of L3 and L4 inputs coloring the 73 genes of interest. Indeed, we can see that those genes are upregulated in the input samples and, as expected, at lower enrichment values as in the fluorescently labeled fraction (given that other tissues make up for a substantial fraction of the worm). Comparing the non-fluorescently labeled fractions (*gfp negative*, Figure 2.7C), we can see even less enrichment than in the input for the 73 genes, supporting the conclusion that those genes are upregulated from L3 to L4 in seam cells.

To find motifs enriched in promoters from different cell types or stages, we ran known and *de novo* motif finding using HOMER (analysis performed with Sarah Carl). The set of known, *C. elegans*-specific motifs was downloaded from the CIS-BP database (Weirauch et al., 2014). HOMER was run comparing promoters from a set of genes highly enriched (log2 fold-change >= 4) in L4 seam cells vs. L4 vulva, L4 vulva vs. L4 seam cells, and seam cells in L4 vs. L3. The tables in the Appendix A list the top 10 hits for such analysis. Notably, *elt-1* is the top hit among the seam cell-enriched motifs compared to vulva cells and it is known for being important for *let-7* family transcription exactly in this cell type (Cohen et al., 2015). Transcription factors of the nuclear hormone receptor (nhr) type, such as nhr-23, which are known for being expressed in the hypodermis (Hayes et al., 2006), are enriched in our analysis, too. The analysis of seam cells at the two developmental stage did not show any significant enrichment.

In conclusion, our data sets provide evidence that our FACSeq approach can be used to study tissue-specific expression patterns in different developmental stages. Analysis of the transcriptome of different sorted cell types identifies genes that are enriched in distinct tissues and allows detection of subtle differences among them.

To demonstrate the usefulness of our protocol, we also used the tissue specific data to understand a peculiar trait of *C. elegans* biology, namely its rhythmic gene expression (Hendriks et al., 2014; Kim et al., 2013).

2.3.2 Tissue-specific oscillatory gene expression

Temporal gene expression analysis with hourly resolution has recently revealed that in *C. elegans* about 20% of the transcriptome (>2700 genes) has a peculiar oscillating behavior (Hendriks et

al., 2014; Kim et al., 2013). In particular, genes can be divided in three categories depending on their behavior over time: oscillating, flat or rising. Flat genes, as the name suggests, do not show any particular change over time, and rising genes, which are mostly germ line specific, tend to go up steadily from the time the gonads start developing. Oscillating genes tend to be expressed in the soma and peak once per larval stage with a period of 8h at 25 °C. Their fold change in oscillatory expression is usually very high, with genes changing from two to more than 10 fold, and they peak across a continuum of phases, which means that at any point in development, some genes are peaking. Furthermore, pairs of different genes maintain a constant phase difference from one to the next oscillation (oscillations are phase-locked). Oscillating transcripts probably arise from periodic transcription because pre-mRNAs show a similar oscillatory gene expression and eventually lead to a rhythmic translation (Hendriks et al., 2014). However, the transcription factors that regulate this oscillatory behavior are unknown.

The periodic gene expression seems to be coordinated with the worm molting cycle (Kim et al., 2013), which every 8h induces formation of a new cuticle, shedding of the old one and progression to the next larval stage. However, it would be simplistic to conclude that the role of oscillating genes is only to drive molting, as only some of the rhythmic genes are coding for obviously molting-related proteins. Moreover, the diversity of phases suggests another role for some genes because if they were only involved in molting cycle, they would only peak around the molts.

Given that the experiments in (Hendriks et al., 2014; Kim et al., 2013) have been performed on whole worm lysates, it is possible that measurements of amplitudes are rather an underestimation. Different genes might indeed oscillate only in one or few cell types, without showing any rhythm in others. Therefore, the observed amplitude would be the sum of the hypothetical tissue-specific amplitudes dampened by the tissues where the genes are not oscillating. A similar problem applies to phases. The observed continuum of phases could be caused by the sum of different tissues, each oscillating with a certain phase. Alternatively, oscillating genes might not be enriched in any particular tissue and sport a diversity of phases. To gain insight into those questions, we combined information about oscillatory status of genes with information about tissue specificity.

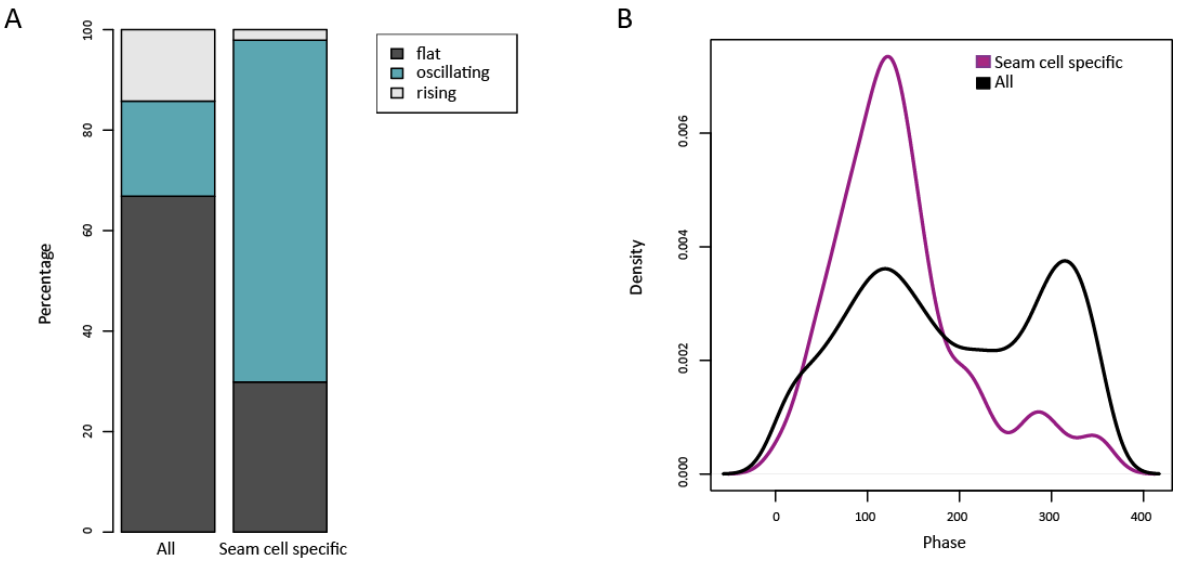


Figure 2.8 Most of the seam cell-enriched genes are oscillating (A) 18.9% of all *C. elegans* genes are classified as oscillating in contrast to 68% of the seam cell-enriched genes (hypergeometric distribution, p-value~0). **(B)** Phase distribution shows that the seam cell specific genes strongly peak between 25 and 225 degrees. (*All* = all *C. elegans* oscillating genes).

Starting from seam cells as the first test case, we addressed the question if there is any link between tissue specificity and oscillatory status. We tested if genes enriched in the L3 seam cells showed a higher or lower prevalence of oscillatory gene expression than expected in all the genes. While in total, 18.9% of the genes are classified as oscillating during development (Hendriks et al., 2014), in seam cells, 68% of the genes showed a rhythmic expression pattern (Figure 2.8A). This suggested that there is a link between tissue specificity and oscillatory gene expression. Furthermore, when examining the phases of those seam cell specific genes, we noticed that they tended to be strongly polarized and lied preferentially within a range of 25 and 225 degrees at the whole worm level (Figure 2.8B, purple line). Genes overall covered the full range of 0 and 360 degrees with some preference for the two phases 100 and 315 degrees (Figure 2.8B, black line).

The striking result that most genes in the seam cells have rhythmic gene expression has been partially validated generating some reporter strains harboring GFP driven by oscillating promoters, which indeed show oscillatory expression in the hypodermal tissue (Hauser and Grosshans, unpublished). This suggested that oscillating genes are enriched in a specific tissue, the seam cells, where they are not only abundant, but where they also show a distinctive behavior as exemplified by their “early phase” preference (25 < phase < 225).

To identify transcription factor motifs enriched in the promoters of seam cell-specific genes, we ran known and *de novo* motif finding using HOMER (analysis performed with Sarah Carl). The top-scoring *de novo* motif (p=1e-21) was predicted to correspond to the transcription factor BLMP-1, while the top three known motifs (p=1e-15, p=1e-14 and p=1e-13) were all variants of the NHR-23 motif.

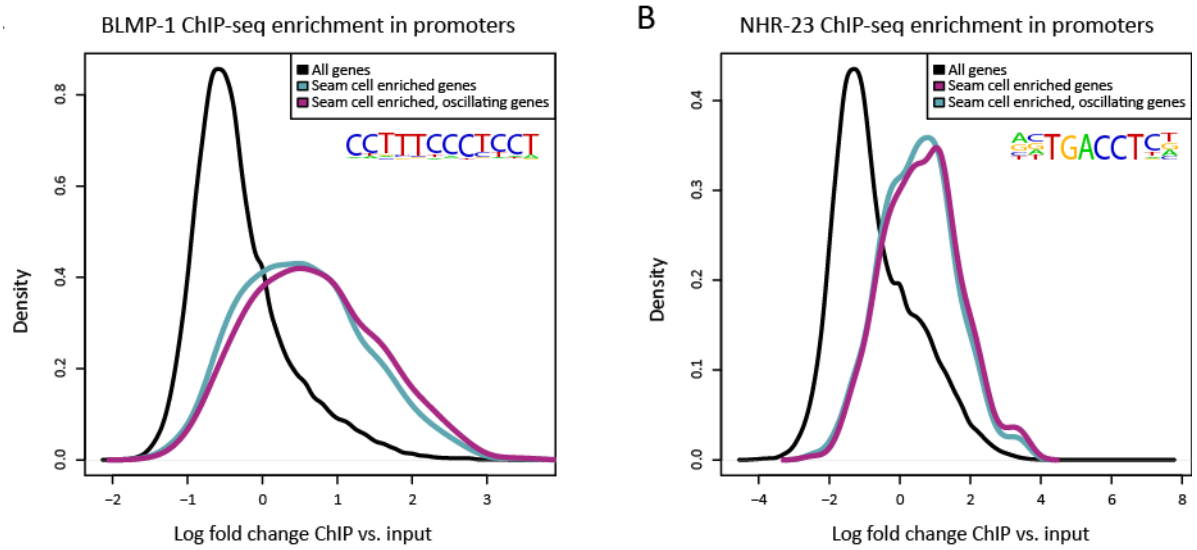


Figure 2.9 BLMP-1 and NHR-23 motif and ChIP peaks are enriched in the promoters of seam cell genes. Enrichment of ChIP peaks for the transcription factors (A) BLMP-1 and (B) NHR-23 in promoters of seam cell-enriched genes and seam cell enriched and oscillating genes. The motif logo for each TF is shown below.

Given the strong enrichments observed for motifs of these two factors, we downloaded ChIP-seq data for both BLMP-1 and NHR-23 in L3 larvae from modENCODE (Contrino et al., 2012). To visualize the distribution of ChIP enrichments in promoters, we first selected seam-specific promoters from all genes. Then, to ask if seam cell oscillating genes would show different transcription factors signatures than seam cell genes in general, we further selected oscillating genes from among the seam-specific promoters. The density functions of enrichments in each of these sets of promoters were then plotted for both BLMP-1 and NHR-23, showing a clear shift towards higher enrichments in both seam-specific and oscillating seam-specific promoters for both factors. We conclude that BLMP-1 and NHR-23 might drive the expression of seam cell enriched genes in general and oscillating seam specific genes, in particular.

Additional tissues that feature rhythmic gene expression

The discovery that seam cells are specifically enriched in genes that oscillate with early phase, induced us to extend our analysis to all the other tissues, which we had profiled, to ask if there is a second tissue that is oscillating with a "late phase" (225 < phase < 25). Moreover, to generalize the link between tissue expression and oscillatory status, we calculated the prevalence of oscillatory behavior in our FACS-sorted tissues (labeled as "GB") and in additional tissue specific gene sets obtained from (Spencer et al., 2011) and labelad as "L" (Fig. 2.10). Since we have previously seen phase preference in addition to amplitude, we included both of those readouts to categorize the genes into six groups. Two of those represented non-oscillatory expression (flat and rising during development) and four of those represented oscillatory expression, namely early phase (25 < phase < 225) with high amplitude (ampl > 1.5), early phase (25 < phase < 225) with low amplitude (0.55 > ampl < 1.5), late phase with high amplitude and late phase with low amplitude.

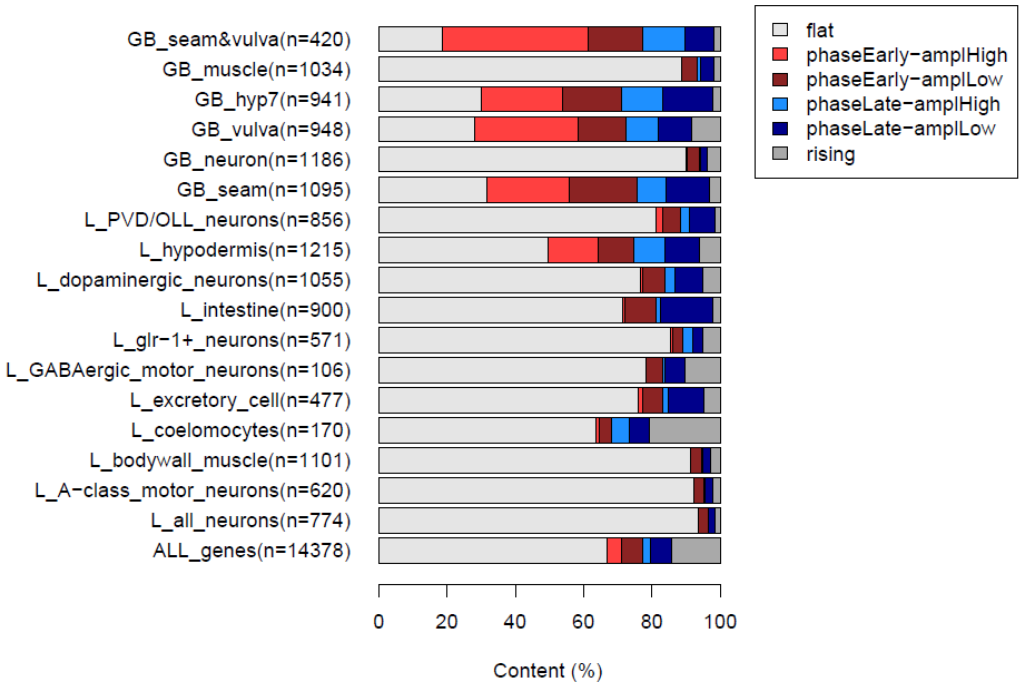


Figure 2.10 Oscillatory status of genes expressed in different tissues Content (%) of flat, rising or oscillating genes (split by phase and amplitude) in different tissue-specific transcriptomes. "GB" = FACS sorted tissues, "L" = data from Spencer 2011, "ALL_genes" = complete *C. elegans* transcriptome. N = number of genes in each tissue.

Figure 2.10 shows a quantification of the oscillatory status for all the genes expressed in a particular tissue. This revealed multiple tissues that showed a tendency to express genes with oscillatory expression during development in a phase-specific fashion, namely seam cells, hyp7 and

vulva, which as observed before are enriched for the early phase. Notably, some of the tissues showed substantial depletion for oscillatory gene expression. These were neurons and muscles. The fact that seam cells, hyp7 and vulva cells all together show this enrichment could be explained by lineage and because of their similarity at the level of the transcriptome (Figure 2.6). The issue of tissue similarity in this case poses a substantial challenge as it prevents us from exactly pinpointing the tissue with the greatest link to oscillatory gene expression. Concerning the late phase of oscillatory gene expression, to this point we have found no tissue that enriches for that fraction.

Taken together, this suggests that oscillatory gene expression during development originates from a potentially small set of tissues, possibly seam cells, hyp7 or vulva and is not taking place in neurons or muscle cells.

Paragraph discussion

Tissue specific gene expression analysis is a resource useful to reveal networks of genes that collaborate in determining cell identity or execution of cellular programs, development etc. Here, we used the data obtained through FACSeq to answer a question recently raised in the lab. In fact, others and we described that 20% of *C. elegans* genome has a rhythmic expression. However, we do not yet understand the mechanism. Moreover, because experiments have been performed on whole worm lysates it is not possible to know if the oscillating genes are specifically expressed in one tissue or if they are rather broadly expressed, and if the oscillation features, such as amplitude or phases, change across tissues.

Analysis of our sorted tissues and other public datasets (Spencer et al., 2011) revealed that oscillating genes are a feature of the hypodermal tissues, where most of the genes peak with one specific phase. De novo motif analysis, intersected with published ChIP peaks, revealed that the motifs of two transcription factors, namely BLMP-1 and NHR-23, are enriched in seam cell genes. Considering that most of the seam cell genes are oscillating (~70%), it is not surprising that these transcription factors are the same that preferentially bind to oscillating genes in general (S. Carl and H. Grosshans unpublished).

Our evidence suggests that oscillatory gene expression takes place in specific tissues, possibly seam/hyp7/vulva and not in tissues such as muscles and neurons. However, multiple scenarios could give rise to the oscillations that are observed during development. It could be the case that within

seam/hyp7/vulva only a small set of genes are transcribed rhythmically, which would be the genes that show particularly high amplitudes in the bulk experiment. An alternative hypothesis could be that there is a much larger number of genes oscillating in seam/hyp7/vulva than what can be detected in the bulk developmental time course. This could e.g. occur for genes that are not exclusively expressed in seam/hyp7/vulva but also present in another tissue. In this case, tissue bleed-through would result in dampening of the amplitude in the bulk. However, it is difficult to conceive that all the genes in a tissue, including housekeeping genes, might show such oscillatory behavior. Accordingly, preliminary data obtained through confocal imaging of two ubiquitously expressed reporters suggests that only a subset of genes is oscillating. For example, the ubiquitously expressed reporter *Pdpy-30::gfp::unc-54* does not show oscillatory gene expression in the hypodermis of L4 animals (Hauser and Grosshans unpublished). To investigate this issue further, a tissue-specific time course would be necessary. In fact, profiling of a tissue, e.g. the seam cells, in a time course manner would allow analysis of the transcriptome with the time resolution needed to appreciate the oscillations. Therefore, through FACSeq, we have isolated seam cells four times every two hours to cover one period of oscillations. Bioinformatics analysis is in progress.

In addition, considering that it is now possible to perform single-cell RNA sequencing (next paragraph and (Cao et al., 2017)), it could design an experiment aimed at obtaining the transcriptomes of every worm cell simultaneously. In fact, such "time course at single-cell resolution" experiment might reveal which additional tissues are responsible for the peculiar rhythmic gene expression observed during *C. elegans* development and extend our preliminary analysis (Fig. 2.10).

It has been speculated that oscillatory or sustained expression of certain factors can maintain cells in a self-renewal or differentiated states. In particular, rhythmic expression has been linked to the undifferentiated state (Bielefeld et al., 2017; Giachino and Taylor, 2014; Pfeuty, 2015). For example, neural progenitor cells in mouse are characterized by oscillatory expression of the Notch effector Hes1 and the pro-neural factor *Neurogenin2* (Imayoshi et al., 2013). In contrast, in differentiated neurons, Hes1 is not expressed and *Neurogenin2* acquires steady expression (Shimojo et al., 2008). In *C. elegans* oscillatory gene expression characterizes larval development and stops after the L4 stage, when the worm reaches adulthood (Hendriks, 2015). Additionally, we observed rhythmic expression in the hypodermal tissue and in seam cells in particular. Notably, seam cells are progenitor cells that divide in an asymmetric fashion once per larval stage and differentiate when the worm becomes adult. Hence, although this is pure speculation, it is intriguing to imagine that the oscillations of genes in the seam cells might be necessary to sustain their stem-like status during larval development. In this

scenario, the rhythmic expression observed in the whole animal would derive from this pulsing tissue. Intriguingly, in *lin-29(-)* mutants in which seam cells fail to differentiate (Rougvie, 2001), the oscillations do not stop at the L4 stage (Aeschimann, Hendriks, Carl and Grosshans unpublished).

In contrast to the hypothesis that the seam cells are the main pulsing tissue in the worm, our transcriptome analysis suggests that at least hyp7 and vulva cells are enriched in oscillating genes, too. However, seam, hyp7 and vulval cells are so similar in their transcriptomes (Figure 2.6) that we cannot distinguish if the seam cells have a stronger link to the oscillations than the other tissues.

2.3.3 Small RNA sequencing of seam cells at the L3 stage

Our initial attempt to profile miRNAs in different cell types was unsuccessful because the amount of RNA isolated from the sorted cells was too low to perform small RNA sequencing at that time (~1ug minimum input material). Nevertheless, we later collaborated with Dr. Dominik Jedlinski from the Zavolan lab at University of Basel, as the lab had established a protocol to prepare libraries from as little as 100ng total RNA. Therefore, by collecting 100,000 seam cells of L3 staged worms we were able for the first time to study the miRNA milieu of a tissue at a specific developmental stage.

For small RNA sequencing of L3 seam cells, we collected two biological replicates of seam cells and input (the cell suspension from which the seam cells derive).

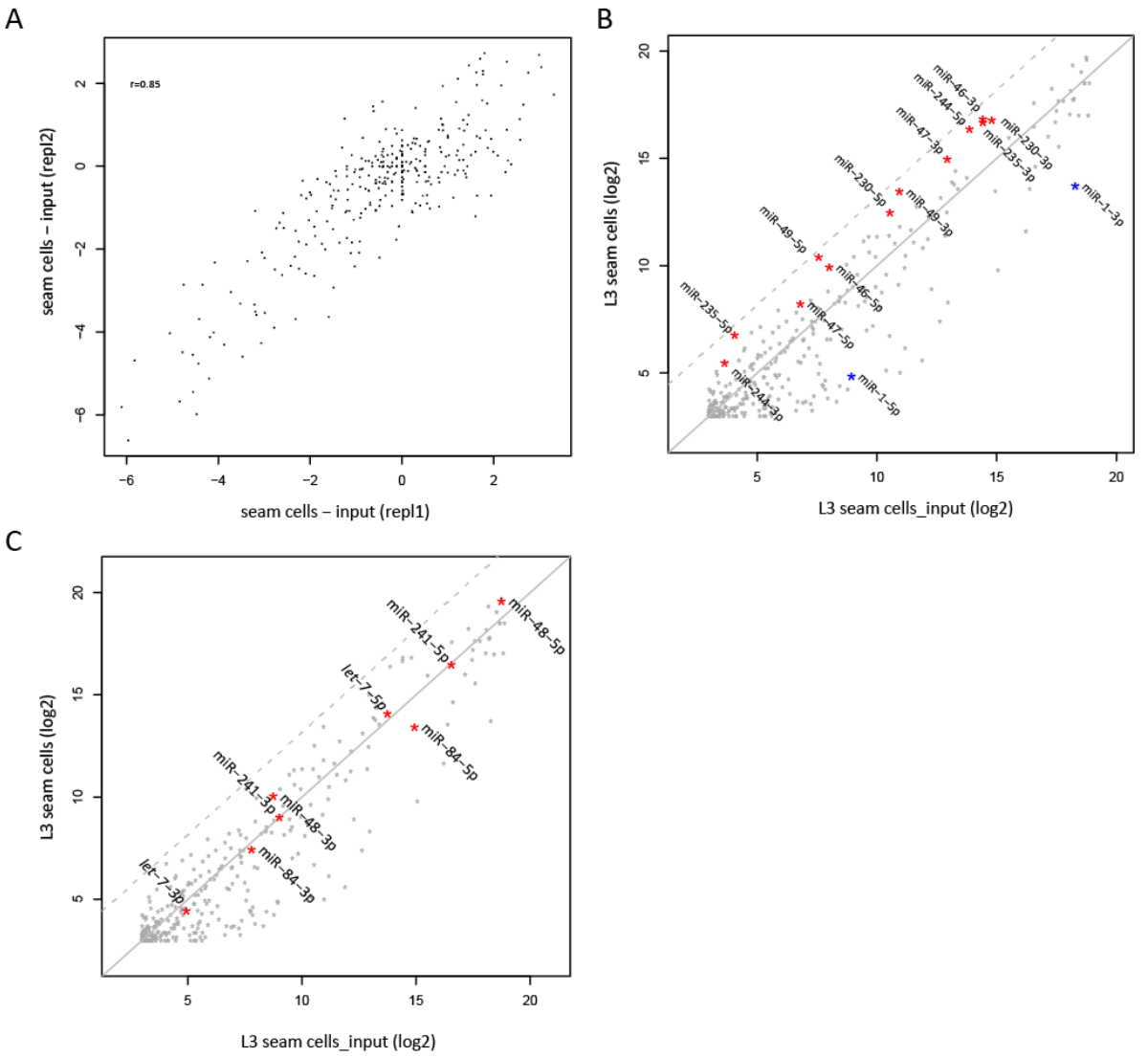


Figure 2.11 Small RNA sequencing of L3 seam cells (A) Expression change in two biological replicates. **(B)** Comparison of miRNA expressed in seam vs. input identifies six highly enriched seam cell miRNAs **(C)** Same as in (B), but highlighting the *let-7* family. Read counts (log2).

As shown in Figure 2.11A, the changes in expression are highly reproducible between replicates (R= 0.85). The top enriched miRNAs are shown in Figure 2.11B. The fact that we see upregulation for the mature as well as the passenger strand (miR*) provides further validation for the selected differentially expressed genes. In addition, this suggests that the high expression of those miRNAs in seam cells is likely caused by increased transcription of the miRNA loci. The six enriched

miRNAs are listed in the following table and for each of them the expression pattern, as inferred previously through transcriptional reporters (Martinez et al., 2008) is listed.

miRNA	**Expression** (Adapted from (Martinez et al., 2008))
miR-46	Expressed from early embryo continuing through adulthood. In the embryonic stage, expression is detected on the lateral sides. Later on, expression is seen in hypodermis an seam cells
miR-47	Expression detected from late embryos to adults. In embryos, expression is detected on lateral sides and in larval stages on, in all hypodermis, rectum and vulval cells.
miR-49	n.a.
miR-230	Expressed from L1 to adulthood. Strong expression in rectum, seam cells (L1-L4) and weak expression in hypodermis (not expressed in vnc and dnc). Also, expressed in vulva muscles
miR-235	Expressed from late embryos to adulthood. Expression detected in hypodermis, especially at L1-L2. Also detected in vulva, rectum and some amphid neurons.
miR-244	Expressed in seam cells from late embryos until adults.

In accordance with their enrichment, five of the six identified "seam-specific" miRNAs (no data are available for miR-49) have been previously shown to be expressed in the seam cells through transcriptional reporters (Martinez et al., 2008), thus corroborating their expression in this tissue. Interestingly, miR-46 and miR-47 are the only two members of a miRNA family. Even if their apparent seam-specific localization could be intriguing and suggest some function in this tissue, a double mutant animal does not show any obvious phenotype (Alvarez-Saavedra and Horvitz, 2010). Among the seam-depleted miRNAs we identified miR-1, which is known for being muscle specific in *C. elegans* and in higher eukaryotes (Figure 2.11B) (Andachi and Kohara, 2016; Simon et al., 2008). We also looked at the *let-7* family in the L3 seam cells and noticed that all the family members are quite abundantly expressed (Figure 2.11C).

In conclusion, through optimization of the FACS-sorting protocol, we were able to profile the miRNAs of seam cells at the L3 stage. Considering that many miRNAs are post-transcriptionally regulated, their quantification in the tissue of interest is necessary. Our results provide the first evidence of miRNA profiling in seam cells of *C. elegans*.

2.3.4 A single-cell worm atlas

Even if only a few years ago dissection of a whole worm in its constituent cells followed by single-cell sequencing would have sounded impossible, it is now reality. In principle, such experiments

allow recovery of all tissues simultaneously, without the need to sort specific cell types sequentially and therefore avoiding batch-effects. Moreover, unbiased clustering of cells based on their gene expression can lead to identification of new cell types and markers.

Single-cell sequencing protocols have improved dramatically and they allow sequencing of individual cells at a reasonable cost (Grün and van Oudenaarden, 2015). More recently, oil droplet based methods that ensure efficient barcoding have been described (Macosko et al., 2015; Zilionis et al., 2017). They are based on the formation of aqueous droplets that form after mixing a cell suspension with micro particles coated with barcoded polyT primers in oil. Although instructions for self-made chips are available, similar methods to encapsulate cells together with barcodes have been developed and commercial products are available. Figure 2.12 shows the strategy utilized by the 10X Genomics.

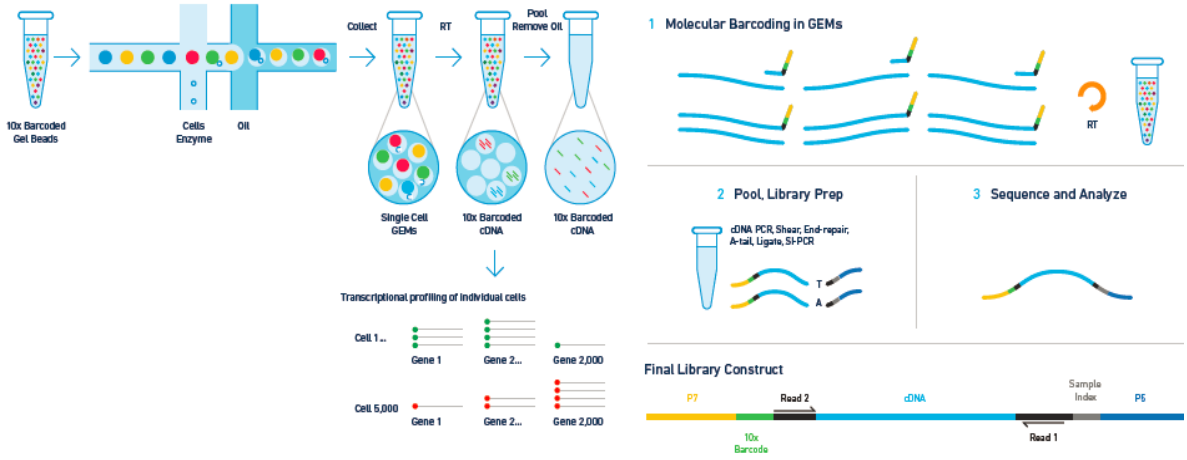


Figure 2.12 Chromium™ Single-cell 3' Solution. The cell suspension is mixed with gel beads containing individual barcodes and oil in a microfluidic device. Each "GEM" contains an encapsulated single cell, a barcode and the enzymes for reverse transcription. After the RT, pooled barcoded cDNA can be amplified and used for library preparation in bulk. Adapted from https://www.10xgenomics.com/

To perform our single-cell experiment using the 10X technology and create an "atlas of each cell of the worm", we collaborated with Guglielmo Roma, Global Head of Genomics Science & Technologies at Novartis. Our dissection protocol works best for L3 staged larvae, thus we grew worms for 24h at 25°C. To obtain a good single cell suspension and avoid debris, we used worms carrying a ubiquitously expressed GFP marker (*eft-3::gfp*) and FACS sorted the GFP(+) cells. In this way, we could also select

for viability through propidium iodide staining. We loaded about 4,000 of cells on five lanes of the cartridge and recovered 13,002 cells with a median number of UMI/cell of 4,109 and a median of 1,088 genes/cell, which is better than in (Cao et al., 2017) (mean = 1,121 UMIs and 431 genes per cell). Considering that the worm has only ~800 cells at the L3 stage, we oversampled the animal (16x).

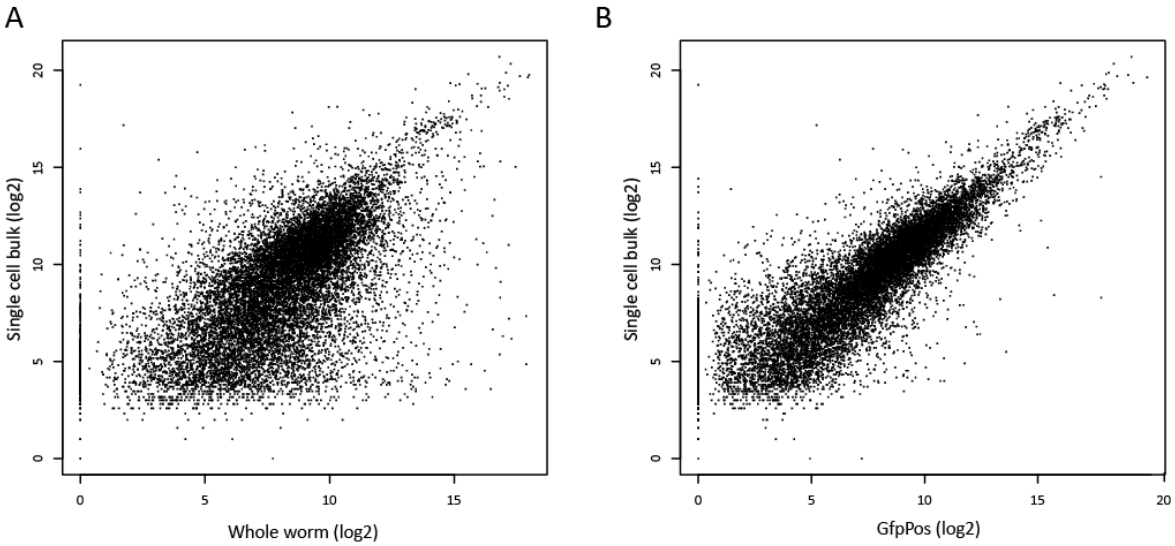


Figure 2.13 The single cells recapitulate a whole worm
Scatter plots showing the comparison between the single-cell bulk, obtained by aggregating all the single cells together, and whole worm from where they originated **(A)** or the GFP positive (GfpPos) cell suspension before it was subjected to droplet formation **(B)**. Shown are read counts (log2) for every gene.

As a reference, we sequenced RNA obtained from L3 larvae before starting the dissociation procedure (*whole worm*) and the GFP positive cells before the droplet formation (*GfpPos*). Both samples were profiled in bulk using the SmartSeq2 protocols. We then compared these reference datasets to an artificial bulk (*single-cell bulk*), which we obtained by aggregating together the transcriptomes of all the individual single cells (Figure 2.13). The comparison revealed overall a good correlation between the *single-cell bulk* and the *whole worm* sample, considering that they were prepared with two different protocols. The *GgfPos* sample, i.e. the cell suspension made up of all GFP positive cells, correlated better to the single cells, as expected given the fact that the two samples have gone through the same methodological procedure (except the droplet formation). In contrast, the *whole worm* sample contains some genes that are absent in the cell suspensions. We hypothesize that the "lost" genes belong to cell types that, like neurons, tend to be depleted in our preparations.

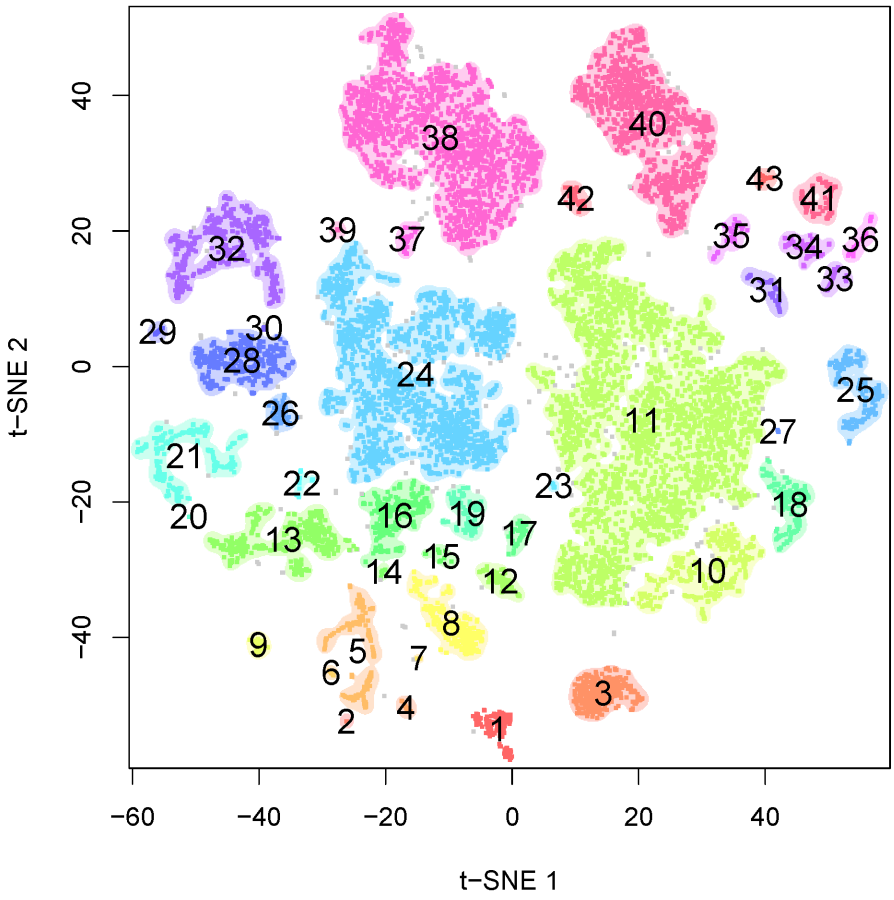


Figure 2.14 t-SNE map of 13,002 cells identifies 43 independent clusters

In order to identify individual tissues within the single-cell dataset, we performed dimensionality reduction of the single-cells gene expression profiles. Next, we clustered the cells by applying a 2D density estimate on the tSNE map and thresholding the result. Figure 2.14 shows a 2D projection using tSNE considering 13,002 cells. This revealed extensive separation into a large number of smaller clusters, which is expected given the large number of cell types in a whole worm. We obtained 43 clusters of variable size.

Figure 2.15 Projection of the 27 tissues (identified in (Cao et al., 2017)) on our preliminary t-SNE map

Given that a single-cell atlas of L2 larvae was recently published (Cao et al., 2017), we compared it to our dataset. Figure 2.15 highlights in our t-SNE map the 27 cluster identified in (Cao et al., 2017). Some clusters, like the cluster corresponding to the body wall muscles, seam cells or hyp7, can be easily identified. We noted, as our colleagues, that the muscles seem to be more abundant and more easily recovered than any other cell type (Cao et al., 2017). The remaining of the clusters will require finer analysis to be confidently assigned (work in progress).

Taken together, our results suggest that the FACSeq protocol can allow profiling of transcripts and small RNAs with tissue-specific resolution. Remarkably, it can also be successfully coupled to single-cell technologies.

3 Discussion

3.1 Towards a *C. elegans* single cell atlas

Although *C. elegans* is a great model organism that allows easy manipulation and elegant genetic perturbations, it has the main limitation of not allowing tissue-specific investigations. Some protocols have been optimized to profile the transcriptomes of different tissues, but miRNA profiling has not been described yet.

Compared to the other approaches based on nuclei purification or pulldown, we believe that our FACSeq method is very appealing for several reasons. First, it allows analysis of viable cells, thus avoiding artifacts due to dead or fixed cells. Second, in contrast to nuclear sorting, it allows sequencing of all isoforms of mRNAs, assuming they are expressed at reasonably high levels. Third, and most importantly for people interested in small RNAs, it allows simultaneous analysis of both mRNA and small RNA species, which are lost with the other protocols. Additionally, given that there are so many publicly available strains carrying fluorescent transgenes, which are specific to cell types or individual cells, it is potentially possible to FACS sort and profile all the cells of *C. elegans* without the need to create additional transgenic lines. We acknowledge that our dissociation protocol could be optimized further to reduce the time between the harvesting of the worms, their treatment and the FACS sorting. Currently, this takes about 4 hours and carries the risk of introducing unwanted changes in gene expression, which can affect the transcriptomic analysis. Moreover, as discussed previously, the transcriptome of the cell suspension is similar to a whole worm, but lacks some tissues that seem to be "lost" during the preparation. Alternative methods, possibly more gentle, might overcome this problem, even if we believe that some cells, like the neurons, will not be efficiently recovered from dissociated worms, due to their long neurite extensions.

Our initial analysis revealed the expression profile of different tissues (seam cells, hyp7, muscles and vulva) at a certain developmental stages (L3), and of the same tissue in two different stages (seam cells in L3 and L4). Because the samples were always treated in the same way, we could perform comparative analysis that identified putative tissue specific-genes. This led to identification of new cell-type marker genes that are particularly useful for the vulva, for which a gene that has exclusive expression is missing. We are currently generating transgenic animals that harbor promoters of some of these tissue-specific genes to confirm their expression profile.

We are also using FACSeq to investigate the gene expression changes in specific tissues of mutant animals. For this, we started a collaboration with Dr. Jan Padeken and Dr. Anna Mattout from the Gasser Lab. Our preliminary results are very promising and more experiments are in progress to validate them.

Additionally, we have been able for the first time to identify the miRNAs that are expressed in a particular tissue, the seam cells, at a certain stage in worm development, namely the L3. Considering that new protocols for small RNA-seq that start from low input amounts or even a single cell are now available, we believe that FACSeq can be coupled to those new methodologies to expand the analysis to more tissues, ideally with single cell resolution (Trilink and (Faridani et al., 2016)).

The differential analysis of distinct transcriptomes revealed similarities among tissues of hypodermal origins, but just hit the surface of more complex comparative studies that could be performed if all the different worm tissues were available and simultaneously recovered. Hence, we overcome the limitation of our own "bulk RNA-seq" by performing single-cell sequencing. This experiment has the great advantage that the data for each tissue are obtained in a unique experiment thus avoiding potential artifacts derived from the variability of different cell preparations performed on distinct days. Moreover, it greatly expands the number of tissues that can be sorted simultaneously and it allows identification of rare or unknown cell types.

Although our cluster assignment is still in progress, we are confident that we will soon appreciate the value of our newest data set. The recently published companion experiment, single cell sequencing of L2 worms, has in fact already shown the value of single-cell sequencing by identifying 27 cell types, whose transcriptome is now available for further investigations (Cao et al., 2017). Notably, these authors fixed the cells, as a necessary step in their library preparation protocol, whereas we sequenced RNA from cells that were selected for viability by FACS sorting. Whether the two approaches are similar or if one of the two reflects a more physiological condition is still under investigation. As a proof of principle, such approach can be extended to analyze mutant animals that lack a muscle-specific transcription factor, UNC-120. Analysis of such mutant, for which muscle-specific RNA-seq data are already available (Steiner and Henikoff, 2015), might reveal how cells in different tissues respond to the lack of a necessary transcription factor and how they rewire to compensate for the muscular defects.

In conclusion, we believe that the optimization of FACSeq is a milestone for the *C. elegans* community, which has so far lacked such tissue-specific resource. Here, we describe only some of the applications for FACSeq, namely "bulk" tissue-specific profiling, miRNA profiling and single-cell sequencing. We envision that a "single cell atlas of *C. elegans"* containing cell-type specific profiles of mRNAs, long non-coding RNAs, small RNAs, transcription factor binding profiles and chromatin modifications and accessibility will soon be reality. In fact, single-cell or low-input protocols for all these analysis are being optimized. In this "single-cell era", the worm is a very special animal: each cell is known, their lineage and division pattern have been described, and we know their position in time and space (Jorgensen and Mango, 2002). The possibility to enrich the knowledge we have on each worm cell with its own molecular milieu, which will help understanding the molecular physiology of the whole organism, is just a few steps away.

3.2 Rethinking miRNA target prediction and validation

The seed sequence of a miRNA has been traditionally considered sufficient and necessary for miRNA-mediated repression (Bartel, 2009) and in fact, functional studies confirmed such hypothesis (Brennecke et al., 2005; Lai, 2002; Lai, 2004; Lewis et al., 2005). Nevertheless, if such sequence were truly the only determinant for targeting, *in silico* prediction tools should not have high rate of false positives nor would miRNAs sharing the seed sequence have non-redundant activity. In addition, some miRNAs, such as *let-7*, are fully conserved from worms to humans, suggesting that each nucleotide might be relevant for function.

In the following paragraphs, we discuss what our in *in vivo* validation of non-canonical targets revealed about the modulation that site architecture and miRNA abundance exert on silencing. We dissected the *let-7* binding site of the *C. elegans lin-41* 3'UTR and found that while the 3' pairing establishes specific binding, the seed match has a key modulating role. Moreover, we discuss the advantages of physiological assays for target validation and how they have revealed new features of miRNA binding sites. Lastly, we discuss some challenges associated with miRNA therapeutics.

MiRNA target specificity within a miRNA family is determined by the 3' pairing and modulated by the seed match

It was hypothesized that miRNA specificity arises in the presence of an imperfect seed match compensated by extensive 3' distal pairing, as exemplified by the *let-7/lin-41* sites in *C. elegans*, and

that such “compensatory sites” are only rarely present in the genome (< 5% of all miRNA sites, (Bartel, 2009; Brennecke et al., 2005)). However, recent Ago-CLIP chimeric reads suggested that specificity arises thanks to only the 3’ pairing of the miRNA (Broughton et al., 2016; Moore et al., 2015).

To clarify this issue, we reanalyzed Ago-chimeras from (Broughton et al., 2016). We tested if sites with an imperfect seed match to a miRNA family were more specific to individual miRNAs than sites with canonical seed matches (Brancati et al., 2017). First, we found that in contrast to previous hypothesis, sites harboring a seed mismatch were as abundant as perfect sites in Ago-iCLIP chimeras. In addition, regardless of the amount of 3’ pairing and the seed match type, perfect and imperfect sites showed the same degree of specificity, as they were mostly bound by one miRNA family member rather than multiple ones. This suggests and confirms that 3’ pairing alone drives specific targeting within a miRNA family. However, if this were true, how can we explain that the *let-7* sites in the *C. elegans lin-41* 3’UTR carry a bulge and a wobble in the seed match, if the 3’ pairing might suffice in instructing specificity?

Our results suggest that the imperfect seed match is required to establish robust specificity. We found that the whole *let-7* family could repress sites with a perfect seed match, revealing redundant activity for the four miRNAs. However, they retained residual specificity towards the miRNA that could pair its unique 3’ sequence. In contrast, stronger specificity could be observed in sites carrying seed match “imperfections”, being it a central bulge (position 4-5 along the seed match) or a peripheral G:U wobble (position 8). Hence, in accordance with previous reports, we propose that the 3’ pairing is the first determinant for preferential binding of individual miRNAs belonging to a family. In addition, we propose a new role for the seed match in modulating the observed miRNA target specificity.

In light of such results, the name “compensatory sites”, often used to refer to sites with imperfect seed match and extensive 3’pairing, such as the ones in the *lin-41* 3’UTR of *C. elegans*, is misleading (Bartel, 2009; Brennecke et al., 2005). In fact, it implies that the additional pairing is needed to *compensate* for the imperfect seed match. In contrast, moving away the focus from the seed match, we propose that the role of the seed distal pairing in the *lin-41* 3’UTR is to determine specific repression by *let-7* and that the imperfect seed match confers a stronger specificity than a perfect seed match would allow. We refer to such sites as “centrally imperfect” because of the central bulge and wobble they have in the seed/seed match duplex.

MiRNA abundance can override specificity in some cases

It was also speculated that "compensatory sites" allow discrimination of changing miRNA concentrations (Brennecke et al., 2005). We provide a first evidence for such hypothesis, but find that *in vivo* the situation is more complex, as it does not only apply to compensatory sites only, but to all binding sites.

Initially, we observed specificity for sites with 3' seed distal pairing and a perfect seed match, although only marginal. In contrast, sites with imperfect seed matches displayed a stronger specificity, suggesting that they enhance preferential binding by individual miRNAs. Remarkably, the role of the seed match was more evident when we studied silencing after perturbation of the levels of a miRNA. In particular, when we over-expressed *let-7* 2-fold we could differentially modulate target specificity according to the architecture of the binding site. While a site with perfect seed match and extensive distal pairing to miR-48 was, under normal condition, specific to miR-48, it could be repressed by *let-7* when its level increased. Sites that contained a peripheral mismatch showed the same behavior. In contrast, sites with central imperfections were highly specific to miR-48 and did not allow abundant *let-7* to override the sequence-instructed specificity.

Hence, we propose that specificity of miRNA binding sites increases in this order: seed match-only sites (non-specific) < perfect seed match + distal 3'pairing < peripheral imperfect seed match + distal 3'pairing < central imperfect seed match + distal 3'pairing.

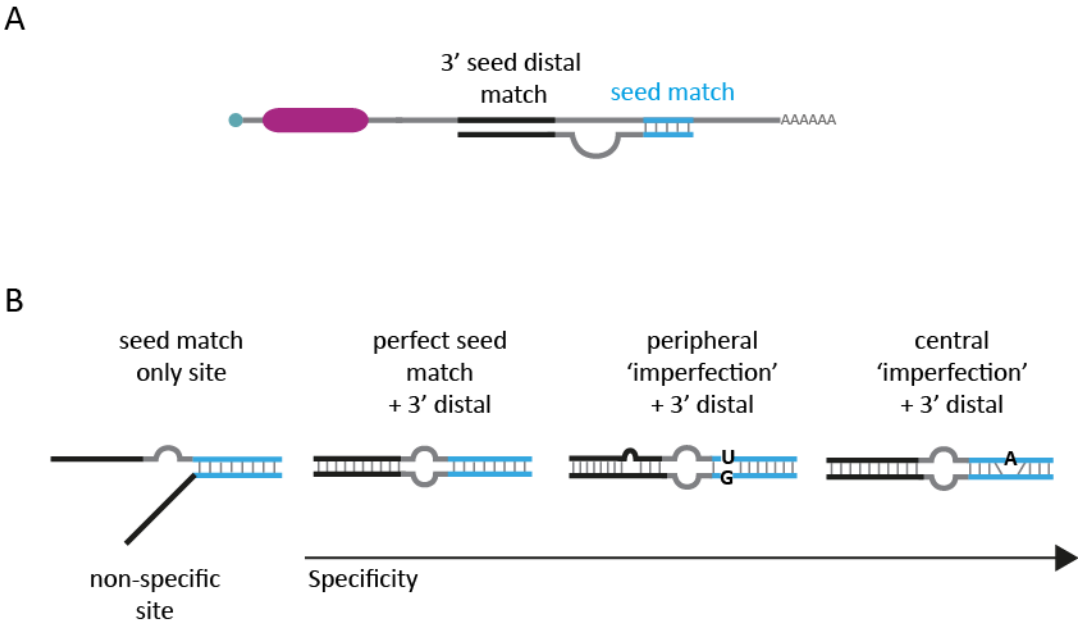


Figure 3 Different miRNA binding sites architectures. Careful analysis of the duplex is required when evaluating miRNA binding sites. (A) The seed match and the 3' (non-seed) distal match both have a role in silencing of an mRNA. (B) Seed match-only sites are not specific to any individual miRNA. The specificity of sites with 3' distal pairing increases according to the seed match. Sites with central imperfections are the most specific and are insensitive to miRNA levels.

Possibly, at lower miRNA concentrations, sites with perfect match or peripheral mismatches and extensive 3' pairing are bound by one specific miRNA, but as miRNA levels increase they cannot discriminate anymore among paralogues. We propose that this does not apply to imperfect sites that supposedly arose to "lock" silencing to one unique miRNA in any conditions. Hence, some sites might have evolved to mediate repression by a certain miRNA, but also engage in effective silencing with another highly abundant miRNA to ensure robustness, maybe in different cell types or under stressful conditions. In contrast, some other binding sites might be characterized by a more stringent specificity, possibly reflecting a key function and highly controlled regulation of the host gene. Such case is exemplified by the role that the imperfect sites in the *lin-41* 3'UTR have on *C. elegans* development (described below).

Taking together our results, we observed that imperfect sites seem as specific as the perfect ones in the captured Ago-chimeras, while *in vivo* robust specificity requires an imperfect seed match (Brancati et al., 2017). These observations might be harmonized considering the context in which miRNA/target interactions occur, that is often overlooked for simplicity. The chimeric reads in fact derive from whole worm lysates: although they can reveal interactions that happen *in vivo*, they lack the resolution to identify tissue-specific events. As discussed extensively, the *C. elegans* community lacked methods to assess the precise expression of miRNAs across cell types. Hence, we cannot exclude that different family members have distinct expression patterns and that miRNA-specific iCLIP reads derive from confined co-expression of the miRNA with the target, independently of the binding site quality. However, we have developed FACSeq, a technique that allows successful cells isolation and RNA sequencing. For example, we have already profiled the miRNAs present in one tissue, namely the seam cells, and found that there the four *let-7* family members are all expressed. Thus, we believe that soon we will be able to profile cell type-specific miRNAs in more cell types and address this issue.

Alternatively, it is possible that individual miRNAs or families have binding preferences, as shown for miR-30a and miR-30c or miR-122 (Luna et al., 2017; Moore et al., 2015). In this scenario,

some miRNAs might prefer imperfect bulged sites independently of specificity. For example, we have observed peculiar targeting behavior for distinct miRNA families (Brancati et al., 2017). While Ago iCLIP chimeric reads with imperfect sites exhibited preferential binding of individual miRNAs of the *let-7* and miR-72 families, promiscuous binding by the whole family was observed in the presence of perfect seed matched sites. In contrast, some other families, such as the miR-58, did not discriminate between perfect and imperfect sites and multiple miRNAs were bound to both kinds of site. It could be hypothesized that this might at least partially be explained by miRNA abundance. Highly expressed miRNAs are predicted to bind more non-canonical sites than lowly abundant ones, and display a broader spectrum of target sites (Khorshid et al., 2013). Notably, the *let-7* and miR-72 miRNAs, which exhibited specific binding, are among the most abundant miRNA in the nematode. However, the same is true for the miR-58 family, which in contrast is more promiscuous.

Taken together, our results affect the way miRNA targets are currently predicted. In fact, we describe silencing instances that happen *in vivo* only under certain conditions, but are difficult to identify with *in silico* tools that do not take into account the context in which the interactions occur. We believe that new mathematical predictions, trained on chimeric-CLIP data, are needed and should include differential expression and abundance of miRNAs, affinity of miRISC for canonical and non-canonical sites, target levels and their isoforms including alternative poly (A) usage that alters miRNA repression and RBPs that, in tissue-specific ways, can modulate site accessibility.

Additional experiments to confirm and expand our findings

In this study, we used the *let-7* family in *C. elegans* as a representative example, and showed, as a proof-of-principle, that over-expression of a family member could override target specificity and rescue the associated lethality, provided the seed match quality can allow for it.

To confirm that this happens globally, more experiments will be needed. For example, in the 3'UTR of the *akap-1* gene, three miRNA binding sites have recently been validated *in vivo*, two of them showing extensive 3' pairing to miR-791 and a 6nt match to both miR-790 and miR-791(Drexel et al., 2016). As we would predict, the 3' pairing establishes specific repression of *akap-1* by miR-791 in the CO_2 sensing neurons of *C. elegans*, although the sister miR-790 is co-expressed (Drexel et al., 2016). To further test our hypothesis and the effect that abundant miRNAs have on such specific binding sites, it

would be interesting to test if overexpression of miR-790 in such neurons, would restore *akap-1* downregulation in the absence of miR-791.

Considering how strongly the worm relies on *lin-41* repression to proceed into adulthood, additional experiments based on genome editing of the *lin-41* native sites could be designed. First, we could pursue a "candidate" approach by exchanging the *lin-41* sites with other sites putatively specific to another miRNA family, as identified in the Broughton dataset of chimeras. For example, assuming that the mir-72 family members are as abundant as the *let-7* family and have a similar expression pattern, especially in the vulva, we could insert miR-72 specific sites in the *lin-41* 3'UTR (with or without seed match bulges and complementary 3' pairing). Worm death, following miR-72 deletion, would be the expected outcome if specificity to this miRNA were indeed established. Overexpression of a sister and subsequent observation of the induced phenotype, similarly to what we observed for *lin-41*/*let-7* and miR-48 (Brancati and Grosshans, 2018), will then provide additional evidence against or in favor of our hypothesis that centrally mismatched sites are highly specific, and "locked", to one miRNA. Alternatively, a different approach could be pursued. MiR-48 mutant worms could be injected with CRISPR mixes that contain homologous recombination oligos to edit the *lin-41* sites. Oligos with different bulged nucleotides in distinct positions along the seed match and with diverse degrees of 3'pairing might be designed to induce and modulate miR-48 specificity. If such specificity is established, miR-48(-) animals will die and sequencing of their *lin-41* locus will allow identification of the nucleotide composition for the edited sites. Overexpression of *let-7* will then reveal sensitivity or insensitivity to increased levels of a paralogue.

However, some messengers could be differentially regulated across tissues. Although we do not believe that sites with centrally imperfect seed matches, which are insensitive to high miRNA level in the worm vulva, will be repressed in other cell types, we cannot exclude that tissue-specifically expressed RNAs or RBPs could influence site accessibility and, more generally, repression. To test this, a screen could be performed to identify "specificity factors" that can modulate specific targeting of miRNAs. We hypothesize that such factors might comprise RNA binding proteins (RBPs) that bind 3'UTRs, or transcription factors and RBPs that interfere with the biogenesis of the miRNAs. Candidates for such screen are genes that interfere with the specific *lin-41*/*let-7* axis and would drive *lin-41* repression towards the sisters. Such candidates could be the 25 hits identified as enhancers of precocious adult development in (Hayes and Ruvkun, 2006). Alternative candidates are the genes identified by our lab (Rausch et al., 2015) that suppress *let-7* lethal phenotype and can induce

precocious *col-19*::GFP accumulation, which is a typical sign of seam cell terminal differentiation that we would expect if the sisters were precociously repressing *lin-41*.

Lastly, assays in different models are required to confirm and extend our results to higher eukaryotes. For example, to test the interplay between site architecture and levels, mammalian cells could be stably transfected with reporters harboring different miRNA binding sites and increasing miRNA amounts. To test the specificity of the different binding sites (seed-only sites, perfect seed + 3'pairing to an individual miRNA, peripherally or centrally imperfect seed match + 3' pairing) the silencing of reporters should be evaluated in the presence and absence of the putative specific miRNA. Then, the system could be challenged with increasing amounts of a paralogous miRNA to test if, as described before, the over-expression can override the sequence-determined specificity in all cases, except the centrally imperfect sites. The limitations of such cell-based assay should anyway be kept in mind when drawing conclusions from such experiments.

3.3 Target validation in physiological conditions

Apart from revealing that miRNA target specificity relies on both the sequence complementarity and the miRNA levels, our data have a main implication, of possibly broader concern, regarding the methods commonly used to validate miRNA targets.

Predicted targets have been traditionally validated in heterologous contexts by introducing a reporter gene fused to the 3'UTR of interest, the miRNA or both, often through transient transfections (Brennecke et al., 2005; Doench and Sharp, 2004; Krek, 2005; Lewis et al., 2005). However, ectopic over-expression of miRNAs and their putative targets can induce artificial interactions between molecules that have the potential to bind. Remarkably, such effect has already been described for transfected reporters harboring a seed match to *let-7* that failed to be repressed unless more *let-7* molecules were provided exogenously (Doench and Sharp, 2004). Hence, such experimental systems could be far from physiological.

It was already proposed that miRNA levels and sequence complementarity together determine mRNA repression in different cells (Bartel and Chen, 2004). The rheostat model predicts that highly complementary sites can be bound by a miRNA, regardless of its levels, whereas low complementary sites (i.e. seed match only) would be bound solely by highly abundant miRNAs. In accordance with such overlooked model, our *in vivo* CRISPR assay and reporters analysis reveal that *let-7* overexpression can

repress a gene that is usually not silenced at the endogenous levels of the miRNA, similarly to what has been observed in cells. This is also consistent with recent mathematical and experimental studies describing how occupancy of non-canonical sites (such as 6nt seed sites) and miRNA binding overall increases at higher miRNA concentrations (Bosson et al., 2014; Denzler et al., 2014; Khorshid et al., 2013). Notably, in our case, the over-expression was only 2-fold, thus nearly physiological, and yet it induced an effect. Thus, we argue that if such small increase in miRNA levels can affect silencing *in vivo*, then some of the targets validated in systems based on ectopic over-expression of the miRNA need to be cautiously reconsidered.

Our results are an important piece of evidence that more physiological assays are needed to validate miRNA targets and corroborate the concerns already raised by the Hobert Lab in 2006 (Didiano and Hobert, 2006) and which we share. Particularly, experimental strategies that validate targets in non-physiological systems might have incorrectly scored as functional interactions that are not happening *in vivo.* This might have "biased" the way we think about miRNA targets.

Genome editing through CRISPR/Cas9 allowed us to study the effects that alteration of one specific miRNA/target interaction has on the physiology of the animals. In contrast to cell-based assays, such approach has the advantage of testing miRNA interactions in the context where they happen. For example, we revealed that two single point mutations, which are supposed to favor binding of the whole *let-7* family to *lin-41* 3'UTR, are deleterious for worm development (see below "*Why specificity matters*"). Such approach had already revealed its power when showing that *let-7* controls vulval development through one target, namely *lin-41*, against the broadly accepted belief that miRNA exert their function through downregulation of multiple targets (Ecsedi et al., 2015). Hence, we strongly believe that more studies focused on endogenous perturbations of miRNA/target interactions are needed to fully understand miRNA mediated silencing.

Physiological reporter assays in live animals

Additionally, to complement the *in vivo* data, we used a reporter assay that allows monitoring of miRNA activity with cell type resolution in living animals. Even in this case, we studied silencing in a physiological context, as we evaluate repression of modified *lin-41* sites in the cells where *lin-41* and *let-7* physiologically interact, namely the vulva (Ecsedi et al., 2015). Importantly, miRNA levels and the

cellular context are unaffected. Although we ectopically express fluorescent reporters fused to a 3' UTR harboring the binding sites of interest, we integrate such reporters in single copy. This is important if we consider that highly expressed targets can saturate the miRNA pool (Mukherji et al., 2011). We acknowledge that the system is imperfect because we might still over-express the transgene and reach higher levels than the real endogenous mRNA.

Our reporter approach (modified from (Ecsedi et al., 2015)) builds on the previous silencing sensor described by Didiano and Hobert (Didiano and Hobert, 2006; Didiano and Hobert, 2008). These authors studied miRNA targeting rules in nearly physiological conditions by looking at the repressive activity of the miRNA *lsy-6* in the left gustatory neuron (called ASEL) of *C. elegans*. *Lys-6* in fact is expressed specifically in such ASEL neuron, where it silences the gene *cog-1*. By modifying the *cog-1* sites and studying ~ 150 transgenes, they raise concerns about considering the seed match sequence as the main predictor for targeting (Didiano and Hobert, 2006). In fact, in contrast to *cog-1*, other 13 predicted targets harboring a seed match to *lsy-6* were not repressed. This approach is very elegant, but relied on the use of overexpressed multi-copy arrays. Although the authors studied the low-expressing transgenes, the lack of single-copy integration does not allow the full standardization necessary to compare different conditions. In addition, they are limited by the specific neuron (ASEL) and miRNA (*lsy-6*) that can be studied. In contrast, our approach is based on MosSCI single-copy integration and can be applied to other miRNAs. For example, by grafting the *lin-4* site present in the *lin-28* 3'UTR in our heterologous system, we could show that the reporter faithfully recapitulates *lin-4* specific binding and is insensitive to the seed-related miR-237 (Appendix A).

Finally, it was also shown that the context, i.e. the 3'UTR in which a miRNA site resides, is important for silencing (Didiano and Hobert, 2008; Grimson et al., 2007). This notion was already proposed specifically for the *lin-41* 3'UTR (Vella et al., 2004a), and for silencing in general (Grimson et al., 2007).

One of the observations reported about the importance of the 3'UTR context is, among others, the fact that the *lsy-6* sites found in the cog-1 3'UTR were unable to redirect silencing and induce repression when grafted in the heterologous *unc-54* 3'UTR (Didiano and Hobert, 2006; Didiano and Hobert, 2008). Surprisingly, the same sites could redirect silencing of other two unrelated 3'UTRs, namely *lin-28* and act-1. Hence, it was hypothesized that different 3'UTRs can have peculiar features

that allow or prevent miRNA mediated silencing. However, our own reporter system is centered on the successful engineering of the very same *unc-54* 3'UTR that we can "activate" for silencing by the *let-7* family or *lin-4*. Hence, we compared the published vectors to ours and found that a longer version of the *unc-54* 3'UTR was used, which likely induced the authors to inadvertently misplace the *lsy-6* sites. Expression in the ASEL neuron of our reporter carrying our engineered *unc-54* 3'UTR with grafted *lsy-6* sites would clarify this issue. We predict it will show that the *lsy-6* sites can redirect the *unc-54* 3'UTR to silencing by this miRNA, too.

The *let-7* sites in the *C. elegans lin-41* 3'UTR

It is worth mentioning that Vella and colleagues, who were the first to validate the *let-7* binding sites on the *lin-41* 3'UTR of *C. elegans* (Vella et al., 2004a), proposed also that such binding sites need to be in a very specific configuration for proper repression (Vella et al., 2004b). Specifically they suggest that both the two *let-7* sites are necessary for robust targeting and that they have to be spaced by the 27nt linker, as suggested by failed repression of the reporter carrying a mutated form of the *let-7* sites or the linker. However, such observations might need some more testing. In fact, we observed that exchanging the wild type 27-nt linker by the *C. briggsae* 12nt string did not affect silencing of our reporter (Appendix A, "Additional miRNA reporters"). Likewise, worms carrying a *lin-41* allele with only one *let-7* binding site do not show any evident phenotype, which we would expect if *let-7* regulation was compromised (characterization of such strains is in progress).

Although some regions of 3'UTRs can generally disfavor silencing, probably by modulating binding sites accessibility, we confirm that the miRNA sites and their peculiar molecular architecture have a main role in targeting. In fact, we could graft 111nt from the *lin-41* 3'UTR into the *unc-54* 3'UTR and induce silencing that strongly mimics the one of the wild type *lin-41* in time and space. Moreover, such repression was specifically *let-7* dependent, suggesting that the short region that we grafted contains all the information for robust and specific silencing. Additionally, when we exchanged the *let-7* sites with the *lin-4* sites from *lin-28*, we could observe specific *lin-4* silencing (Appendix A, "Additional miRNA reporters"). Hence, although RBPs could bind UTRs and modulate silencing, most of the information required to drive efficient repression is imparted by the miRNA binding site itself, at least in these two cases.

Taken together, other authors' observations and ours suggest that some of the targets solely predicted on the presence of a seed match in their 3'UTRs and validated in commonly used cell culture-based assays might need further validation. Such issue should concern not only future scientists who will have to face the challenge to validate new miRNA targets, but in particular people working on opening the way to miRNA-based therapeutics from the lab to the clinics. Generally, we believe that cell-based assays should be one of several experiments aimed to confirm functional interactions and that more studies testing the "rheostat model" (Bartel and Chen, 2004) and focusing on the interplay between sequence complementarity and miRNA levels are needed.

3.4 Why specificity matters: lessons from *C. elegans*

It has been hypothesized that the imperfect seed match in the *lin-41* 3'UTR is necessary to avoid redundancy of the *let-7* family (Bartel, 2009). Indeed, when we modify the sites to perfectly match the whole *let-7* family, *let-7* is not essential and its sisters can silence *lin-41* (Brancati and Grosshans, 2018). Moreover, it was believed that such particular sites, of which one with a bulge between nucleotide 4-5 the seed match sequence, rarely found in other transcripts or organisms (in humans for targets of miR-33 and miR-374 (Agarwal et al., 2015), evolved because of the *lin-41* function in worm development. Specifically, *lin-41* needs to be downregulated at an exact time to allow triggering of adult programs through accumulation of the transcription factor LIN-29 and its cofactor MAB-10. Accordingly, precocious or delayed *lin-41* repression results in heterochronic defects (Slack et al., 2000). To our surprise, two single point mutations that confer a perfect seed match to the *let-7* family induce premature repression of *lin-41* that results in precocious accumulation of LIN-29. However, such effect did not lead to any other obvious phenotypes, as if the induced LIN-29 accumulation was not sufficient to alter development to an extent that we could detect in the lab.

It is possible that gross defects were not detected because the imperfect sites are only one of many layers of regulation that protects the worms from developmental abnormalities. In fact, miRNA activity is usually part of regulatory networks in which complex feedbacks ensure robust regulations and buffering of fluctuations (Ebert and Sharp, 2012). In these cases, analysis of the network in a sensitized background, where a node is defective, can reveal the role of other network components (Li et al., 2009; Ren and Ambros, 2015). Similarly, we reasoned that the *lin-41(xe83 [perfect])* allele might not show obvious phenotypes, because other genes are involved in preventing the onset of adult development.

The *hbl-1* gene induces partially penetrant precocious adult phenotypes when knocked down on its own (Abrahante et al., 2003; Lin et al., 2003), but it leads to stronger defects in combination with *lin-41* RNAi. Similarly, we showed that sensitization via *hbl-1* knockdown leads to stronger and almost completely penetrant precocious adult phenotypes (scored as seam cell fusion and alae secretion) when in combination with the *lin-41(xe83 [perfect])* allele (Brancati and Grosshans, 2018). This reveals that the regulation through imperfect sites is a mechanism to ensure robustness. In addition, we want to test if the imperfect seed match generally confers robustness when different genes involved in the heterochronic pathways, beyond *hbl-1*, are mutated. Thus, we plan to test enhancement of precocious phenotypes in the *lin-41(xe83 [perfect])* animals after knock down of other genes that are known to induce premature adult development, such as *lin-4*2 and *lin-28*.

To explain why obvious premature development can be observed only in sensitized background, we hypothesize that even if *lin-41(xe83 [perfect])* is prematurely downregulated, the repression could be suboptimal to efficiently trigger adult development on its own. The *lin-41* downregulation by the sisters might be suboptimal because, even if the sisters are already expressed earlier than *let-7*, they are not highly abundant. In fact, their maximum accumulation occurs about the time when *let-7* is highly expressed (Vadla et al., 2012). Hence, although the sisters start repressing *lin-41*, the optimal downregulation level might be reached only in L4, as in wild type animals. Another explanation is that *lin-41* is dosage sensitive, such that only when its level decreases under a certain threshold, then it releases the repression on its targets, such as LIN-29 and MAB-10 (Slack et al., 2000, Aeschimann et al. 2017). Alternatively, LIN-29 or its cofactor MAB-10 might need to reach a certain concentration in the hypodermis, resulting in an on/off-like response to trigger precocious adult development. Supporting this model, we observed that LIN-29 accumulation was stronger in animals subjected to *lin-41* RNAi than in the animals carrying the *lin-41(xe83 [perfect])* allele. Similarly, we expect that worms carrying the *lin-41(xe83 [perfect])* treated with *hbl-1* RNAi will show higher levels of LIN-29 that precociously accumulates in the hypodermis. Lastly, we know that *LIN-41* specifically represses LIN-29A (Aeschimann et al., 2017) while the function of LIN-29B, the other isoform, is unknown. It is intriguing to believe that LIN-29B, whose regulation is independent of *LIN-41* (Aeschimann et al., 2017), might be partially redundant to LIN29A and thus responsible for the partially penetrant phenotypes of *lin-41* hypomorphic alleles.

In conclusion, our working hypothesis is that the imperfect seed match is needed to avoid precocious *lin-41* regulation by the sisters, which would trigger premature adult development if other

components of the heterochronic pathways were affected. Although the imperfect seed matches are not essential for viability, they are necessary to confer robustness to the heterochronic pathway. Notably, the imperfect seed matches in the *let-7* sites in the *lin-41* 3'UTR are typical of the genus *Caenorabditis*, but lost in the *trim71* 3'UTR in higher eukaryotes.

Many miRNA sites have been found to share pairing beyond the seed in conservation analysis (Brennecke et al., 2005; Grimson et al., 2007). Biochemical assays suggest that target specificity of miRNA family members could arise through 3' pairing (Broughton et al., 2016; Moore et al., 2015). Our results suggest that strong specificity develops through centrally imperfect seed match. We speculate that the *lin-41* sites might have had originally perfect seed match and extensive pairing to *let-7*. This would have allowed specific silencing by *let-7* in tissues where this miRNA is highly abundant. However, it would have also led to repression of *lin-41* in cells where the level of the *let-7* sisters is high enough to drive repression, as the perfectly matched sites can allow for some redundant activity. Later, the appearance of the bulge and the wobble would have marked the specificity of *let-7* for *lin-41* in any cell type or condition, such as when one of the genes of the heterochronic pathway is affected, essential to ensure robustness of developmental programs.

3.5 Open questions

Function of miRNA families

The presence and origin of miRNA families in animals, like the *let-7*, remains unclear. Seed-related miRNAs possibly originated through duplication, especially when the different isoforms are closely related to each other, such as the mammalian *let-7* family (Roush, 2008). After gene duplication, individual miRNAs might have been mutated and subjected to positive selection that led to sub-functionalization or neo-functionalization, maybe through divergence of their 3'sequence or change in their expression patterns (Berezikov, 2011; Wolter et al., 2016). It is unlikely that silencing by more miRNAs with the same seed sequence would result in stronger repression of a target, as one miRNA may drive efficient downregulation on its own. Hence, highly similar miRNAs might have arisen to ensure robustness (Ebert and Sharp, 2012) or to repress distinct targets following neo-functionalization. Redundancy of miRNAs with the same seed sequence is based on the assumption that the seed is enough for targeting and the observation that mutations of individual miRNAs induce

no phenotypes (Miska et al., 2007). For example, the miR-290-295 family (located in a cluster) represses retinoblastoma-like 2 protein in embryonic stem cells (Benetti et al., 2008; Sinkkonen et al., 2008) or the miR-51 or miR-35 family show redundant activity in *C. elegans* (McJunkin and Ambros, 2017; Shaw et al., 2010). However, some non-redundant activity has been described for miRNA sharing the same seed. In worms, for example, of the co-expressed miR-790 and miR-791, only miR-791 is needed for CO_2 detection in worms (Drexel et al., 2016) and only *lin-4* mutation and not miR-237 induces severe heterochronic defects (Lee et al., 1993; Tsialikas et al., 2017). Likewise, in mammals, only *let-7*i is targeting pro-neuronal differentiation genes in neural precursor cells (Cimadamore et al., 2013). Generally, the lack of consensus about miRNA families' function arises from our inability to predict miRNA targets with confidence.

Non-redundant miRNA families challenge the generalization of the seed sequence as the main determinant for targeting and let us wonder if we should not rather consider individual miRNAs as unique regulators with their own features, as proposed by Moore and colleague (Moore et al., 2015). Interestingly, we could show that *lin-41* specificity to *let-7* depends on the 3' pairing and on the imperfect seed match created by the target sequence (Brancati and Grosshans, 2018). Therefore, we suggest that both features in a duplex should be attentively studied to determine if a target is repressed by an individual miRNA, or redundantly silenced by miRNAs with the same seed sequence.

Reasons to study the *let-7* family in *C. elegans*

Studying the biology of miRNA families, the targets and regulation of each member in a model system such as *C. elegans* is not only helpful to understand miRNA targeting in general, but might hold therapeutic potential, too. In fact, it seems that 85% of conserved human miRNAs belong to a family (Wolter et al., 2016). Additionally, at least a third of *C. elegans* miRNA families is conserved across species (Miska et al., 2007).

Several parallelisms exist between *C. elegans* and mammals. For example, some *let-7* isoforms can be found misregulated in distinct cancers like individual *let-7* family members can be mis-regulated in worms, following *lin-28* mutation (Balzeau et al., 2017; Vadla et al., 2012). In addition, non-redundant and specific activity for some mammalian *let-7* isoforms has been described (Cimadamore et al., 2013) similarly to *lin-41* specific repression by *let-7* (Vella et al., 2004a). Lastly, over-expression of some isoforms can have different consequences on liver regeneration and tumor suppression (Wu et al., 2015), resembling the diverse phenotypes that we observed when increasing *let-7* doses

(Brancati and Grosshans, 2018). Hence, studying the *let-7* family in *C. elegans* allows investigations in a system that is easy to manipulate and in a physiological setting, and can lead to identification of transcriptional and post-transcriptional mechanisms that tightly control the expression of miRNA families or their target specificity. In addition, although evolutionary distant, studying worms holds the potential to identify mechanisms that can be translated to humans and bring miRNAs closer to the clinics, especially for those cancers in which the *let-7* family is affected (Thammaiah and Jayaram, 2016).

MiRNA-based therapies

Although some miRNAs are slowly finding their way to the clinic (www.clinicaltrial.gov), we still lack a general understanding of their targets. Our limitations derive from diverse issues, some specific to cancer research, others broadly applicable to other pathologies. First, miRNAs and their targets need to be studied in samples derived from patients, but identification of a "good miRNA candidate" is extremely hard because biopsies from cancer patients do not always mirror the tumor landscape and, in addition, hypoxia or inflammation can change gene expression locally (Rupaimoole and Slack, 2017). Second, even if a miRNA candidate is chosen, we do not understand enough of its biology. On the one hand, given that miRNAs are supposed to have many targets, they could repress a set of genes with opposite effect (like oncogenes and tumor suppressors or completely unrelated genes) making it hard to find the balance between favoring and diminishing the miRNA function through small molecules or by other means. Additionally, as we observed in worms and others described in mouse models of liver cancer (Wu et al., 2015), overexpression of a miRNA can be deleterious and induce repression of targets that under ordinary physiological conditions would be left untouched. Therefore, replenishing of a miRNA through its overexpression might not always be the most straightforward way to address its pathological loss. On the other hand, as extensively discussed, miRNA targets are largely unknown, thus precise disruption of particular miRNA/mRNA duplex is still not possible.

Nowadays, several CLIP-datasets from cancer cell lines or animal models have been deposited and they are useful together with other strategies, such as high-throughput screens, to identify and then validate relevant candidate targets. For example, Ago-CLIP of miR-122 mutant mice led to the identification of new targets (mostly harboring non-canonical binding sites) that could harbor predictive power for hepatocellular carcinoma patients (Luna et al., 2017). This suggests that such

target identification approach has therapeutic value. Such results are in accordance with our main conclusion that miRNA targets should be identified and validated with assays that resemble physiological conditions.

Final Remarks

Poorly validated miRNA targets might explain why we are still not able to predict targets with high confidence. Experimental validation is key to generalize results and identify common rules, but what if some of our conclusions are biased by the "over-expression" effect in the assays that have been used? What if we have over-interpreted the significance of the mere presence of a miRNA site? After all, transcription factors teach us that the presence of their motif on DNA does not always result in functional interactions (Slattery et al., 2014). Similarly, the presence of a seed match might not always be sign of targeting. This applies to the miRNA binding sites identified in the CLIP datasets, too. The simple evidence that Ago was found on a transcript does not imply functionality of the binding site, likewise the binding of a transcription factor to a site does not always modulate transcription.

Solving the conundrum about how to best predict miRNA targets is not easy. In fact, the seed-match seems consistently involved in miRNA-mediated repression. For example, proteomic studies performed after miR-223 deletion in neutrophils (Baek et al., 2008) revealed that seed-match containing sites were upregulated upon the miRNA removal. However, considering that upregulation of targets is usually minor, it remains to be determined whether it has any relevant functional outcome. Additionally, it was reported that the endogenous seed match-containing-messengers are upregulated after transfection of siRNAs or miRNA mimics due to competition of the transfected and the endogenous small RNAs (Khan et al., 2009), suggesting that, although gene expression is clearly altered after transfections, the seed match must have some roles in silencing. However, growing amounts of reports suggest that the seed match is not enough to predict targets.

In conclusion, there is a big amount of false positives among the *in silico* predicted seed-match-containing transcripts and potentially among the targets validated in systems characterized by ectopic expression of miRNA and/or targets. Hence, it is reasonable to wonder whether the so-called *bona fide* targets, predicted and validated through these approaches, are more informative than the non-canonical sites identified *in vivo,* or whether we should rewrite the rules of miRNA targeting to include

the variety of unconventional sites that have been identified. In fact, some of the best validated targets studied *in vivo* or in nearly physiological conditions carry non-canonical sites (*let-7*/*lin-41(Vella et al., 2004a)*, *lin-4*/*lin-14* (Ha et al., 1996), *lsy-6*/cog1 (Didiano and Hobert, 2006)), which have been shown to be generally more abundant than previously thought (Broughton et al., 2016; Helwak et al., 2013; Khorshid et al., 2013; Luna et al.; Moore et al., 2015).

As suggested before "*if there is one lesson to be learned from the history of the miRNA field, it is that one should not discard individual case studies as mere oddities*"(Didiano and Hobert, 2008; Ruvkun, 2004).

Some miRNAs are already in the clinics, but we do not seem to appreciate their mechanism of action with the precision needed to make miRNA therapeutics reality. Despite decades of research, the most fundamental question remains unanswered: what makes a miRNA target?

Materials and Methods

Worm handling and strains
Worms were grown using standard methods at 25 °C. The transgenic *unc-54+miRNA sites* reporter strains were obtained by single-copy integration into the *ttTi5605* locus on chromosome II (Frokjaer-Jensen et al., 2012). Injected plasmids were cloned using the MultiSite Gateway Technology (Thermo Fisher Scientific) and the destination vector pCFJ150 (Frokjaer-Jensen et al., 2008) or Gibson assembly (Gibson et al., 2009). All strains are listed in the "Worm Strains" table.

***Unc-54+miRNA sites* reporters**
All *unc-54+miRNA sites* reporters were constructed using the MultiSite Gateway Technology (Thermo Fisher Scientific) and the destination vector pCFJ150 (Frokjaer-Jensen, 2008) or Gibson assembly (Gibson et al., 2009). First, the pGB0 vector was obtained via site-directed mutagenesis (Zheng et al., 2004) of the pDONR P2R-P3_p37 vector to insert the AscI restriction site. Then, the pGB01 plasmid was obtained via LR reaction (Gateway LR Clonase II Enzyme mix, Thermo Fisher Scientific; 11791020) of the three entry vectors pdpy-30 x pGFP::H2B x pGB0 and the pCFJ150 backbone.
All the plasmids listed in the Plasmids table were obtained via Gibson assembly of the digested pGB01 plasmid and gBlocks® Gene Fragments (Integrated DNA Technologies) listed below. All plasmids were verified by sequencing. Transgenic worms were obtained by single-copy integration into the *ttTi5605* locus on chromosome II, following the published protocol for injection with low DNA concentration (Frokjaer-Jensen et al., 2012). We optimized our previous mCherry reference transgene (Ecsedi et al., 2015) by replacing the artificial 3'UTR with an endogenous *unc-54* 3' UTR, to achieve more physiologic and brighter expression. The resulting *Pdpy-30::mCherry::H2B::unc-54* transgene was integrated on chromosome I to yield strain HW1454.

***let-7*(++) strains:** *let-7(++)* strain (HW 1909 [*xeSi287*, V]) was obtained by injection of the plasmid pGB26, obtained via Gibson assembly (Gibson et al., 2009) of the PCR amplified minimal rescue fragment from (Reinhart et al., 2000) and the pIK37 plasmid. Transgenic worms were obtained by

single-copy integration into the *oxTi365* locus on chromosome V (universal MosSCI strain #EG8082 (Frokjaer-Jensen et al., 2014).

CRISPR strains

Mutations in the endogenous *lin-41* 3'UTR sequence or *miR-235* locus were obtained by CRISPR-Cas9 to generate the *lin-41(xe83 [perfect]) allele, lin-41(xe76 [ap427_W-C]), lin-41(xe99 [48-zed]) and miR-235(xe181).* Wild-type worms were injected with a mix containing 50 ng/µl pIK155, 100 ng/µl of each pIK198 with a cloned sgRNA, 20 ng/ µl repair oligo for gene of interest, co-crispr mix containing 100 ng/ml pIK208 and 20 ng/ml AF-ZF-827 oligo PAGE purified (IDT). Single F1 roller progeny of injected wild-type worms were picked to individual plates and the F2 progeny screened for deletions using PCR assays. After analysis by DNA sequencing, the alleles were outcrossed three times to the wild-type strain.

Reporter Quantification

For confocal assays, worms were grown at 25 °C. *Let-7ts* worms were maintained at 15 °C and adults were transferred to 25 °C for 48 h before imaging. Z-stacks of 0.313 µm µm thickness were acquired in green, red and transmitted light channels at 40x magnification on a Zeiss LSM700 confocal microscope coupled to Zeiss Zen 2010 software equipped with a multi-position tile scan macro. The z-stacks were stitched together and compiled into a single image using scripts in Matlab and Fiji (Schindelin et al., 2012).

For data analysis, late L4 worms were selected based on visual inspection of gonad length and vulva morphology (Mok et al., 2015). 10-14 vulva cells were selected in the 'cell counter' macro in Fiji. Images around these seed points were de-noised using a Richardson-Lucy algorithm and segmented using an Otsu global threshold. Remaining holes were filled using a morphological filter. Signal intensity in the green channel was divided by the red signal intensity for each cell; relative signal intensities were then averaged for each worm. 10-12 vulva cells in 5-10 worms per genotype were quantified; mean signal intensity and SD were calculated and graphed using GraphPad Prism software.

Confocal analysis of LIN-29 precocious accumulation

Synchronized arrested L1 larvae of animals carrying endogenously tagged LIN-29, *lin-29(xe61[lin-29::gfp::3xflag]*) (Aeschimann et al., 2017), in wild type or *lin-41(xe83[perfect])* background, were plated on food and incubated at 25 °C on 2% NGM agar plates with Escherichia coli OP50 bacteria and imaged at the L3 stage (20-22h after plating). Images were acquired in green and transmitted light channels (with Differential Interference Contrast, DIC) with 40x/1.3 oil immersion objective on a Zeiss LSM700 confocal microscope coupled to Zeiss Zen 2010 software. Further image processing was performed with Fiji (Schindelin et al., 2012).

Ago iCLIP reads analysis

Reproducible miRNA-target sites as defined in (Broughton et al., 2016) were extracted from supplementary table S2. Target sequences were retrieved from the UCSC October 2010 (ce10) genome assembly (Rosenbloom et al., 2015), using the BSgenome.Celegans.UCSC.ce10 package in R. Sites in 3' UTRs or CDSs were identified by intersecting all sites with annotated 3' UTRs or CDSs from the ce10 Ensembl gene annotations. MicroRNA family information, including mature miRNA and seed sequences, was downloaded from TargetScanWorm release 6.2 (Agarwal et al., 2015). Chimeras were predicted by calling RNAhybrid on all miRNA-target pairs using the command 'RNAhybrid -b 1 -c -s

3utr_worm' (Rehmsmeier et al., 2004). Perfect sites were identified by searching for an exact match to the corresponding seed in the predicted bound miRNA and target sequences for each chimera. Imperfect sites were defined as any site containing a single bulged nucleotide, a single G: U wobble or a single mismatch in the target between positions 2-8.
The number of paired 3' nucleotides for each chimera was determined by counting the number of nucleotides in the mature miRNA predicted to be bound by RNAhybrid downstream of the seed match. For imperfect sites where an exact seed match was not present, 3' paired nucleotides were considered as any predicted to be bound downstream of the 7th paired nucleotide after trimming an initial U if present (corresponding to position 8 of the seed).
RNA-seq data from *alg-1(-)* vs. WT were downloaded from the SRA (accession number: SRP078368). Reads were aligned against the ce10 genome assembly using the qAlign function from the QuasR R package (Gaidatzis et al., 2015) with default settings. Reads were counted in all annotated transcripts from the ce10 Ensembl gene annotations. Counts were normalized by the mean number of reads in transcripts in both libraries, and a log2 fold-change for each transcript was calculated between *alg-1(-)* and WT. Transcripts were considered to be targets of a particular miRNA if a corresponding miRNA-target site was found in either the 3' UTR or CDS, respectively. The empirical cumulative distribution of the log fold-change for each class of sites was calculated using the ecdf function in R.
Minimum free energy predictions were taken directly from RNAhybrid. For seed-only minimum free energy, RNAhybrid was run with the same parameters but using only the seed sequence as input for the miRNA. All computations were performed using R (version 3.4.0) in the RStudio environment (version 1.0.143).

MiRNA Promoter strains
The promoters *miR-241* and *miR-235* (~2kb upstream of the miRNA hairpin) were cloned from *C. elegans* genomic DNA into entry clones pGB27 & pGB28 and pGB33 & pGB34 using the Gateway BP Clonase II Enzyme mix (Thermos Fisher Scientific; 11789020). Entry plasmids containing the promoters, gfp::H2B and the pCM5.37 were then recombined with the pCFJ150 vector backbone (Gateway LR Cleanse II Enzyme mix, Thermos Fisher Scientific; 11791020) into pGB29, pGB30 and pGB35, pGB36. Transgenic worms were obtained by single-copy integration into the *ttTi5605* locus on chromosome II, following the published protocol for injection with low DNA concentration (Frokjaer-Jensen et al., 2012).

FACSeq
Worm dissociation: *C. elegans* strains carrying integrated fluorescent reporters (see strains table) and non-fluorescent wild type worms were used in all experiments. A synchronized L3 population was obtained by bleaching gravid adults and allowing the eggs to hatch in the absence of food to generate a population of starved L1 animals. Around 150,000 L1 larvae were plated on each 15cm peptone rich petri plates seeded with NA22 bacteria and grown at 25 °C for 24 hours (early L3 larvae), or 48h at 20 °C (late L4). Dissociated cells were recovered following a published protocol (Zhang, 2013) with modification. Specifically, L3 or L4 stage worms were collected by adding 10 ml sterile M9 buffer to each plate. The collected larvae were pelleted by centrifugation at 1300 g for 1 min. The larval pellet was let sediment on ice for 30min, washed and sediment again to remove bacteria. After washing with sterile ddH_2O, the resulting pellet was transferred to a 1.6 ml micro centrifuge tube. Around 80 µl of the final compact pellet was used for each cell dissociation experiment. The worm pellet was treated with 200 µl of SDS-DTT solution (20 mM HEPES pH8, 0.25% SDS, 200 mM DTT, 3% sucrose) for 4 min. Immediately after SDS-DTT treatment, egg buffer (25mM HEPES pH 7.3, 118mM Nalco, 48mM Kill, 2mM $CaCl_2$, 2 mM $MgCl_2$, osmolarity to 340±5 mOsm) was added to the SDS-DTT treated worms.

Worms were pelleted at 500 g for 1 min, and then washed 5 times with egg buffer). Pelleted SDS-DTT treated worms were digested with 200 µl of 15 mg/ml pronase (Sigma-Aldrich, St. Louis, MO) for 20 min. Pipette the larvae suspension in each tube and transfer singularly to individual wells of a 96-well plate and pipetted by the Hamilton Starlet robot (3 x 99strokes). Treated worms are then transferred to a fresh tube and treatment was finalized by pipetting the suspension with a modified glass Pasteur pipette. After 25min the reaction was stopped by adding 900µl L-15 medium (Thermos Fisher Scientific #21083027, supplemented with 10% fetal bovine serum and 50 U/ml penicillin + 50 µg/ml streptomycin (Sigma P4458); osmolarity 340±5 mom). The pronase treated worms were washed 3x with L-15 medium by centrifugation at 150g for 5min at 4°C. After the last centrifugation step, the samples were left on ice for 30min and the supernatant-containing cells (800µl) was filtered with 20 µm and 10 µm filters (Celtrics®) and transferred to a fresh tube. Cell were treated with DAPI or propidium iodide and sorted with a BD FACS aria with 70 µm nozzle.

RNA Sequencing: 10,000 sorted cells were lysed in 100 µl of Norgon lysis buffer (+ β-mercaptoethanol) (Norgen Biotek # 17200) and flash frozen in liquid nitrogen. Libraries were prepared following the SmartSeq2 protocol (Picelli et al., 2014) and then sequenced on an Illumina HiSeq 2500. The total RNA sequencing data was analyzed as previously described (Hendriks et al., 2014).

Transcription factor motif and ChIP peak Analysis on seam cell-enriched genes

Known and *de novo* motif finding was run using HOMER and comparing seam cell promoters (n=1,143) against promoters of all other coding genes (n=19,051). The set of known *C. elegans*-specific motifs was downloaded from the CIS-BP database (Weirauch et al., 2014).

Next, we downloaded ChIP-seq data for both BLMP-1 and NHR-23 in L3 larvae from modENCODE (Contrino et al., 2012). Reads were aligned against the ce10 genome using QuasR with default parameters (Gaidatzis et al., 2015). Aligned reads were counted in promoters of all protein-coding genes, defined as the 2 kb upstream of the TSS, and a log2 enrichment score of ChIP over input was calculated for each factor.

Small RNA sequencing

For small RNA sequencing of seam cells, total RNA was extracted from 100,000 sorted cells using TRIzol reagent (Thermofisher, #15596026) from two biological replicates (performed on different days). 100ng total RNA were used as input. The protocol was adapted from (Vigneault et al., 2001) by the Zavolan lab (University of Basel). Briefly, small RNAs (15-30nt) were separated on a 15% PAA gel 1XTBE buffer in RNA loading dye, after denaturation (30s at 95°C). A radiolabeled 5 bp ladder (low Molecular Weight Marker, Affymetrix) was used for size selection. Gel pieces from 15nt to 30nt were cut and RNA was elute from PAA gel overnight in 450 µl 0.4M NaCl at 4°C. 1 µl of glycogen-carrier was added followed by 1.125 ml of 100% ethanol; precipitation was done for 1h at -80°C. Samples were spun down at max speed at 4°C for 10 min; pellet was then washed with 70% ethanol and air-dried for 7-8 min. For ligation of 3'adenylated adapters, the pellet was dissolved in 8µl of the adapter-containing mix (1µl T4 RNAligase Buffer, 1µl 20uM 3-rApp-adpaters, 2µl 50% DMSO, 4µl RNA in H_2O). Denaturation was performed at 90°C for 30 sec and 4°C for at least 30 sec. 0.5µl RNase Inhibitor (40 U/µl) and 1.5µl T4 RNA Ligase 2tr (200 U/ µl) were added directly to the ligation reactions on ice and incubate overnight (18-20 h) at 4°C. Annealing of RT primer was done adding 1µl 20µM RT Primer to each reaction on ice, Incubating at 90°C for 30 sec, then 65°C for 5 min, then 4°C for at least 30 sec. For 5'RNA adapter ligation, 1.5µl 10 mM ATP, 1.5µl T4 RNA Ligase 1 (20 U/µl) and 1 µl of 20 µM 5' RNA Adapter (prepared by incubating ~5µl at 70°C for 2 min, then 4°C for at least 30 sec) were added to the ligation reaction. The ligation was perform incubating at 20°C for 1 h + 37°C for 30 minutes. Reverse transcription of captured MicroRNAs was done using Superscript III (200 U/µl, Invitrogen) and 15µl of ligated miRNAs. Pilot PCRs to determine optimal cycle conditions for library amplification were done

with increasing numbers of cycles (12-22). Large scale PCRs with the determined cycle number were loaded on an agarose gel and the amplified library (135bp-160bp) was excised from the gel and eluted in 400 µl pure H_2O overnight. For precipitation, 1 µl Glycogen and NaCl to a concentration of 0.3 M were added followed by 3 volumes of 100 % ethanol. Pellet was washed once with 70% ethanol, air-dried and dissolved in 20 µl H_2O. Libraries were ran on a bio analyzer and sequenced on Illumina HiSeq 2500. MiRNA expression levels were quantified as described before (Miki et al., 2014).

Single-cell RNA sequencing preparation

Synchronized L1 larvae carrying the transgene P*eft-3::gfp* (HW1057) were grown for 24h at 25 °C on 15cm peptone rich plates seeded with NA22 bacteria. Cell suspension was obtained following the cell dissociation protocol described previously.

GFP positive and propidium iodide negative cells were collected in cold L-15 medium. Cell number and viability were evaluated with the Countess II FL Automated Cell Counter (Invitrogen) and ~4,000 cells were loaded 5 channels of a 10X genomics cartridge. Library preparation was performed following the manufacturers' instructions (https://www.10xgenomics.com/).

Worm strains

Name	Genotype	Strain number (HW)
let-7ts	*let-7(n2853) X*	18
unc-54 (Ctrl)	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi104[Pdpy-30::GFP(PEST)-H2B::unc-54 3'UTR, unc-119 (+)] II*	1542
lin-41 reporter	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi78 [Pdpy-30::GFP(PEST)-H2B::lin-41 3'UTR, unc-119 (+)] II*	1529
lin-41 reporter let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I, xeSi78 [Pdpy-30::GFP(PEST)-H2B::lin-41 3'UTR, unc-119 (+)] II, let-7 (n2853) X*	2179
lin-41 reporter miR-48/241 (-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi78 [Pdpy-30::GFP(PEST)-H2B::lin-41 3'UTR, unc-119 (+)] II; miR-48/miR-241 (nDf51) V*	2268
lin41 miR-48/241/84(-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi78 [Pdpy-30::GFP(PEST)-H2B::lin-41 3'UTR, unc-119 (+)] II; miR-48/miR-241 (nDf51) V; miR-84 (n4037) X*	1964
unc-54+let-7 sites - gonads off	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi139[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 3'UTR, unc-119 (+)] II*	1316
unc-54+let-7 sites let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi139[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 3'UTR, unc-119 (+)] II; let-7 (n2853) X*	1306
unc-54+let-7 sites miR-48/241 (-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi139[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 3'UTR, unc-119 (+)] II; miR-48/miR-241 (nDf51) V; miR-84 (n4037) X.*	1326
unc-54+let-7 sites miR-48/241/84(-)	*xeSi176 [Pdpy30::mCherry::H2B::unc-54] I; xeSi139[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 3'UTR, unc-119 (+)] II, miR-48/miR-241 (nDf51) V; miR-84 (n4037) X.*	2266
unc-54+miR-48 sites wt	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi183[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 48ized 3'UTR, unc-119 (+)] II*	1305
unc-54+miR-48 sites let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi183[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 48ized 3'UTR, unc-119 (+)] II; let-7(n2853) X*	1300
unc-54+miR-48 sites miR-48 (-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi183[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 48ized 3'UTR, unc-119 (+)] II; mir-48(n4097)V*	1646
unc-54+miR-48 sites mir-48 (-) let-7(++)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi183[Pdpy-30::GFP(PEST)-H2B::unc-54 LCS 48ized 3'UTR, unc-119 (+)] II; mir-48(n4097) xeSi287 [Plet-7::let-7] V*	2267
let-7(++)	*xeSi287 [Plet-7::let-7] V*	1909
unc-54+lin-4 sites	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi403[Pdpy30::GFP-H2P-PEST::unc-54 3'utr with lin4 binding site wt] II*	2270
unc-54+lin-4 sites; lin-4 (-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi403[Pdpy30::GFP-H2P-PEST::unc-54 3'utr with lin4 binding site wt] II, lin-4(e912)*	2271
unc-54+lin-4 sites; mir-237(-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi403[Pdpy30::GFP-H2P-PEST::unc-54 3'utr with lin4 binding site wt] II; miR-237(n4296) X*	2272
unc-54+let-7 sites perfect seed	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi368 [Pdpy30::GFP-H2P-PEST::unc-54 3'utr LCSs with perfect seed match, unc-119 (+)] II*	2058
unc-54+let-7 sites perfect seed; let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi368 [Pdpy30::GFP-H2P-PEST::unc-54 3'utr LCSs with perfect seed match, unc-119 (+)] IIlet-7 (n2853) X*	2059
unc-54+let-7 sites perfect seed; miR-48/241/84(-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi368 [Pdpy30::GFP-H2P-PEST::unc-54 3'utr LCSs with perfect seed match, unc-119 (+)] IImiR-48/241 (nDf51) V; miR-84 (n4037) X #4*	2060
unc-54+miR-48 site perfect seed	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi404[Pdpy30::GFP-H2P-PEST::unc-54 3'utr 48z with perfect seed match, unc-119 (+)] II*	2273

wt (unc-54+miR-48 sites)		
unc-54+miR-48 sites_perfect; let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi404[Pdpy30::GFP-H2P-PEST::unc-54 3'utr 48z with perfect seed match, unc-119 (+)] II; let-7 (n2853) X*	2274
unc-54+miR-48 sites_perfect; miR-48/241/84(-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi404 [Pdpy30::GFP-H2P-PEST::unc-54 3'utr 48z with perfect seed match, unc-119 (+)] II; miR58/miR-241(ndf51) V; miR-84(n4037) X*	2275
unc-54+dot-1 sites site	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi375 [Pdpy30::GFP-H2P-PEST::unc-54 dot-1.1zed (ap427)], unc-119 (+)] II*	2151
unc-54+dot-1 sites; let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi375 [Pdpy30::GFP-H2P-PEST::unc-54 dot-1.1zed (ap427)], unc-119 (+)] II; let-7(n2853) X*	2142
unc-54+dot-1 sites; miR-48(-)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi375 [Pdpy30::GFP-H2P-PEST::unc-54 dot-1.1zed (ap427)], unc-119 (+)] II; miR-48(n4097) V*	2143
unc-54+dot-1 sites; miR-48(-) let-7(++)	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi375 [Pdpy30::GFP-H2P-PEST::unc-54 dot-1.1zed (ap427)], unc-119 (+)] II; mir-48(n4097) xeSi287 V*	2276
unc-54+let-7 sites 13-1648zed	*xeSi176[Pdpy30::mCherry::H2B::unc-54; unc-119+] I; xeSi405 Pdpy30::GFP-H2B-PEST::unc-54 3'utr with LCS 1/2 (lin-41) LCS 13-1648zed II;*	2269
lin-41(xe83)	*N2;lin-41 perfect seed (xe83) I*	2103
lin-41(xe83); let-7ts	*N2; lin-41 perfect seed (xe83) I; let-7 (n2853) X*	2138
ap427	*N2, lin-41(ap427) I (PQ570)*	(Broughton et al., 2016)
ap427 miR-48(-) let-7(++)	*N2, lin-41 (ap427), I; miR-48(n4097) V xeSi287 [Plet-7::let-7] V;*	2102
ap427GU	*N2; (xe76) lin-41 (ap427, GU corrected at position 8) I*	2069
ap427 GU miR-48(-)	*N2; (xe76) lin-41 (ap427, GU corrected at position 8) I; mir-48 (n4097) V*	2140
ap427 GU miR-48(-) let-7(++)	*N2; (xe76) lin-41 (ap427, GU corrected at position 8) I; mir-48 (n4097) xeSi287 [Plet-7::let-7] V*	2117
ap427 GU let-7ts	*N2; (xe76) lin-41 (ap427, GU corrected at position 8) I; let-7 (n2853) X*	2139
lin-41(48zed)	*N2; (xe99); lin-41(48zed, 3'UTR to contain miR-48 specific sites (48zed as the reporters = specific to mir-48)*	2228
LIN29::GFP	*N2, lin-29(xe61[lin-29::gfp::3xflag]) II*	1822
LIN29::GFP, lin-41(xe83)	*N2, lin-41 perfect seed (xe83) I; lin-29(xe61[lin-29::gfp::3xflag]) II*	2147
Plet-7::GFP	*xeSi117[Plet-7::GFP(PEST)-H2B::unc-54 3'UTR, unc-119 (+)] II*	1191
Pmir-48::GFP	*xeSi100[Pmir-48::GFP(PEST)-H2B::let-858 3'UTR, unc-119 (+)] II*	1223
Pmir-241::GFP (2kb)	*xeSi406[Pmir-241(esquela)::GFP::H2B::PEST::unc-54] II*	2278
Pmir-241::GFP (1.3kb)	*xeSi407[Pmir-241(1.3kb)::GFP::H2B::PEST::unc-54] II*	2279
Pmir-84::GFP	*xeSi61[Pmir-84::GFP(PEST)-H2B::let-858 3'UTR, unc-119 (+)] II*	1127
Plet-7t; miR-48/241/84(-)	*xeSi117[Plet-7::GFP(PEST)-H2B::unc-54 3'UTR, unc-119 (+)] II; mir-48/mir-241(nDf51) V; mir-84(n4037) X [HW1191 x 803]*	1963
vt1189	*maIs140 [mir-241p::GFP + unc-119(+)]*	1645
PmiR-235::GFP (~1.5kb)	*xeSi408[Pmir-235long(pGB33.2)::GFP::unc-54] II*	2280
PmiR-235::GFP (~1kb)	*xeSi409[PmiR-235(pGB34.1)::GFP::H2B::unc-54] II*	2281
scm::gfp	*unc119(e2598) III; lin-15(n765) X; mjIs15[ajm-1::mCherry::LIN-15]; wIs51[scm::gfp, unc-119(+)]*	647

scm::gfp; lin-41(xe83)	*unc119(e2598) III; lin-15(n765) X; mjIs15[ajm-1::mCherry::LIN-15]; wIs51[scm::gfp, unc-119(+)]*	2144
dpy-5:GFP	*ttTi5605, xeSi339 [pDpy-5 - npp-9::GFP::BLRP::3xFLAG - unc-54 3'UTR] II*	1978
rab-3::GFP	*jsIs682[rab-3p::GFP::RAB-3; pJM23]*	1374
egl-17::mCherry	*rdvIs1 [egl-17p::Myri-mCherry::pie-1 3'UTR + egl-17p::mig-10::YFP::unc-54 3'UTR + egl-17p::mCherry-TEV-S::his-24 + rol-6(su1006)] III.*	RDV55
myo-3::RFP	*GW76, gwIs4[myo-3::RFP baf-1::GFP lacI let-858]*	GW76 (Towbin et al., 2012)
Peft-3::GFP	*ttTi5605; oxSi221[eft-3::GFP] II*	1057
unc-54+let-7 sites + C.br linker	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi209[Pdpy30::GFP-H2P-PEST::unc-54 3'utr LCS with linker from C. briggsae, unc-119 (+)] II*	1378
unc-54+let-7 sites + C.br linker; let-7ts	*xeSi176[Pdpy30::mCherry::H2B::unc-54] I; xeSi209[Pdpy30::GFP-H2P-PEST::unc-54 3'utr LCS with linker from C. briggsae, unc-119 (+)] II; let-7 (n2853) X*	1379
miR-235(-)	*miR-235(n4504) I*	1072
miR-235 seed mutant	*miR-235(xe181) I*	2643

Plasmids

Name	**Plasmid**	**Insert**	**Cloning technique**	**Cloning Reagents (Oligos, plasmids or References)**
pCM1.151	pENTR_L1-L2	*mCherry-H2B*		(Merritt et al., 2008)
pCM5.37	pENTR_R2-L3	*unc-54 3'UTR*		(Merritt et al., 2008)
pdpy-30	pENTR_L4-R1	*Pdpy-30*		(Ecsedi et al., 2015)
pFA173	pCFJ210	*Pdpy30::mCherry::H2B::unc-54 3'UTR*	LR reaction	pdpy-30 x pCM1.151 x pCM5.37
pGB0	pENTR_R2-L3	*unc-54 3' UTR with AscI site*	Site directed mutagenesis	Based on (Zheng et al., 2004)
pGB01	pCFJ150	*Pdpy30::GFP::H2B-PEST::unc-54 3'UTR with AscI site*	LR reaction	pdpy-30 x pGFP::H2B x pGB0
pGB02	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with LCS 1/2 (lin-41)*	Gibson	pGB01 x LCSwt
pGB07C	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with LCS 1/2 (lin-41) 48ized*	Gibson	pGB01 x LCS 48ized (ATT bulges)
pGB11A	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with LCS 1/2 (lin-41) linker from C. briggsae*	Gibson	pGB01 x LCS brig linker
pGB19	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with LCS 1/2 (lin-41) LCS 13-1648zed*	Gibson	pGB01 x GBgblock01
pGB26	pIK37	*Plet-7::let-7 (Promoter + minimal rescue fragment from Slack)*	Gibson	GB108/GB105
pGB27	pDONR_P4-P1	*Pmir-241 (esquela Kercher boundaries)*	BP reaction	GB116/GB117
pGB28	pDONR_P4-P1	*Pmir-241 short*	BP reaction	GB117/GB118
pGB29	pCFJ150	*Pmir-241(esquela)::gfp::H2B/Pest::unc-54*	LR reaction	pGB27 x pGFP::H2B x pCM5.37

pGB30	pCFJ150	*Pmir-241(1.3kb)::gfp::H2B/Pest::unc-54*	LR reaction	pGB28 x pGFP::H2B x pCM5.37
pGB31		*Plin-48::GFP (pTJ1157)*		(Johnson et al., 2001)
pGB33.2	pDONR_P4-P1	*PmiR-235long_deleted tRNAs*	BP reaction	GB123/GB124
pGB34.1	pDONR_P4-P1	*PmiR-235 (NO blmp1 site)*	BP reaction	GB124/GB125
pGB35.2	pCFJ150	*PmiR-235(long)::GFP::H2B/PEST::unc-54*	LR reaction	pGB33.2 x pGFP::H2B x pCM5.37
pGB36.2	pCFJ150	*PmiR-235(short)::GFP::H2B/PEST::unc-54*	LR reaction	pGB34.1 x pGFP::H2B pCM5.37
pGB37	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with lin4 binding site wt*		pGB01 x GBgblock06
pGB45	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with LCS 1/2 (lin-41) 48ized perfect seed*	Gibson	pGB01 x GBgblock09
pGB46	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with LCS 1/2 (lin-41) LCSs perfect seed*	Gibson	pGB01 x GBgblock10
pGB48	pIK198	*sgRNA targeting lin-41 LCS1 (same as in Broughton 2015)*	Gibson	LCS1 sgRNA sense/LCS1 sgRNA antisense in NotI site
pGB49	pCFJ150	*Pdpy30::GFP-H2P-PEST::unc-54 3'utr with dot1.1 sites (as in ap427)*	Gibson	pGB01 xGBgblock11
pGFP::H2B	pENTR_L1-L2	*GFP(PEST)-H2B (pBMF2.7)*		Wright et al., 2011 (Wright et al., 2011)
pIK155		*Peft-3::Cas9::tbb-2 3'UTR*		(Katic et al., 2015)
pIK208	pIK198	*sgRNA targeting dpy-10 (sgRNA sequence GCTACCATAGGCACCACGA)*		
p*lin-41*sgRNA		*pU6::lin-41sgRNA*		(Ecsedi et al., 2015)

Oligos

Name	**Purpose**	**Sequence**
GB66 *n2853* fwd	to genotype *let-7* locus_66/68 (n2853) 67/68 (wt)	tacactgtggatccggtgaca
GB67 wt fwd		tacactgtggatccggtgacg
GB68 rev		atacagttcttgcgactccga
GB97	To sequence the *unc-54* 3'UTR	agagctccgcatcggccg
GB98	*mir-48(n4097)* deletion genotype	ggtagcaccacgttattgaatgaaac
GB99		caccttgatgacgataccatcgg
GB105	amplifies the minimal rescue *let-7* fragment (Reinhart et al., 2000) flanked by pIK37 sequences for gibson	gcgtgtcaataatatcactcggtaccctccctcttttaagcctg
GB108		atccagtcactatggcggccctgaaaactaaaaacactaacaaagaattg
GB109	genotype of chr V: 109/110 (transgenic band) and 109/111 (wt band)	ccgtcgaagacccaataaga
GB110		tatcgtaaatcggcgcgagc
GB111		atgaaaggcaagcgtgaact
GB116	Pmir-241 cloning (long)	ggggacaactttgtatagaaaagttggtgtcgtcgtgttctaaatgttcc

GB117		ggggactgctttttgtacaaacttgcactttgacacccccgcggtttg
GB118	Pmir-241_ 1.31 kb (short)	ggggacaactttgtatagaaaagttggcataccttctcagcttcaccag
GB124	Pmir-235_ GW attB1r	ggggactgctttttgtacaaacttgctgatcctgatatcttcggacg
GB125	Pmir-235 (short) GW_ GW attb4 fwd	ggggacaactttgtatagaaaagttgtcattcatcgttatgtcattgtg
GB128	*mir-235(n4504)* genotyping	tctccttctttggcagtcgatc
GB129		gcaaacaagatgggcgtactc
GB130		gggcggaccttcaattggtac
GB142	to amplify *let-7*(++) inside the universal mossci site on chr V	gcccgggcgtgtcaataatatc
GB143		atatccagtcactatggcggcc
GB149	Amplifies the LCSs in *lin-41* 3'UTR	ttcccatccattcatatggc
GB153		taaaattgggtgcgcaagaag
GB155	to genotype *lin-41(xe76)*, use with gb149	ggtttcaatggttcatgaggtug
GB160	to genotype *lin-41*(xe99)_use with gb149	gaggcagaattcagtagatgc
mosSCI_fw_outside	5' integration mossci site (chr II or universal)	cag aat gtg aac aag act cg
mosSCI_rv_unc-119		caaggacttggataaattggc
mosSCI_fw_GFP	3' integration of GFP reporter in chr II mossci site	ggc cgt cac caa gta c
mosSCI_rv_outside		ggaggcgaacctaactg
Right rec II rv	To genotype wild type chr II (680bp in wt; 1970 with Mos1)	tgaatttggcttgtaacgcgga
Left rec fw		agacgacgagccacttgctca
mir-48 mir-241 fwd	genotyping of *mir-48/241(nDf51)*_fwd/rev (mutant) fwd/rev int (wt)	ttgggtttgttttggctctc
mir-48 mir-241 rev		cgttcgcactctctgttctg
mir-48/241 del int rev		cagatgtgtgtagacggcaaag
mir-84 fwd	genotyping of *mir-84(n4037)*, amplifies a shorter product in the mutant, Ta=52C (wt = 1694bp, about 905bp if mutant)	gcaacgggaagctctgttac
mir-84 rev		aagtatcattcagcttcaattttgtc
right rec I fw	genotyping of wild type chr I	atagctctctcgcatactcgaattcc
left rec I rv		cgaggtgtcggccgctaaa

gBlocks and other oligos used for CRISPR editing

Description	Name	Sequence
LCS1 & 2 wild type	LCS wt	ttctcttaatttctttgtGGATTGCACCAACTCAAGTATACCTTTTATACAACCGTTCTACACTCA ACGCGATGTAAATATCGCAATCCCTTTTTATACAACCATTCTGCCTCTGAACCATTGAAAC CTTCCGCGCCTtttttagcttctttttaagtc
48zed sites (site1 mutated to att bulge - wt has gtt)	LCS 48ized (ATT bulges)	ttctcttaatttctttgtGGATTGCACCAACTCAAGTATACCTTGCATCTACTGAATTCTACACTC AACGCGATGTAAATATCGCAATCCCTTTGCATCTACTGAATTCTGCCTCTGAACCATTGA AACCTTCCGCGCCTtttttagcttctttttaagtc
LCSs 13-16nt 48zed_att bulge	GBgblock01	ttctcttaatttctttgtGGATTGCACCAACTCAAGTATACCTTTTATtactCCaTTCTACACTCAA CGCGATGTAAATATCGCAATCCCTTTTTATtactCCATTCTGCCTCTGAACCATTGAAACCT TCCGCGCCTtttttagcttctttttaagtc
LCS 1-2 48ized_ATTbulges_Perfect seed	GBgblock09	ttctcttaatttctttgtGGATTGCACCAACTCAGTATACCTTGCATCTACTGAATTCTACCTCAA CGCGATGTAAATATCGCAATCCCTTTGCATCTACTGAATTCTaCCTCTGAACCATTGAAAC CTTCCGCGCCTtttttagcttctttttaagtc
LCS 1-2 wt_perfect seed	GBgblock10	ttctcttaatttctttgtGGATTGCACCAACTCAAGTATACCTTTTATACAACCGTTCTACCTCAA CGCGATGTAAATATCGCAATCCCTTTTTATACAACCATTCTaCCTCTGAACCATTGAAACC TTCCGCGCCTtttttagcttctttttaagtc
dot1.1 sites in LCS environment containing sites exactly as in ap427 (pGB47&Gblock 11 are slightly different) for reporter	GBgblock13	ttctcttaatttctttgtGGATTGCACCAACTCAAGTATACCTTtttgcatctgaactcccttacctcatCGC GATGTAAATATCGCAATCCCTTTttgcatctgaactcccttacctcatGAACCATTGAAACCTTCC GCGCCTtttttagcttctttttaagtc
LCS1&2(lin41,Celegans)with 12nt from C.briggsae_for Gibson	LCS brig linker	ttctcttaatttctttgtGGATTGCACCAACTCAAGTATACCTTTTATACAACCGTTCTACACTCA aaaacacccctTTATACAACCATTCTGCCTCTGAACCATTGAAACCTTCCGCGCCtttttagctt cttttaagtc
lin-4 site(lin-28)_wt	GBgblock06	ccccctatttttgttattatcaaaaaacttctcttaatttctttgtGGcctcctcaaattgcactctcagggattctttttt tttttcaaaCGCGCCtttttagcttctttttaagtcacctctaacaatgaaattgtgt

Oligos		
Description	**Name**	**Sequence**
dpy-10 repair oligo for co-crispr mix	AF-ZF-827 (Arribere et al., 2014)	CACTTGAACTTCAATACGGCAAGATGAGAATGACTGGAAACCGTACCGCAT GCGGTGCCTATGGTAGCGGAGCTTCACATGGCTTCAGACCAACAGCCTAT
miR-235 sgRNA	mir235 sg sense	GCTGGAA
	mir235 sg antisense	TTGCAATT
sgRNA from Broughton, 2016 to cut LCS1 (modified from oligos A3427&3428)	LCS1 sgRNA sense	AATTGCAAATCTAAATGTTTttctacactcaacgcgatgtGTTTAAGAGCTATGCTGG AA
	LCS1 sgRNA antisense	TTCCAGCATAGCTCTTAAACacatcgcgttgagtgtagaaAAACATTTAGATTTGCA ATT
mir-235 CRISPR seed mutant (seed mutation in both mir-235 and mir-235*)	mir235 HR oligo 2	gatggttcttaattttctatctattccttattcttccaagtgctatctattcgtttgctgaaaatcgtccgaa gatatcaggatcaggccttggctgattTACCaattgttcaccgtgaaaattaaataGGTAactctcc ccggcctgatctgagagtaaggcgaagctgaattgactttaattcgacaggtttgttttt
perfect seed match in LCS1&LCS2 in lin-41 3'UTR (as in GBgblock10)_sense	lin41 HR oligo_perfect seed_sense	AAGTATACCTTTTATACAACCGTTCTACCTCAACGCGATGTAAATATCGCAAT CCCTTTTTATACAACCATTCTaCCTCTGAACCATTGAAACCTTCTCCCGTACTC CCACCAATAGATTATTGCACTTTTCTGAGAGTTTTTCTG
dot1.1 sites in LCS environment for CRISPR - GU corrected to GC	GB lin41 HR oligo1	GAATTAAAACACCCACAATAGCACCTCTTTTCCTCAAATTGCACCAACTCAAG TATACCTTtgcatctgaactccc**C**tacctcatCGCGATGTAAATATCGCAATCCCTTTt tgcatctgaactccc**C**tacctcatGAACCATTGAAACCTTCTCCCGTACTCCCACCAAT
dot1.1 sites in LCS environment for CRISPR - with A bulge and GU wobble - endogenous GU corrected to GC	GB lin41 HR oligo2	GAATTAAAACACCCACAATAGCACCTCTTTTCCTCAAATTGCACCAACTCAAG TATACCTTtgcatctgaactccc**C**tac**A**ctcatCGCGATGTAAATATCGCAATCCCTTT ttgcatctgaactccc**C**t**G**cctcatGAACCATTGAAACCTTCTCCCGTACTCCCACCAA TAGATTATTGCACTTTTCTGAGAGTTTT
lin-41(48zed) allele, sites with *imperfect* seed matches	48zed site HR oligo	AAGTATACCTTGCATCTACTGAATTCTACACTCAACGCGATGTAAATATCGC AATCCCTTTGCATCTACTGAATTCTGCCTCTGAACCATTGAAACCTTCTCCCG TACTCCCACCAATAGATTATTGCACTTTTCTGAGAGTTTTTC
lin-41(48zed) allele, sites with *perfect* seed matches	48zed site perfect HR oligo	AGTATACCTTGCATCTACTGAATTCTACCTCAACGCGATGTAAATATCGCAAT CCCTTTGCATCTACTGAATTCTaCCTCTGAACCATTGAAACCTTCTCCCGTACT CCCACCAATAGATTATTGCACTTTTCTGAGAGTTTTTCTG

Appendix A (Additional results)

a) MiR-235: a seam cell specific miRNA

Of the six seam-enriched miRNAs, miR-235 caught our attention due to a recent publication that describes its involvement in blast cell quiescence (Kasuga et al., 2013).

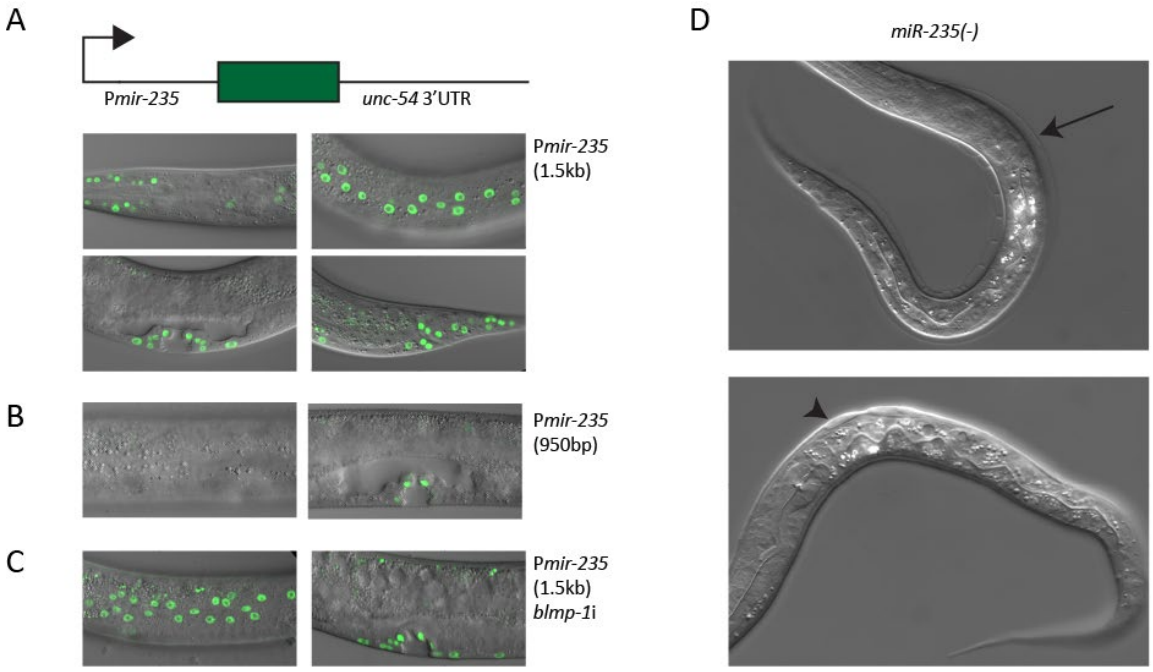


Figure 4.1 miR-235 characterization (A) Confocal analysis of a P*miR-235*::GFP reporter confirms miR-235 hypodermal expression and extends it to head, tail and vulva. **(B)** A reporter with a shorter putative P*miR-235* sequence loses only hypodermal expression. **(C)** Worms carrying the same reporter as in (A) but treated with blmp-1 RNAi show no altered GFP expression. **(D)** Starved L1s lacking miR-235 (*miR-235 (-)*) grown in poor nutritional medium supplemented with cholesterol remain trapped in their cuticle (top panel) or display over-proliferation of unidentified cells (bottom).

To confirm that miR-235 was indeed expressed in the seam cells, we generated two transgenic lines carrying a putative promoter, nuclear GFP and unregulated *unc-54* 3'UTR. A ~1.5kb promoter drove GFP expression in the head and in about five cells close to the anus, likely neurons, in L3 and L4 stage. GFP expression was detected also in the vulva, starting already in precursor cells in L3, as well as in the hypodermis, both hyp7 and seam cells, where it started in young larvae and remained high throughout development (HW2280, Figure 4.1A).

However, when we used a shorter putative promoter sequence (~1kb, HW 2281) not only GFP hypodermal expression was low until adulthood, but also GFP was undetectable in seam cells at any

stage (Figure 4.1B). Outside of the hypodermis, the expression pattern was unchanged. This suggested the presence of a hypodermal-regulatory element in the 500bp sequence that differentiates the two reporter strains (Figure 4.1A and B). We then scanned the promoter sequence of miR-235 for matches to the blmp-1 motif and nhr-23 using the matrix-scan tool from RSAT (Thomas-Chollier et al., 2008) (performed by Sarah Carl). The analysis indicated the presence of a putative site for the transcription factor BLMP-1 (Horn et al., 2014), which is known for being expressed in the hypodermis and whose ChIP peaks are often found in seam-enriched genes (see "2.3.2 *Oscillatory gene expression*"). However, we could not detect any change in the expression of the transgene containing the long promoter when BLMP-1 was depleted via RNAi (HW2280, Figure 4.1C) or when using a null *blmp-1* allele (data not shown), suggesting that another factor must be involved in the hypodermal expression of miR-235.

MiR-235 couples regulation of blast cells and nutritional cues

Usually worms go through four stages of larval development and molt before progressing to the next stage. However, this happens only in the presence of food. In fact, after hatching in a food-deprived environment, worms arrest development to enter a quiescent state (L1 arrest) until food is available again. In contrast to wild type, animals lacking miR-235 do not arrest when grown in nutritionally poor condition, but seem to molt and enter the L2 stage, as indicated by several markers: *lin-4*, *lin-42* and *mlt-10* accumulation, absence of L1- alae, precocious blast cell development (P cell ventral migration and M cell division) (Kasuga et al., 2013).

Surprisingly, when we repeated the experiment described by Kasuga and colleagues, which involves growing synchronized L1 larvae without food for 3 days, we made the unexpected observation that miR-235 mutant worms grew differently in different solutions. When grown in S-medium + cholesterol, we noticed that most of the L1 mutant worms had the cuticle attached, probably about to shed it, while some others had patches of over-proliferated cells whose origin was unknown (Figure 4D, bottom panel). On the contrary, wild type worms were normal and showed none of these phenotypes. Mutants grown on S-medium without further cholesterol addition did not shed the cuticles, but were rather trapped in it close to their mouth. The same was true for worms grown in S-basal. Interestingly, *miR-235(-)* larvae in S-basal supplemented with cholesterol showed yet a different phenotype. In fact, they looked rather dry or "dissolved" in the medium, as if they shed the cuticle without having a new one to protect them. Lastly, when miR-235 mutant worms were grown in M9 buffer, which is the equivalent of PBS for cells, their tissue degenerated, whereas when cholesterol

was supplemented, they either dried out (as observed for S-basal growing worms) or got trapped in their cuticle (as in S-medium). Wild type worms looked normal or slim or had a swollen pharynx, at most, in all conditions.

In conclusion, our data confirm and extend the function of miR-235 in controlling blast cell quiescence (Kasuga et al., 2013). The phenotypes of miR-235 mutant animals in nutritionally poor media supplemented with cholesterol confirms that the miRNA is important to stop development of starved larvae and avoid improper L1/L2 molting in stressful environments, where animals would not survive. The presence of unshed cuticle or of worms somehow "dissolved" in the media suggests that the animals tried to molt, but then failed, as if their molting timing was wrong. Moreover, we hypothesize that the different phenotypes that miR-235 mutant worms show in the different solutions (M9, S-basal and S-medium) reflects a function of this miRNA in osmolarity sensing. We predict that among the direct and indirect targets of the miRNA there could be cuticle proteins or components of sensory organs, whose misregulation would affect the ability of the worm to sense and respond to change in the osmolarity of the environment. However, this needs further testing.

Putative miR-235 targets

To find putative miR-235 targets we performed RNA sequencing of miR-235 null animals and compared their transcriptome to wild type. Because we found miR-235 enriched in seam cells at the L3 stage, we specifically looked at this stage. To avoid artifacts due to the oscillating gene expression of the worms (Hendriks et al., 2014), which could lead to misleading gene expression profiles when analysis is limited to one time point, we sequenced RNA extracted from worms harvested at 24h, 26h, 28h and 30h of development at 25°C and then combined in equal amounts for library preparation (in collaboration with Stephane Thierry, FMI genomic facility).

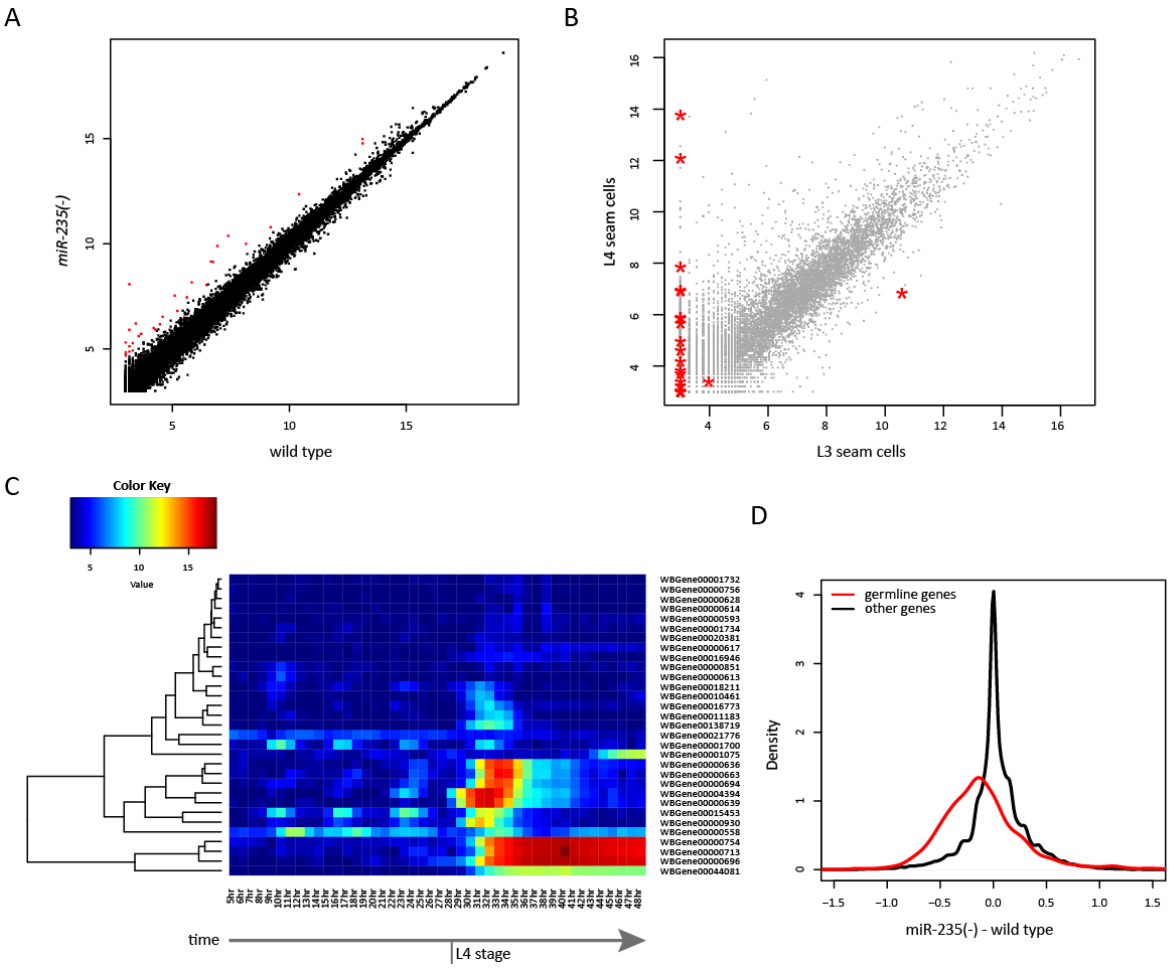


Figure 4.2 Genes upregulated in *miR-235 (-)* mutants (A) Comparison of *miR-235 (-)* mutant and wild type animal identifies 33 upregulated genes. **(B)** The genes upregulated in *miR-235 (-)* mutants are enriched in seam cells and specific to L4 animals. **(C)** The genes upregulated in *miR-235 (-)* mutants are high in L4 larvae or adult wild type animals. Time course from 5h to 48h after hatching at 25°C. **(D)** The miR-235 mutant animals are younger than the wild type counterpart is. Germline genes (red line) expression increases as the animal ages.

Comparison of mutant and wild type animals shows that very few genes are overexpressed in the absence of the miRNA (gene list in the table below and Figure 4.2A). *Nhr-91*, which was previously identified as a target important to maintain blast cell quiescence (Kasuga et al., 2013), was not among those, but its repression might not be relevant in L3 or its change might be lower than the variability that we detect in our experiment.

Since miR-235 is highly expressed in the seam cells, we wondered if the upregulated genes that we identified are expressed in the same tissue. Figure 4.2 shows seam cells expression in L3 and L4 stage, highlighting the upregulated genes upon miR-235 deletion. Notably, they are specifically

expressed only in the seam cells at L4 stage. We confirm this result in Figure 4.2C where we displayed those genes in the context of a full developmental time course in wild type.

We hypothesize that miR-235 loss allows in the hypodermis of younger L3 worms precocious expression of some genes that are usually only expressed at the L4 stage. To rule out the possibility of slight differences in timing between the wild type and the mutant samples, we quantified the expression of previously described germ line genes, which increase gradually their expression with time (Hendriks et al., 2014). If our miR-235 worm sample were developmentally older than the wild type sample, then we would expect higher expression of the germ line genes in the putatively older mutant. In contrast to this possibility, Figure 4. D shows that miR-235 sample is slightly younger than the wild type because the germ line genes are slightly reduced in expression. This supports our hypothesis that miR-235 causes precocious seam cell development. Nevertheless, we cannot exclude that miR-235 has an impact on the germline; possibly leading to a smaller tissue that would also result in reduced expression of germ line genes. Of note, we observed gonad migration defects in miR-235 mutant animals and reduced brood size count (90 in miR-235(-) vs. 182 in wild type).

Next, we scanned the upregulated genes to identify a seed match to miR-235 using TargetScan (Agarwal et al., 2015), but only 2/35 genes had a canonical seed match (*cut-1* and *grl-23*). In addition, none of those genes was found ligated to miR-235 in Ago-chimeras (Broughton et al., 2016). Given the absence of strong evidences to identify miR-235 targets, we concluded that the genes that upregulated in the RNA-seq are likely misregulated as an indirect effect of the miRNA loss.

Discussion

The function of miR-235 in the arrest of blast cell quiescence in response to nutritional stimuli is intriguing. Considering the effect that the nutritional state has on animal development, it could be helpful to understand how nutritional cues and miRNAs affect development using the worms as a model system. It has recently been shown that ethanol and amino acids provided to arrested L1 larvae grown in poor nutritional media awaken blast cells from their quiescent state by activating the insulin like signal in the hypodermis (Fukuyama et al., 2015). We observed a similar effect when larvae were grown in media supplemented with cholesterol. Because we cannot rule out whether the effect is purely cholesterol dependent, or rather ethanol dependent, as cholesterol is dissolved in ethanol, it would be necessary to repeat our experiments in the presence of a control set of worms grown in media supplemented only with the alcohol.

Additionally, it would be worth to characterize further miR-235 by investigating how its expression changes in poor nutritional conditions by growing the P*miR-235*::GFP strain in the absence or presence of ethanol. Maybe the shorter transgene, which showed no hypodermal expression (Figure 4B), will be expressed under nutritional stress. Furthermore, RNA-seq of miR-235 mutant worms grown in the different poorly nutritional conditions might help understanding the mechanism underlying the supposedly altered osmolarity sensitivity of the mutants and the changes involved in the L1/L2 transition induced by ethanol/cholesterol. We hypothesize that miR-235 might be required in stressful conditions to control the cuticle, thus experiments to test if mutants normally enter or exit the alternative *dauer* developmental stage, where yet another cuticle is needed, or if they have alterations of the molting cycle (e.g. by luciferase assay (Olmedo et al., 2015)) would be required.

Considering that we show that miR-235 is specifically expressed in the hypodermis and that it has already been shown to be important to keep the blast cells quiescent, additional studies focused on the relationship between miR-235, food and the insulin-like signaling and their interplay in the hypodermis will be important to understand how the nutritional state regulates developments.

Lastly, our results suggest that miR-235 activity might be needed beyond L1 stage. We suppose that in L3 larvae miR-235 plays a key role in repressing L4 and adult genes. In fact, some of the upregulated genes in *miR-235(-)* are cuticle genes that are L4 or adult specific collagens, hinting to the possibility that the developmental timing of *miR-235(-)* worms might be altered particularly in the cuticle formation or maybe more generally in the hypodermal tissue. *Rol-1* is among the upregulated genes, this collagen is typical of adult cuticles. The *rol-1(e61)* mutant in fact was used to determine if genes have a precocious or retarded phenotype (Abrahante et al., 2003). Accordingly, if *rol-1* is indeed precociously upregulated, animals double mutant for *miR-235 (-); rol-1(e61)* would roll as young larvae, due to the premature presence of adult cuticle. Even if *rol-1* has no seed match to miR-235, the precocious accumulation of this collagen might be a downstream effect of the loss of this miRNA. It is intriguing to think that miR-235 might be part of the heterochronic pathway. To investigate this, the rol-1 assay should be performed together with analysis of typical heterochronic phenotypes of the hypodermis (seam cell fusion, seam cell number and alae secretion). Further experiments will clarify the effects that lack of miR-235 has on worm physiology and will reveal if miR-235 will join the group of miRNAs whose mutation causes heterochronic defects in *C. elegans*.

It is worth mentioning that the available miR-235 mutant strain, *miR-235 (-)*, harbors a deletion that removes the miRNA stem loop, but affects the downstream gene by removing part of its promoter, too. Even though it seems that the phenotypes are specific to the miRNA loss, as shown by Kasuga and colleagues, we generated via CRISPR an additional mutant strain that specifically lacks the miRNA function without affecting the nearby genomic context, by scrambling the seed sequence of the miRNA (HW2643). RNA-seq of this strain will reveal if the upregulated genes in the *miR-235 (-)* strain are specific to the loss of such miRNA.

Additional Material and Methods

***MiR-235* (-) starvation experiment**
Gravid adults of wild type or *miR-235(-)* were bleached as described previously and larvae hatched and starved in one of the following media (with or without addition of 10ul [5mg/ml] cholesterol per 10ml of medium) for 3 or 4 days on a rotating wheel at room temperature. *S-basal* (1L): 5.9g NaCl; 50ml 1M KPO4 (pH 6.0); 1ml cholesterol (5mg/ml in ethanol); water up to 1L, autoclaved. *S-medium complete* (100ml) : S-Basal + 300ul 1M MgSO4; 300ul 1M CaCl2; 1ml 100X trace metal solution; 1ml 1M potassium citrate (pH 6). *M9* (42 mM Na2HPO4, 22 mM KH2PO4, 86 mM NaCl, 1 mM MgSO4).

RNA sequencing and data analysis of *miR-235*(-) mutants Synchronized L1s of wild type N2 and *miR-235*(-) (~1000/plate) were grown for 24h, 26h, 28h and 30h at 25 °C and then independently harvested in TRIzol Reagent. RNA was isolated following manufacturer's instruction. RNA quality was assessed with an Agilent bioanalyzer prior to library preparation. The RNA from each of the four time points was combined in equal amounts for library preparation with Truseq stranded mRNA sample preparation kit (Illumina) and then sequenced on an Illumina HiSeq 2500 with the help of Stephane Thiery, FMI Functional Genomics). The mRNA sequencing data was analyzed as previously described (Hendriks et al., 2014). The data used in Figure 4.2C and D are published in (Hendriks, 2015) and (Hendriks et al., 2014), respectively.

Genes upregulated in *miR-235*(-) mutants

Gene name	WormBase ID	Description
cnc-4	WBGene00000558	cnc-4 encodes one of eleven *C. elegans* caenacin peptides; cnc-4 expression is strongly induced after infection by the fungus Drechmeria coniospora.
col-2	WBGene00000593	col-2 encodes a member of the collagen superfamily containing collagen triple helix repeats (20 copies); expression peaks during the molt that separates the L2 larval and dauer stages as the dauer cuticle is being formed, and mRNA is expressed at low levels in post-dauer L4 larvae and in adult animals.
col-36	WBGene00000613	col-36 encodes a collagen protein; expressed during the L1 to L2 and L2 to dauer larval stage molts.
col-37	WBGene00000614	col-37 encodes a cuticle collagen.
col-40	WBGene00000617	col-40 encodes a cuticle collagen protein that belongs to the col-6 cuticle collagen family; col-40 transcript is present in L1 larvae and at the L2d-dauer molt.
col-51	WBGene00000628	col-51 encodes a cuticle collagen.
col-60	WBGene00000636	developmental delay postembryonic

col-63	WBGene00000639	
col-88	WBGene00000663	mutant animals exhibit increased lethality as a result of exposure to environments with a higher solute concentration than their own internal environment, compared to control.
col-120	WBGene00000694	
col-122	WBGene00000696	
col-140	WBGene00000713	
col-181	WBGene00000754	col-181 is homologous to the human gene PRO ALPHA 1(I) COLLAGEN (COL1A1; OMIM:120150), which when mutated leads to osteogenesis imperfecta.
col-183	WBGene00000756	
cut-1	WBGene00000851	cut-1 encodes a component of cuticlin, an insoluble residue of the nematode cuticle required for alae formation and radial shrinking during dauer differentiation; expressed specifically in the cuticle of dauer larvae and is secreted by the seam cells.
dao-4	WBGene00000930	dao-4 encodes a novel protein, conserved amongst nematodes; dao-4 transcripts are expressed at higher levels in wild-type adult animals than in daf-2 mutant adults at 25C, suggesting that dao-4 expression is positively regulated by DAF-2/insulin-like receptor signaling; reduced dao-4 expression in daf-2 mutants is dependent upon DAF-16.
dpy-14	WBGene00001075	dpy-14 encodes a type III (alpha 1) collagen that is required for embryonic, larval, and vulval development, proper amphid morphology, and regulation of body shape and size; a dpy-14 promoter-GFP fusion construct is reportedly expressed in embryonic neurons.
grd-11	WBGene00001700	grd-11 encodes a hedgehog-like protein, with (from N- to C-terminus) a signal sequence, four Ground domains, an short region of low-complexity sequence, and a Hint/Hog domain; the Hint/Hog domain is predicted to cut GRD-2 into two halves and then covalently link cholesterol to the C-terminus of the Ground domain; the four N-terminal Ground domains are predicted to form one or more cysteine-cross-linked proteins involved in intercellular signaling; Ground domains have subtle similarity to the N-terminal Hedge domain of HEDGEHOG proteins; grd-11 has no obvious function in RNAi assays.
grl-23	WBGene00001732	grl-23 encodes a hedgehog-like protein, with (from N- to C-terminus) a signal sequence, a glycine-rich low-complexity region, a Ground-like (Grl) domain, and an acid-rich low-complexity region; the Grl domain is predicted to form a cysteine-cross-linked protein involved in intercellular signalling, and it has subtle similarity to the N-terminal Hedge domain of HEDGEHOG proteins.
grl-25	WBGene00001734	grl-25 encodes a putatively hedgehog-like protein, with an N-terminal signal sequence and an extended glycine-rich low-complexity region; GRL-25 is expressed in intestine and the nervous system.
rol-1	WBGene00004394	rol-1 encodes a nematode cuticular collagen required for normal body morphology at the adult stage of development; rol-1 expression is under the control of the heterochronic pathway, as heterochronic mutants that synthesize adult cuticle early produce animals that roll as larvae, while heterochronic mutants that fail to execute normal adult development never display the roller phenotype.

b) Transcription factor motifs enriched in comparative analysis (paragraph 2.3.1)

Tables listing the transcription factor motifs identified in selected comparisons. Minimum p value of 10^4, except for seam L4 vs. seam L3 where the first two hits had a p value of 10^3. For the "seam vs vulva" and "vulva vs. seam" comparisons, the genes identified in the "L4 seam cell" sample were used. Strikethrough = genes are absent from the tissue of interest according to the GExplore database (Hutter and Suh, 2016).

seam cells vs vulva

Motif Name	**TF name**	**Consensus**	**P-value**
M4691_1.02	elt-1	AGATRDTATC	1.00E-10
M4713_1.02	elt-1	GATAANVATCT	1.00E-07
~~M5221_1.02~~	~~unc-55~~	~~TGACCTT~~	~~1.00E-05~~
M6454_1.02	nhr-23	TGACCTASTTWT	1.00E-04
M1491_1.02	nhr-41	DTGACCTN	1.00E-04
M1458_1.02	nhr-23	NNTGACCTCN	1.00E-04
M1469_1.02	nhr-23	NTGACCTS	1.00E-04
M1497_1.02	nhr-41	NNTGACCY	1.00E-04
M1459_1.02	nhr-23	RNTGACCTCD	1.00E-04
M3837_1.02	nhr-23	WTGACCTTGATWY	1.00E-04

vulva vs seam cells

Motif Name	**TF name**	**Consensus**	**P-value**
M4343_1.02	daf-19	GGTTGCCATGGCAA	1.00E-07
M4504_1.02	egrh-1	GGBGCGGGGGCGG	1.00E-07
~~M3295_1.02~~	~~fkh-6~~	~~NSNVAGTAAATAAA~~	~~1.00E-06~~
~~M5182_1.02~~	~~ces-1~~	~~CAACAGGTGGT~~	~~1.00E-06~~
M3830_1.02	daf-19	NDGTHRCCATGGCA	1.00E-05
M1971_1.02	egrh-1	GGCGGGGGCGGGG	1.00E-05
M4897_1.02	C27D6.4	GVTBACGTGGCA	1.00E-05
M2943_1.02	hlh-11	AGARYCAGCTGYGG	1.00E-05
M6180_1.02	crh-1	GTGACGTCA	1.00E-04
~~M5270_1.02~~	~~ceh-22~~	~~GCACTTGAGC~~	~~1.00E-04~~

L3seam vs L4seam			
Motif Name	**TF name**	**Consensus**	**P-value**
M5860_1.02	ets-4	ACCCGGATGTR	1.00E-03
M2337_1.02	gei-11	GTGTCGGCCGCT	1.00E-03
M4691_1.02		AGATRDTATC	1.00E-02
M5184_1.02		TKCAACAGGTGGT	1.00E-02
M4577_1.02		YATKNCCAWATAAG	1.00E-02
M1511_1.02		WWGGCCCCCC	1.00E-02
M3960_1.02		GBCCAWATAWGGM	1.00E-02
M4512_1.02		HTTKCCTTATATGG	1.00E-02
M6208_1.02		MMCMGGAAGTSC	1.00E-02
M5163_1.02		STTAATCC	1.00E-02

c) Additional miRNA reporters

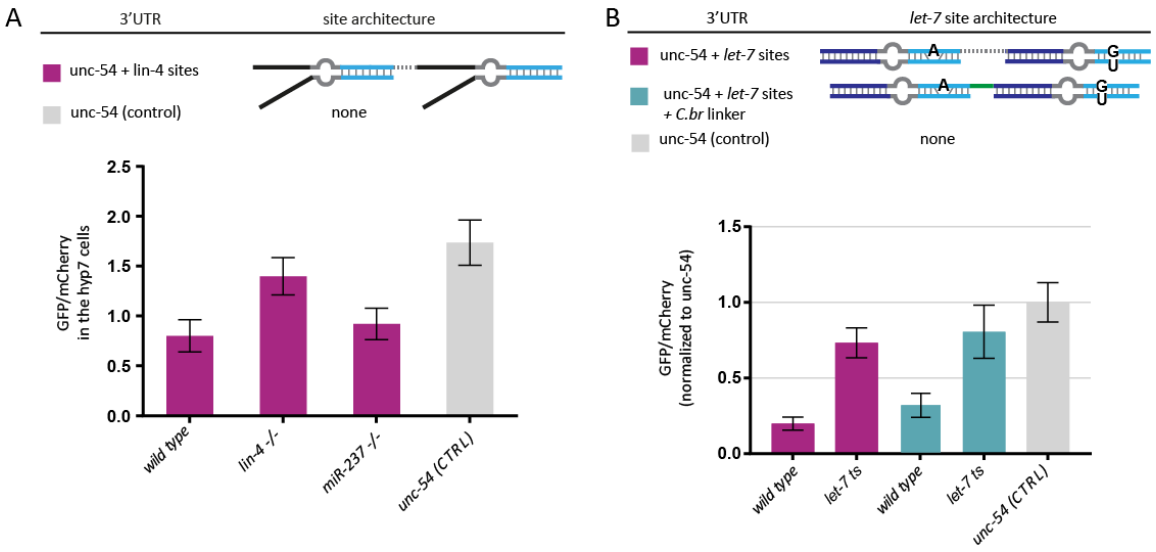


Figure 4.3 (A) Quantification of reporters containing *lin-4* binding sites from the *lin-28* 3'UTR in the hyp7 cells of wild type, *lin-4(-)* and *miR-237(-)* animals at the L4 stage. The sites are specifically silenced by *lin-4*.
(B) Quantification of reporters containing the wild type *let-7* sites spaced by the 12nt linker from *C. briggsae lin-41* 3'UTR sequence (wild type linker is 27nt). De-silencing of the reporter is *let-7* dependent in the *unc-54_let-7 sites +C.br linker*, too, and not affected by the change in the linker.

References

Abbott, A.L., Alvarez-Saavedra, E., Miska, E.A., Lau, N.C., Bartel, D.P., Horvitz, H.R., and Ambros, V. (2005). The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. Dev Cell *9*, 403-414.

Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A., and Rougvie, A.E. (2003). The Caenorhabditis elegans hunchback-like Gene lin-57/hbl-1 Controls Developmental Time and Is Regulated by MicroRNAs. Developmental cell *4*, 625-637.

Aeschimann, F., Kumari, P., Bartake, H., Gaidatzis, D., Xu, L., Ciosk, R., and Grosshans, H. (2017). LIN41 Post-transcriptionally Silences mRNAs by Two Distinct and Position-Dependent Mechanisms. Mol Cell *65*, 476-489 e474.

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*.

Alvarez-Saavedra, E., and Horvitz, H.R. (2010). Many families of C. elegans microRNAs are not essential for development or viability. Curr Biol *20*, 367-373.

Ambros, V. (2011). MicroRNAs and developmental timing. Curr Opin Genet Dev *21*, 511-517.

Ambros, V., and Horvitz, H. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. Science *226*, 409-416.

Andachi, Y., and Kohara, Y. (2016). A whole-mount in situ hybridization method for microRNA detection in Caenorhabditis elegans. RNA *22*, 1099-1106.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Balzeau, J., Menezes, M.R., Cao, S., and Hagan, J.P. (2017). The LIN28/let-7 Pathway in Cancer. Front Genet *8*, 31.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215-233.

Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet *5*, 396-400.

Beadle, G.W., and Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A *27*, 499-506.

Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E., and Meister, G. (2007). Identification of human microRNA targets from isolated argonaute protein complexes. RNA Biol *4*, 76-84.

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P.*, et al.* (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. Nat Struct Mol Biol *15*, 998.

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. Nat Rev Genet *12*, 846-860.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol *11*, R90.

Bethke, A., Fielenbach, N., Wang, Z., Mangelsdorf, D.J., and Antebi, A. (2009). Nuclear Hormone Receptor Regulation of MicroRNAs Controls Developmental Progression. Science *324*, 95-98.

Bielefeld, P., Schouten, M., Lucassen, P.J., and Fitzsimons, C.P. (2017). Transcription factor oscillations in neural stem cells: Implications for accurate control of gene expression. Neurogenesis (Austin) *4*, e1262934.

Blazie, S.M., Babb, C., Wilky, H., Rawls, A., Park, J.G., and Mangone, M. (2015). Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. BMC Biol *13*, 4.

Blazie, S.M., Geissel, H.C., Wilky, H., Joshi, R., Newbern, J., and Mangone, M. (2017). Alternative Polyadenylation Directs Tissue-Specific miRNA Targeting in Caenorhabditis elegans Somatic Tissues. Genetics *206*, 757-774.

Bosson, A.D., Zamudio, J.R., and Sharp, P.A. (2014). Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Mol Cell *56*, 347-359.

Bracht, J., Hunter, S., Eachus, R., Weeks, P., and Pasquinelli, A.E. (2004). Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. RNA *10*, 1586-1594.

Brancati, G., Carl, S.H., and Grosshans, H. (2017). Interplay of target site architecture and miRNA abundance determine miRNA activity and specificity. bioRxiv, 214817. doi: https://doi.org/10.1101/214817

Brancati, G., and Grosshans, H. (2018). An interplay of miRNA abundance and target site architecture determines miRNA activity and specificity. Nucleic Acids Res *46*, 3259-3269.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (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, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol *3*, e85.

Broughton, J.P., Lovci, M.T., Huang, J.L., Yeo, G.W., and Pasquinelli, A.E. (2016). Pairing beyond the Seed Supports MicroRNA Targeting Specificity. Mol Cell *64*, 320-333.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J.*, et al.* (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science *357*, 661-667.

Cech, Thomas R., and Steitz, Joan A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. Cell *157*, 77-94.

Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, 59-69.

Chandradoss, Stanley D., Schirle, Nicole T., Szczepaniak, M., MacRae, Ian J., and Joo, C. (2015). A Dynamic Search Process Underlies MicroRNA Targeting. Cell *162*, 96-107.

Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. Nat Struct Mol Biol *19*, 321-327.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Chin, L.J., Ratner, E., Leng, S., Zhai, R., Nallur, S., Babar, I., Muller, R.U., Straka, E., Su, L., Burki, E.A.*, et al.* (2008). A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. Cancer Res *68*, 8535-8540.

Cimadamore, F., Amador-Arjona, A., Chen, C., Huang, C.T., and Terskikh, A.V. (2013). SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. Proc Natl Acad Sci U S A *110*, E3017-3026.

Cohen, M.L., Kim, S., Morita, K., Kim, S.H., and Han, M. (2015). The GATA factor elt-1 regulates C. elegans developmental timing by promoting expression of the let-7 family microRNAs. PLoS Genet *11*, e1005099.

Contrino, S., Smith, R.N., Butano, D., Carr, A., Hu, F., Lyne, R., Rutherford, K., Kalderimis, A., Sullivan, J., Carbon, S.*, et al.* (2012). modMine: flexible access to modENCODE data. Nucleic Acids Res *40*, D1082-1088.

Crick, F. (1970). Central Dogma of Molecular Biology. Nature *227*.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol *15*, 743-749.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. Mol Cell *54*, 766-776.

Didiano, D., and Hobert, O. (2006). Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nat Struct Mol Biol *13*, 849-851.

Didiano, D., and Hobert, O. (2008). Molecular architecture of a miRNA-regulated 3' UTR. RNA *14*, 1297-1317.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes & development *18*, 504-511.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Drexel, T., Mahofsky, K., Latham, R., Zimmer, M., and Cochella, L. (2016). Neuron type-specific miRNA represses two broadly expressed genes to modulate an avoidance behavior in C. elegans. Genes Dev *30*, 2042-2047.

Easow, G., Teleman, A.A., and Cohen, S.M. (2007). Isolation of microRNA targets by miRNP immunopurification. Rna *13*, 1198-1204.

Ebert, Margaret S., and Sharp, Phillip A. (2012). Roles for MicroRNAs in Conferring Robustness to Biological Processes. Cell *149*, 515-524.

Ecsedi, Matyas. Target specificity and developmental functions of the let-7 microRNA. 2015, Doctoral Thesis, University of Basel, Faculty of Science.

Ecsedi, M., Rausch, M., and Grosshans, H. (2015). The let-7 microRNA directs vulval development through a single target. Dev Cell *32*, 335-344.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell *56*, 104-115.

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, Astrid D., Greene, Emily M., Hannon, Gregory J., and Joshua-Tor, L. (2012). The Structure of Human Argonaute-2 in Complex with miR-20a. Cell *150*, 100-110.

Enright, A.J. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.

Esquela-Kerscher, A., Johnson, S.M., Bai, L., Saito, K., Partridge, J., Reinert, K.L., and Slack, F.J. (2005). Post-embryonic expression of C. elegans microRNAs belonging to the lin-4 and let-7 families in the hypodermis and the reproductive system. Dev Dyn *234*, 868-877.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet *12*, 861-874.

Evans, T.C., ed. (2006). Transformation and microinjection. WormBook, ed The C elegans Research Community, WormBook.

Fang, Z., and Rajewsky, N. (2011). The impact of miRNA target sites in coding sequences and in 3'UTRs. PLoS One *6*, e18067.

Faridani, O.R., Abdullayev, I., Hagemann-Jensen, M., Schell, J.P., Lanner, F., and Sandberg, R. (2016). Single-cell sequencing of the small-RNA transcriptome. Nat Biotech *34*, 1264-1266.

Fay, D.S., Stanley, H.M., Han, M., and Wood, W.B. (1999). A Caenorhabditis elegans homologue of hunchback is required for late stages of development but not early embryonic patterning. Dev Biol *205*, 240-253.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Flamand, M.N., Gan, H.H., Mayya, V.K., Gunsalus, K.C., and Duchaine, T.F. (2017). A non-canonical site reveals the cooperative mechanisms of microRNA-mediated silencing. Nucleic Acids Res *45*, 7212-7225.

Fox, R.M., Von Stetina, S.E., Barlow, S.J., Shaffer, C., Olszewski, K.L., Moore, J.H., Dupuy, D., Vidal, M., and Miller, D.M. (2005). A gene expression fingerprint of C. elegans embryonic motor neurons. BMC Genomics *6*, 42.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome research *19*, 92-105.

Frokjaer-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet *40*, 1375-1383.

Frokjaer-Jensen, C., Davis, M.W., Ailion, M., and Jorgensen, E.M. (2012). Improved Mos1-mediated transgenesis in C. elegans. Nat Meth *9*, 117-118.

Frokjaer-Jensen, C., Davis, M.W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., Pozniakovsky, A., Moerman, D.G., and Jorgensen, E.M. (2014). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat Meth *11*, 529-534.

Fukuyama, M., Kontani, K., Katada, T., and Rougvie, Ann E. (2015). The C. elegans Hypodermis Couples Progenitor Cell Quiescence to the Dietary State. Current Biology *25*, 1241-1248.

Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M.B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat Biotech *33*, 722-729.

Gan, H.H., and Gunsalus, K.C. (2015). Assembly and analysis of eukaryotic Argonaute-RNA complexes in microRNA-target recognition. Nucleic Acids Res *43*, 9613-9625.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol *18*, 1139-1146.

Giachino, C., and Taylor, V. (2014). Notching up neural stem cell homogeneity in homeostasis and disease. Front Neurosci *8*, 32.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth *6*, 343-345.

Gilbert, W. (1985). Genes-in-Pieces Revisited. Science *228* 823-824.

Gorrepati, L., Thompson, K.W., and Eisenmann, D.M. (2013). C. elegans GATA factors EGL-18 and ELT-6 function downstream of Wnt signaling to maintain the progenitor fate during larval asymmetric divisions of the seam cells. Development *140*, 2093-2102.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (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.

Grosshans, H., Johnson, T., Reinert, K.L., Gerstein, M., and Slack, F.J. (2005). The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans. Dev Cell *8*, 321-330.

Grosswendt, S., Filipchyk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol Cell *54*, 1042-1054.

Grün, D., and van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing Experiments. Cell *163*, 799-810.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Ha, I., Wightman, B., and 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, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol *15*, 509-524.

Haenni, S., Ji, Z., Hoque, M., Rust, N., Sharpe, H., Eberhard, R., Browne, C., Hengartner, M.O., Mellor, J., Tian, B*., et al.* (2012). Analysis of C. elegans intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3'-end-seq. Nucleic Acids Res *40*, 6304-6318.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M*., et al.* (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*, 129-141.

Hammell, C.M., Karp, X., and Ambros, V. (2009). A feedback circuit involving let-7-family miRNAs and DAF-12 integrates environmental signals and developmental timing in Caenorhabditis elegans. Proc Natl Acad Sci U S A *106*, 18668-18673.

Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C.S., Han, M., Ding, Y., and Ambros, V. (2008). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. Nat Methods *5*, 813-819.

Hausser, J., Syed, A.P., Bilen, B., and Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. Genome research *23*, 604-615.

Hayes, G.D., Frand, A.R., and Ruvkun, G. (2006). The mir-84 and let-7 paralogous microRNA genes of Caenorhabditis elegans direct the cessation of molting via the conserved nuclear hormone receptors NHR-23 and NHR-25. Development *133*, 4631-4641.

Hayes, G.D., and Ruvkun, G. (2006). Misexpression of the Caenorhabditis elegans miRNA let-7 is sufficient to drive developmental programs. Cold Spring Harb Symp Quant Biol *71*, 21-27.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell *153*, 654-665.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Hendriks, G.J. (2015). Extensive Oscillatory Gene Expression During C. elegans Larval Development. PhD Thesis, University of Basel.

Hendriks, G.J., Gaidatzis, D., Aeschimann, F., and Grosshans, H. (2014). Extensive oscillatory gene expression during C. elegans larval development. Mol Cell *53*, 380-392.

Heo, I., Joo, C., Kim, Y.-K., Ha, M., Yoon, M.-J., Cho, J., Yeom, K.-H., Han, J., and Kim, V.N. (2009). TUT4 in Concert with Lin28 Suppresses MicroRNA Biogenesis through Pre-MicroRNA Uridylation. Cell *138*, 696-708.

Hipfner, D.R., Weigmann, K., and Cohen, S.M. (2002). The bantam gene regulates Drosophila growth. Genetics *161*, 1527-1537.

Horn, M., Geisen, C., Cermak, L., Becker, B., Nakamura, S., Klein, C., Pagano, M., and Antebi, A. (2014). DRE-1/FBXO11-dependent degradation of BLMP-1/BLIMP-1 governs C. elegans developmental timing and maturation. Dev Cell *28*, 697-710.

Hutter, H., and Suh, J. (2016). GExplore 1.4: An expanded web interface for queries on Caenorhabditis elegans protein and gene function. Worm *5*, e1234659.

Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory Control of Factors Determining Multipotency and Fate in Mouse Neural Progenitors. Science *342*, 1203-1208.

Ingolia, N.T. (2016). Ribosome Footprint Profiling of Translation throughout the Genome. Cell *165*, 22-33.

Inoue, T., Sherwood, D.R., Aspock, G., Butler, J.A., Gupta, B.P., and Kirouac, M. (2002). Gene expression markers for Caenorhabditis elegans vulval cells. Mech Dev *119*.

Janssen, H.L., Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patick, A.K., Chen, A., Zhou, Y.*, et al.* (2013). Treatment of HCV infection by targeting microRNA. N Engl J Med *368*, 1685-1694.

Jo, Myung H., Shin, S., Jung, S.-R., Kim, E., Song, J.-J., and Hohng, S. (2015). Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. Molecular Cell *59*, 117-124.

Johnson, A.D., Fitzsimmons, D., Hagman, J., and Chamberlin, H.M. (2001). EGL-38 Pax regulates the ovo-related gene lin-48 during Caenorhabditis elegans organ development. Development *128*, 2857-2865.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., and Slack, F.J. (2005). RAS Is Regulated by the let-7 MicroRNA Family. Cell *120*, 635-647.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature *426*, 845-849.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet *16*, 421-433.

Jorgensen, E.M., and Mango, S.E. (2002). The art and design of genetic screens: caenorhabditis elegans. Nat Rev Genet *3*, 356-369.

Kasuga, H., Fukuyama, M., Kitazawa, A., Kontani, K., and Katada, T. (2013). The microRNA miR-235 couples blast-cell quiescence to the nutritional state. Nature *497*, 503-506.

Katic, I., Xu, L., and Ciosk, R. (2015). CRISPR/Cas9 Genome Editing in Caenorhabditis elegans: Evaluation of Templates for Homology-Mediated Repair and Knock-Ins by Homology-Independent DNA Repair. G3 (Bethesda) *5*, 1649-1656.

Khan, A.A., Betel, D., Miller, M.L., Sander, C., Leslie, C.S., and Marks, D.S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. Nat Biotech *27*, 549-555.

Khorshid, M., Hausser, J., Zavolan, M., and van Nimwegen, E. (2013). A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. Nat Methods *10*, 253-255.

Kim, D., Grun, D., and van Oudenaarden, A. (2013). Dampening of expression oscillations by synchronous regulation of a microRNA and its target. Nat Genet *45*, 1337-1344.

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., and Davidson, G.S. (2001). A Gene Expression Map for Caenorhabditis elegans. Science *293*, 2087-2092.

Konig, I.R., Schumacher, J., Hoffmann, P., Kleensang, A., Ludwig, K.U., Grimm, T., Neuhoff, N., Preis, M., Roeske, D., Warnke, A.*, et al.* (2011). Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. Am J Med Genet B Neuropsychiatr Genet *156B*, 36-43.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res *42*, D68-73.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M.*, et al.* (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.

Kruger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res *34*, W451-454.

Lai, E.C. (2002). Micro RNAs are complementary to 3[prime] UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet *30*, 363-364.

Lai, E.C. (2004). Predicting and validating microRNA targets. Genome Biol *5*, 115.

Lal, A., Navarro, F., Maher, C.A., Maliszewski, L.E., Yan, N., O'Day, E., Chowdhury, D., Dykxhoorn, D.M., Tsai, P., Hofmann, O.*, et al.* (2009). miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. Mol Cell *35*, 610-625.

Lall, S., Grun, D., Krek, A., Chen, K., Wang, Y.L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., Macmenamin, P.*, et al.* (2006). A genome-wide map of conserved microRNA targets in C. elegans. Curr Biol *16*, 460-471.

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

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell *120*, 15-20.

Lewis, B.P., Shih, I.h., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of Mammalian MicroRNA Targets. Cell *115*, 787-798.

Li, M., Jones-Rhoades, M.W., Lau, N.C., Bartel, D.P., and Rougvie, A.E. (2005). Regulatory mutations of mir-48, a C. elegans let-7 family MicroRNA, cause developmental timing defects. Dev Cell *9*, 415-422.

Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A microRNA imparts robustness against environmental fluctuation during development. Cell *137*, 273-282.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X.*, et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-469.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003a). Vertebrate MicroRNA Genes. Science *299*, 1540-1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003b). The microRNAs of Caenorhabditis elegans. Genes Dev *17*, 991-1008.

Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev Cell *4*, 639-650.

Liu, N., Bezprozvannaya, S., Williams, A.H., Qi, X., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2008). microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. Genes Dev *22*, 3242-3254.

Luna, J.M., Barajas, J.M., Teng, K.Y., Sun, H.L., Moore, M.J., Rice, C.M., Darnell, R.B., and Ghoshal, K. (2017). Argonaute CLIP Defines a Deregulated miR-122-Bound Transcriptome that Correlates with Patient Survival in Human Liver Cancer. Mol Cell *67*, 400-410 e407.

Macosko, Evan Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, Allison R., Kamitaki, N., Martersteck, Emily M*., et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Martinez, N.J., Ow, M.C., Reece-Hoyes, J.S., Barrasa, M.I., Ambros, V.R., and Walhout, A.J. (2008). Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. Genome Res *18*, 2005-2015.

McCulloch, K.A., and Rougvie, A.E. (2014). Caenorhabditis elegans period homolog lin-42 regulates the timing of heterochronic miRNA expression. Proc Natl Acad Sci U S A *111*, 15450-15455.

McJunkin, K., and Ambros, V. (2017). A microRNA family exerts maternal control on sex determination in C. elegans. Genes Dev *31*, 422-437.

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

Merritt, C., Rasoloson, D., Ko, D., and Seydoux, G. (2008). 3' UTRs Are the Primary Regulators of Gene Expression in the C. elegans Germline. Current Biology *18*, 1476-1482.

Miki, T.S., Ruegger, S., Gaidatzis, D., Stadler, M.B., and Grosshans, H. (2014). Engineering of a conditional allele reveals multiple roles of XRN2 in Caenorhabditis elegans development and substrate specificity in microRNA turnover. Nucleic Acids Res *42*, 4056-4067.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692-1694.

Min, H., and Yoon, S. (2010). Got target? Computational methods for microRNA target prediction and their extension. Exp Mol Med *42*.

Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007a). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS Genet *3*, e215.

Mok, D.Z., Sternberg, P.W., and Inoue, T. (2015). Morphologically defined sub-stages of C. elegans vulval development in the fourth larval stage. BMC Dev Biol *15*, 26.

Moore, M.J., Scheel, T.K., Luna, J.M., Park, C.Y., Fak, J.J., Nishiuchi, E., Rice, C.M., and Darnell, R.B. (2015). miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. Nat Commun *6*, 8864.

Moss, E.G., Lee, R.C., and 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.

Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. Nat Genet *43*, 854-859.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368-374.

Newman, M.A., Thomson, J.M., and Hammond, S.M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. Rna *14*, 1539-1549.

Olmedo, M., Geibel, M., Artal-Sanz, M., and Merrow, M. (2015). A High-Throughput Method for the Analysis of Larval Developmental Phenotypes in Caenorhabditis elegans. Genetics *201*, 443-448.

Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M., and Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol Cell *33*, 204-214.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P*., et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86-89.

Pauli, F., Liu, Y., Kim, Y.A., Chen, P.-J., and Kim, S.K. (2006). Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. Development *133*, 287-295.

Pennisi, E. (2012). ENCODE Project Writes Eulogy for Junk DNA. Science *337*, 1159-1161.

Perales, R., King, D.M., Aguirre-Chen, C., and Hammell, C.M. (2014). LIN-42, the Caenorhabditis elegans PERIOD homolog, negatively regulates microRNA transcription. PLoS Genet *10*, e1004486.

Pfeuty, B. (2015). A computational model for the coordination of neural progenitor self-renewal and differentiation through Hes1 dynamics. Development *142*, 477-485.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat Protocols *9*, 171-181.

Ratner, E., Lu, L., Boeke, M., Barnett, R., Nallur, S., Chin, L.J., Pelletier, C., Blitzblau, R., Tassi, R., Paranjape, T*., et al.* (2010). A KRAS-variant in ovarian cancer acts as a genetic marker of cancer risk. Cancer Res *70*, 6509-6515.

Rausch, M., Ecsedi, M., Bartake, H., Mullner, A., and Grosshans, H. (2015). A genetic interactome of the let-7 microRNA in C. elegans. Dev Biol *401*, 276-286.

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA-target duplexes. Rna *10*, 1507-1517.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

Ren, Z., and Ambros, V.R. (2015). Caenorhabditis elegans microRNAs of the let-7 family act in innate immune response circuits and confer robust developmental timing against pathogen stress. Proc Natl Acad Sci U S A *112*, E2366-2375.

Resnick, T.D., McCulloch, K.A., and Rougvie, A.E. (2010). miRNAs give worms the time of their lives: small RNAs and temporal control in Caenorhabditis elegans. Dev Dyn *239*, 1477-1489.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M.*, et al.* (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res *43*, D670-681.

Rougvie, A.E. (2001). Control of developmental timing in animals. Nat Rev Genet *2*, 690-701.

Rougvie, A.E., and Ambros, V. (1995). The heterochronic gene lin-29 encodes a zinc finger protein that controls a terminal differentiation event in Caenorhabditis elegans. Development *121*, 2491-2500.

Roush, S.F., and Slack, F.J. (2009). Transcription of the C. elegans let-7 microRNA is temporally regulated by one of its targets, hbl-1. Dev Biol *334*, 523-534.

Roush, S., and Slack, F.J. (2008). The let-7 family of microRNAs. Trends Cell Biol *18*, 505-516.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature *418*.

Rupaimoole, R., and Slack, F.J. (2017). MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov *16*, 203-222.

Ruvkun, G., Wightman, B., and Ha, I. (2004). The 20 years it took to recognize the importance of tiny RNAs. Cell *116*, S93-96, 92 p following S96.

Ruvkun, G. (2008). The perfect storm of tiny RNAs. Nat Med *14*, 1041-1045.

Salditt-Georgieff, M., and Darnell, J.E., Jr. (1982). Further evidence that the majority of primary nuclear RNA transcripts in mammalian cells do not contribute to mRNA. Mol Cell Biol *2*, 701-707.

Salomon, William E., Jolly, Samson M., Moore, Melissa J., Zamore, Phillip D., and Serebrov, V. (2015). Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. Cell *162*, 84-95.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat Meth *9*, 676-682.

Schindler, A.J., and Sherwood, D.R. (2013). Morphogenesis of the C. elegans vulva. Wiley interdisciplinary reviews Developmental biology *2*, 75-95.

Schirle, N.T., and MacRae, I.J. (2012). The Crystal Structure of Human Argonaute2. Science *336*, 1037-1040.

Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. Science *346*, 608-613.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.

Shaw, W.R., Armisen, J., Lehrbach, N.J., and Miska, E.A. (2010). The conserved miR-51 microRNA family is redundantly required for embryonic development and pharynx attachment in Caenorhabditis elegans. Genetics *185*, 897-905.

Sherrard, R., Luehr, S., Holzkamp, H., McJunkin, K., Memar, N., and Conradt, B. (2017). miRNAs cooperate in apoptosis regulation during C. elegans development. Genes Dev *31*, 209-222.

Shimojo, H., Ohtsuka, T., and Kageyama, R. (2008). Oscillations in notch signaling regulate maintenance of neural progenitors. Neuron *58*, 52-64.

Shin, C., Nam, J.W., Farh, K.K., Chiang, H.R., Shkumatava, A., and Bartel, D.P. (2010). Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell *38*, 789-802.

Simon, D.J., Madison, J.M., Conery, A.L., Thompson-Peer, K.L., Soskis, M., Ruvkun, G.B., Kaplan, J.M., and Kim, J.K. (2008). The microRNA miR-1 regulates a MEF-2-dependent retrograde signal at neuromuscular junctions. Cell *133*, 903-915.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol *15*, 259-267.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell *5*, 659-669.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem Sci *39*, 381-399.

Spencer, W.C., McWhirter, R., Miller, T., Strasbourger, P., Thompson, O., Hillier, L.W., Waterston, R.H., and Miller, D.M., 3rd (2014). Isolation of specific neurons from C. elegans larvae for gene expression profiling. PLoS One *9*, e112102.

Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J.*, et al.* (2011). A spatial and temporal map of C. elegans gene expression. Genome Res *21*, 325-341.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila MicroRNA targets. PLoS Biol *1*, E60.

Stefani, G., Chen, X., Zhao, H., and Slack, F.J. (2015). A novel mechanism of LIN-28 regulation of let-7 microRNA expression revealed by in vivo HITS-CLIP in C. elegans. Rna *21*, 985-996.

Steiner, F.A., and Henikoff, S. (2015). Cell type-specific affinity purification of nuclei for chromatin profiling in whole animals. Methods Mol Biol *1228*, 3-14.

Steiner, F.A., Talbert, P.B., Kasinathan, S., Deal, R.B., and Henikoff, S. (2012). Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome research *22*, 766-777.

Subasic, D., Brümmer, A., Wu, Y., Pinto, S.M., Imig, J., Keller, M., Jovanovic, M., Lightfoot, H.L., Nasso, S., Goetze, S*., et al.* (2015). Cooperative target mRNA destabilization and translation inhibition by miR-58 microRNA family in C. elegans. Genome research *25*, 1680-1691.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol *56*, 110-156.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol *100*, 64-119.

Takayama, J., Faumont, S., Kunitomo, H., Lockery, S.R., and Iino, Y. (2010). Single-cell transcriptional analysis of taste sensory neuron pair in Caenorhabditis elegans. Nucleic Acids Res *38*.

Thammaiah, C.K., and Jayaram, S. (2016). Role of let-7 family microRNA in breast cancer. Noncoding RNA Res *1*, 77-82.

Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., Vervisch, E., Brohee, S., and van Helden, J. (2008). RSAT: regulatory sequence analysis tools. Nucleic Acids Res *36*, W119-127.

Tintori, S.C., Osborne Nishimura, E., Golden, P., Lieb, J.D., and Goldstein, B. (2016). A Transcriptional Lineage of the Early C. elegans Embryo. Dev Cell *38*, 430-444.

Towbin, Benjamin D., González-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., Askjaer, P., and Gasser, Susan M. (2012). Step-Wise Methylation of Histone H3K9 Positions Heterochromatin at the Nuclear Periphery. Cell *150*, 934-947.

Tsialikas, J., Romens, M.A., Abbott, A., and Moss, E.G. (2017). Stage-Specific Timing of the microRNA Regulation of lin-28 by the Heterochronic Gene lin-14 in Caenorhabditis elegans. Genetics *205*, 251-262.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science *302*, 1212-1215.

Vadla, B., Kemper, K., Alaimo, J., Heine, C., and Moss, E.G. (2012). lin-28 Controls the Succession of Cell Fate Choices via Two Distinct Activities. PLoS genetics *8*, e1002588.

Van Wynsberghe, P.M., Finnegan, E.F., Stark, T., Angelus, E.P., Homan, K.E., Yeo, G.W., and Pasquinelli, A.E. (2014). The Period protein homolog LIN-42 negatively regulates microRNA biogenesis in C. elegans. Developmental biology *390*, 126-135.

Van Wynsberghe, P.M., Kai, Z.S., Massirer, K.B., Burton, V.H., Yeo, G.W., and Pasquinelli, A.E. (2011). LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in Caenorhabditis elegans. Nat Struct Mol Biol *18*, 302-308.

Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., and Slack, F.J. (2004a). The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. Genes Dev *18*, 132-137.

Vella, M.C., Reinert, K., and Slack, F.J. (2004b). Architecture of a validated microRNA::target interaction. Chem Biol *11*, 1619-1623.

Vigneault, F., Ter-Ovanesyan, D., Alon, S., Eminaga, S., D, C.C., Seidman, J.G., Eisenberg, E., and G, M.C. (2012). High-throughput multiplex sequencing of miRNA. Curr Protoc Hum Genet *Chapter 11*, Unit 11 12 11-10.

Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective Blockade of MicroRNA Processing by Lin28. Science *320*, 97-100.

Wang Y, J.S., Li H, Sheng G, Tuschl T, Patel DJ. (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456(7224)*, 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754-761.

Wee, Liang M., Flores-Jasso, C.F., Salomon, William E., and Zamore, Phillip D. (2012). Argonaute Divides Its RNA Guide into Domains with Distinct Functions and RNA-Binding Properties. Cell *151*, 1055-1067.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K*., et al.* (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431-1443.

Wolter, J.M., Le, H.H., Linse, A., Godlove, V.A., Nguyen, T.D., Kotagama, K., Lynch, A., Rawls, A., and Mangone, M. (2017). Evolutionary patterns of metazoan microRNAs reveal targeting principles in the let-7 and miR-10 families. Genome Res *27*, 53-63.

Wright, J.E., Gaidatzis, D., Senften, M., Farley, B.M., Westhof, E., Ryder, S.P., and Ciosk, R. (2011). A quantitative RNA code for mRNA target selection by the germline fate determinant GLD-1. EMBO J *30*, 533-545.

Wu, L., and Belasco, J.G. (2005). Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. Mol Cell Biol *25*, 9198-9208.

Wu, L., Nguyen, L.H., Zhou, K., de Soysa, T.Y., Li, L., Miller, J.B., Tian, J., Locker, J., Zhang, S., Shinoda, G*., et al.* (2015). Precise let-7 expression levels balance organ regeneration against tumor suppression. eLife *4*, e09431.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotech *29*, 607-614.

Zhang, H., Artiles, K.L., and Fire, A.Z. (2015). Functional relevance of “seed” and “non-seed” sequences in microRNA-mediated promotion of C. elegans developmental progression. Rna *21*, 1980-1992.

Zhang, S.K., Jeffrey R. (2013). Cell isolation and culture The C. elegans Research Community. WormBook. 1551-8507.

Zhao, J.L., Rao, D.S., O’Connell, R.M., Garcia-Flores, Y., and Baltimore, D. (2013). MicroRNA-146a acts as a guardian of the quality and longevity of hematopoietic stem cells in mice. eLife *2*, e00537.

Zheng, L., Baumann, U., and Reymond, J.L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res *32*, e115.

Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M., and Mazutis, L. (2017). Single-cell barcoding and sequencing using droplet microfluidics. Nat Protocols *12*, 44-73.

Zisoulis, D.G., Kai, Z.S., Chang, R.K., and Pasquinelli, A.E. (2012). Autoregulation of microRNA biogenesis by let-7 and Argonaute. Nature *486*, 541-544.

Zisoulis, D.G., Lovci, M.T., Wilbert, M.L., Hutt, K.R., Liang, T.Y., Pasquinelli, A.E., and Yeo, G.W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat Struct Mol Biol *17*, 173-179.

Acknowledgements

To all the people that crossed my path

"I thank you for having met me, for my being able to remember you all my life"

Dostoevskij, White Nights.

I would like to start thanking my supervisor ***Helge*** *Grosshans, for the time we spent together looking at worms or miRNA duplexes, for the freedom you gave me and for the times when you listened (but also the ones when you did not!). In the last four years together, we went through a broad range of situations in which you taught me basic scientific skills, from the importance of control experiments to how to write a paper, but also how to stand for my ideas and handle difficult situations. Grazie, Helge!*

I am grateful to the members of my thesis committee, ***Rene*** *Ketting and* ***Mihaela*** *Zavolan, for their support during my PhD and their helpful feedback on both science and career. Thanks for your time; you have been a precious resource. I want to thank* ***Witek Filipowicz*** *for chairing my exam; I feel honored and look forward to an inspiring discussion. I am grateful to* ***Anne*** *Ephrussi because our meeting motivated me to not give up when I needed it the most and to* ***Amy*** *Pasquinelli for the helpful critical discussions during our meetings.*

A big "thank you" goes to the ***Grosshans*** *lab, old and current members. Life in the lab would have not been as much fun without you. Lene, Stefan, Gert, Flo and Yannick thanks for the fun we had and still have together.* ***Lene****, thanks for helping me navigating the lab;* ***Stefan****, thanks for taking care of me as a big brother;* ***Gert****, thanks for our endless discussions;* ***Flo****, thanks for helping in many occasions, especially for pushing me to write the BIF fellowship when I was about to give up, and* ***Yannick****, thanks for the "Alpenrose". I want to thank* ***Sarah****, for her great help with bioinformatics analysis;* ***Manu****, for his latin optimism and help with RNA work;* ***Monika****, for teaching me the lab basics and being our point of reference in the lab;* ***Jun*** *for his constant smile and encouraging and kind words; thanks also to* ***Milou, Thomas, Ben, Takashi, Hannes, Hrishi and Iskra****. A special thanks goes to* ***Kathrin*** *and* ***Lan,*** *I would have never finished all the things I started without them. Thanks to* ***ChiaretTina****, I would have been lost in this last year without you, so thanks for all the time we spent together. Lastly, Chiara, Flo and Jun, thanks for your helpful comments on this thesis.*

I want to thank the ***Ciosk Lab*** *for the time spent together and particularly* ***Pooja*** *Kumari, for her critical and helpful feedback. I am grateful to* ***Dimos*** *Gaidatzis for teaching me the basics of bioinformatics and for the endless days spent together analyzing data. I want to thank* ***Hubertus*** *Kohler for the long afternoons we spent sorting cells and for accepting to start a new challenging project;* ***Birgitte*** *Lucas, whose work has been*

instrumental for the success of my experiments; ***Kirsten, Stephan, Tim & Sebastien*** *of the functional genomics facility for their essential support;* ***FAIM*** *(Laurent, Steve, Raphael, Jan and Moritz) for help with the confocal microscopy and imaging analysis; the* ***IT*** *department for their excellent support throughout the years and the* ***media kitchen*** *for its superb work that spoils us so much. I want to thank* ***Elida*** *for her help navigating the institute and the bureaucracy of Basel-Stadt, I would be lost without you! And thanks to* ***Sandra*** *Ziegler, it has been great working with you in the LIBRA team.*

I want to thank the ***student representatives*** *because with them I learned how to team up and support each other while trying to improve the PhD program. "****Firsties****": Ivana, Christina, Valeee, Johannes, Julia (thanks for saving me from endnote!), David, Romain, Silvi, Domi, Aris, Isabella, we have come a long way together since we met at the PhD selections and now we are almost done. Thanks for the support along the way, dear friends.*

I am grateful to ***Boheringer Ingheleim Fonds****, which funded my PhD project and introduced me to wonderful people.* ***Maria****,* ***Chiara*** *and* ***Mirjam*** *thanks for our retreats around the world. I want to thank the* ***Antelope program*** *at University, which helped me improving my soft skills and put me in contact with a great group of talented women and my mentor* ***Julie****, with whom I have enjoyed several helpful discussions. I also want to thank some of the other people who made my life in Basel so amazing: Tam, Sjoerd&Stef, Ale, Chris&Edo, Vroni, Mike, Nathalie&Melina, Lynn, Yannick, Bolla&Celi, Nic, Francesca&Celeste, Lena, Enrico, Melissa, James, Jan, Monica, Silvia, Conni, Dani, Daniele, Anais, Federico, Kasia, my former flatmates and the FMI crew.*

A special thanks to ***Chrissi*** *&* ***Ivča****, for all the great time we spent together, I am sure distance will not break us apart.* ***Perito****, grazie per essere la mia pillola del buon umore da dieci anni!*

Grazie ad ***Alberto****, per aver condiviso gioie e dolori durante questo dottorato, per avermi aiutato, ascoltato, per essere stato il mio personal consultant e per aver creduto in me fin dal primo momento. Grazie di tutto!*

Grazie al ***nonno Aldo*** *per il suo orgoglio quando ripete che ha "una nipote scienziata". E infine grazie a* ***mamma, papa'*** *e* ***Manu****. Grazie per avermi ascoltata, incoraggiata, rincuorata, supportata e per aver gioito con me in tutti questi anni lontano da casa. E' bellissimo incontrarsi e ritrovarsi insieme come se il tempo non fosse mai passato, che sia a Basel, a Palermo o in Rwanda. Non ce l'avrei mai fatta senza di voi, vi voglio bene!*