

## 4. Heterochronic genes and regulatory RNAs

Genetic and molecular analyses in *Drosophila* and *C. elegans* have shown that gene pathways specific to one axis of the developing animal—for example, anterior/posterior, dorsal/ventral or temporal-- generate control gene activities which form a gradient over that developmental axis to specify patterned arrays of cells. Mutations that disrupt their normally asymmetric pattern of expression or activation lead to homeotic or heterochronic changes.

Such explicit genetic control of temporal patterning has been revealed by analysis of the heterochronic gene pathway in *C. elegans*. During wild type development, blast cells execute stage-specific patterns of cell lineage and cell differentiation at each of the four larval stages and during the adult stage. Mutations in the heterochronic genes *lin-4*, *lin-14*, *lin-28*, *egl-35*, *let-7*, *lin-41*, *daf-12*, *lin-46*, and *lin-29*, cause temporal transformations in these stage-specific patterns of cell lineage in many tissues and cell types. Depending on the mutation, stage-specific developmental events in heterochronic mutant animals occur at either earlier or later stages than they would normally.

An instructive role for *lin-14* in temporal pattern formation was established by the observation that *lin-14* gain-of-function and loss-of-function mutations have the opposite phenotypes. Loss-of-function (lf) *lin-14* alleles cause the precocious execution at early larval stages of cell fates appropriate for later larval stages. Gain-of-function (gf) *lin-14* alleles cause the opposite transformation in temporal cell fate—reiterations of early cell fates at later stages.

*lin-14* controls stage-specific cell lineages by generating a temporal gradient of the LIN-14 proteins. The *lin-14* gene encodes three related nuclear proteins of unknown biochemical function. Antibodies which detect all three proteins show that they accumulate in the nuclei of most cells of wild type late embryos and L1 larvae, but the LIN-14 protein levels fall dramatically during the late L1 stage and remain low at all subsequent stages. The *lin-14(gf)* mutations interfere with normal *lin-14* temporal down-regulation, causing inappropriately high LIN-14 protein levels at post-L1 stages, which imposes L1 patterns of cell lineage on post-L1 blast cells. Because LIN-14 is a nuclear protein we postulate that the proteins regulate the expression of genes that mediate L1-specific and L2-specific patterns of cell lineage.

*lin-4* is a negative regulator of *lin-14*. A null mutation in *lin-4* results in L1-specific cell lineage reiterations, and cause inappropriate LIN-14 protein to be present at post-L1 stages. The functional products of the *lin-4* gene are two small (22 nt and 61 nt) untranslated RNA molecules. The abundance of these RNAs is temporally regulated—they are expressed late in the L1 stage, just before LIN-14 protein levels begin to fall.

The temporal regulation of LIN-14 protein abundance during wild type development occurs at a post-transcriptional step and is mediated by the *lin-14* 3' UTR. First, *lin-14(gf)* mutations delete sequences from the 3'UTR, showing that regulatory elements in the 3' UTR are normally responsible for mediating the down-regulation of LIN-14 protein abundance at later larval stages. Second, the *lin-14* 3' UTR is sufficient to confer this post-transcriptional temporal regulation on an unrelated transcript. The *lin-4* gene product is required for the post-transcriptional temporal regulation of both the *lin-14* gene and chimeric reporter genes bearing the *lin-14* 3' UTR. Both the function and the nucleotide sequence of particular regions of the *lin-14* 3' UTR are conserved in the related species *C. briggsae*. Among the conserved sequences are seven copies of a 14 to 19 nucleotide sequence that is complementary to the *lin-4* RNA and necessary for the temporal gradient generating capacity of the *lin-14* UTR and for binding in vitro to the *lin-4* RNA. These data show that the temporal gradient in LIN-14 protein is triggered by the up regulation of *lin-4* RNA expression which in turn acts to down-regulate LIN-14 translation by binding directly to the *lin-14* 3' UTR.

### The molecular dissection of the *lin-4/lin-14* RNA duplex

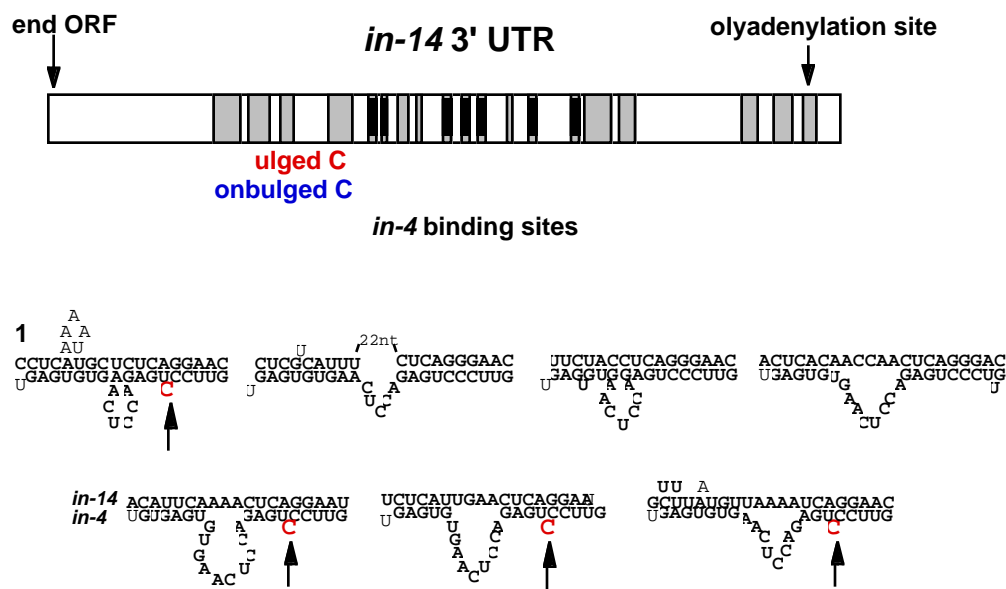
Ilho Ha (Hanwa Chemical Co, Korea) and Bruce Wightman (now on the faculty at Muhlenburg College)

The *lin-14* mRNA 3' UTR mediates the generation of the LIN-14 temporal gradient and is regulated by *lin-4*. Expression of the *lin-4* regulatory RNA is up-regulated during the late L1 stage to down-regulate LIN-14 translation by specific base-pairing of multiple *lin-4* RNAs to the *lin-14* mRNA. The *lin-14* 3' UTR has seven copies of a 14 to 19 nucleotide sequence that is complementary to the *lin-4* RNA. The complementarity between *lin-4* and *lin-14* implies that

*lin-4* is synthesized in the many (and probably all) somatic cells which accumulate LIN-14, but this has not been formally established.

The *lin-4/lin-14* RNA duplex model is supported by mutations in *lin-4* and in *lin-14*, as well as by *in vitro* binding studies which show that the RNAs form a complex that is disrupted by mutations in the complementary region. Chemically synthesized *lin-4* RNA (24 mer) binds to *in vitro* transcribed wild-type *lin-14* 3'UTR RNA. However the same *lin-4* RNA fails to bind to the *lin-14* 3'UTR with mutations in all seven *lin-4* complementary regions (three substitution mutations (UCA to AGU) in the duplex regions of each of the seven elements of *lin-14* 3'UTR). We assayed the temporal regulation of a lacZ reporter gene bearing the same mutant *lin-14* 3' UTR that fails to bind *lin-4* RNA *in vitro*. Animals carrying the mutant *lin-14* 3'UTR show 10 fold less down regulation of reporter gene activity than those with wild type *lin-14* 3'UTR. However both reporter genes are not down regulated in *lin-4* null background. Thus the *lin-4* binding sites are necessary for temporal down regulation.

#### Four of *lin-4* binding sites in *lin-14* 3'UTR bulge a *lin-4* C residue.



The strong *lin-4(ma161)* allele substitutes a U for a C at position 6 of the *lin-4* RNA. Thermodynamic calculations suggest that the substitution of a G::U base pair in a *lin-4(ma161)/lin-14(+)* RNA duplex for the central G::C base pair of a wild type RNA duplex would lower the binding constant of each by a factor of 100. But in addition, while three of the proposed *lin-4/lin-14* RNA duplexes would base pair perfectly in the core region, four of these sites would base pair with the *lin-4* RNAs only by bulging out one of the conserved C residues of the *lin-4* sequence (sites 1, 2, 4, and 6). The *lin-4(ma161)* mutation should also affect the four bulged C RNA duplexes, most likely bulging the U rather than a C.

We showed that only bulged C *lin-4* binding sites are functional in temporal gradient formation *in vivo*. We constructed *lin-14* 3' UTRs that bear either perfect duplex or bulged duplex regions with *lin-4*. The bulged RNA duplexes are down-regulated in a *lin-4* dependent manner *in vivo*, whereas a *lin-14* 3' UTR with non bulged duplexes is not downregulated. Paradoxically, *lin-4* RNA binds *in vitro* to non-bulged *lin-14* RNA more avidly than to the bulged *lin-14* RNA, as predicted from thermodynamic calculations. One model to explain this is that a specific secondary structure of *lin-4/lin-14* RNA duplex may be recognized by an accessory protein rather than an RNA duplex *per se* is required *in vivo* for the generation of the LIN-14 temporal gradient.

The sequence conservation of both the *lin-4* and the multiple complementary regions in the *lin-14* 3' UTR also highlights possible functional residues in addition to those which are base paired in the proposed *lin-4/lin-14* RNA duplexes. For example, all seven proposed RNA duplexes loop out the conserved residues ACCUCA from the *lin-4* RNA. Because this loop is

conserved, we hypothesize that it must remain in a loop to down-regulate translation. For example, the conserved loops or bulges on the duplex could be recognized by accessory proteins or may be essential to any catalytic activity the *lin-4/lin-14* RNA duplex may possess. RNA duplexes are A-form double helices which have equally sized major and minor grooves neither of which is wide enough to allow entry of an alpha helical protein domain, unlike B-form DNA duplexes which have a much larger major groove and correspondingly smaller minor groove. Bulges and loops in RNA structures have been shown to form the recognition elements for RNA binding proteins. Some of the experiments proposed explore the function of the features decorating the *lin-4/lin-14* and *let-7/lin-41* RNA duplexes.

How could the *lin-4/lin-14* RNA duplexes inhibit translation? For instance, the 7 bound *lin-4* RNA molecules could sequester the *lin-14* message or provide sequence-specific binding for a protein that prevents its translation. Such "masking" of maternal messages in sea urchin eggs has been observed. Because the *lin-14* 3' UTR can also down-regulate translation of a hybrid *col-10/lacZ/lin-14* mRNA, the regulatory activity of the *lin-14/lin-4* RNA duplex does not depend on 5' sequences peculiar to the *lin-14* mRNA. Because the level of the *lin-14* mRNA and the *col-10/lacZ/lin-14* 3' UTR mRNA do not change during *lin-4* -mediated post-transcriptional down-regulation, the *lin-4/lin-14* RNA duplex does not destabilize the *lin-14* mRNA.

### **Genetic identification of components necessary for *lin-14* 3' UTR mediated down-regulation**

#### **i. A *lin-4* allele-specific suppressor screen**

**Ilho Ha (Hanwa Chemical Co, Korea) and Amy Pasquinelli**

We sought mutations that suppress the heterochronic phenotype of a *lin-4* point mutant, *lin-4(ma161)*. This mutant *lin-4* RNA produces a product (albeit at lower abundance) but is not functional in assays for downregulation of translation of either the normal *lin-14* 3' UTR or reporter genes bearing this *lin-14* 3' UTR. Because this mutant *lin-4* RNA is expected to bulge a U rather than a C and because we showed that the bulged nucleotide is essential for *lin-4/lin-14* RNA duplex mediated downregulation, we reasoned that mutations in particular proteins that recognize the bulged C might allow, for example, a relaxed specificity mutant protein to recognize the bulged U. We realized that it is a lot to expect a change of specificity mutant at any reasonable frequency, but the genetic selections available were very powerful, allowing rare mutations to be detected. From a screen of over 1,000,000 chromosomes (this was a selection actually), we isolated two independent suppressor mutations, *srb-1,2* (for suppressor of RNA bulge) that can suppress the *lin-4* point mutant but not a *lin-4* deletion mutant. These mutations map to the same genetic region and may in fact be in the same gene (because these mutations appear to be dominant, allelism tests have not yet been possible). We do not yet know if mutations in this gene causes heterochronic phenotypes on its own, but the allele specificity of this suppression is so striking that we have decided to molecularly analyze the gene.

#### **ii. A screen for trans-acting factors necessary for *lin-4/lin-14* 3' UTR regulation**

**Ilho Ha (Hanwa Chemical Co, Korea)**

We developed phenotypic reporter genes for *lin-14* 3' UTR activity, to allow a genetic screen for mutations that fail to allow the *lin-4/lin-14* RNA duplex to function. We fused the *lin-14* 3' UTR to the dominant *rol-6(su1006)* reporter gene. Normally *rol-6* causes animals to roll on plates at all post L1 stages. Fusion of the *lin-14* 3' UTR to the *rol-6* promoter plus coding region causes animals to no longer rolled in a wild type background but animals bearing this fusion gene do roll if they also carry a *lin-4* mutation. We used the non rolling phenotype in wild type to screen for mutations in other genes that are necessary for LIN-4 action. In a genetic screen of more than 10,000 chromosomes, we isolated 10 new suppressor mutations, *slu-1* to *10* (for suppressor of lin-14 UTR), including multiple alleles of 1 gene. We found subtle heterochronic phenotypes in many of these mutants and are now exploring these phenotypes in combination with sensitized heterochronic genetic backgrounds (ie *lin-14(n179)*, a ts mutant at various temperatures). These new genetic loci promise to identify proteins or RNAs that may interact with the *lin-4/lin-14* RNA duplex to in turn mediate down-regulation of mRNA translation in cis. We plan to test whether the new genes identified in this screen also affect

the *let-7* RNA regulation of *lin-41* 3' UTR function. We will molecularly analyse any genes that have heterochronic phenotypes and/or also affect the other heterochronic RNA duplex.

### Genetic discovery of a second regulatory RNA in the heterochronic pathway

Brenda Reinhart, Frank Slack (now the faculty at Yale), and Amy

Pasquinelli, in collaboration with Michael Basson and Bob Horvitz at MIT

We explored other genes in the heterochronic regulatory network by genetic screens for mutations that suppress the heterochronic phenotypes conferred by decreased *lin-14* gene activity. Two mutations caused retarded heterochronic defects, and mapped to the position of the canonical *let-7* lethal allele *let-7(mn112)*.

Based upon patterns of hypodermal cell divisions, the first evidence of a *let-7* heterochronic defect was reiteration of L4 patterns of division at the "adult" stage. The opposite phenotype resulted from increasing the gene dosage of *let-7*: hypodermal cells express adult fates following the L3-to-L4 molt. The opposite heterochronic phenotypes caused by reducing or increasing *let-7* activity suggest that *let-7* functions as a temporal switch between late larval and adult fates. The retarded hypodermal cell development observed in *let-7* mutant animals is reminiscent of the phenotype caused by mutations in the heterochronic gene *lin-29*. *lin-29* encodes a zinc finger transcription factor that specifies adult-specific patterns of cell lineage and cell differentiation. In wild-type animals, LIN-29 protein expression begins during the L4 and adult stages in the hypodermis. LIN-29 expression in the hypodermis of L4 stage *let-7* animals was reduced relative to the level of LIN-29 in L4 stage wild-type animals, showing that *let-7* acts upstream of *lin-29* in the heterochronic pathway.

By a combination of transgene complementation, RNA expression analysis, and mutant allele sequencing, we mapped *let-7* to a 2.5 kb region that could fully complement a *let-7* null mutant. Three *let-7* mutations map to a 20 base interval in this region. Comparison of the *C. elegans* and *C. briggsae* sequences capable of rescuing a *let-7* null mutant identified a 26 bp region flanking the *let-7(n2853)* point mutation is conserved between *C. elegans* and *C. briggsae*.

We detected a 21 nt RNA transcript by northern analysis of small RNAs separated on polyacrylamide gels and probed with an oligonucleotide from the conserved region flanking the *let-7* mutations. This 21 nt RNA is undetectable in the *let-7(mn112)* deletion mutant, strongly suggesting that *let-7(mn112)* is a null mutant. The *let-7(n2853)* mutation substitutes A for G at the fifth nucleotide of the *let-7* RNA. The weak mutation *mg279* is a 27 bp deletion upstream of the *let-7* transcript (Fig.2) that may regulate the level of *let-7* transcript expression.

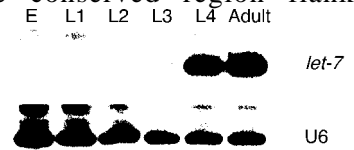


Figure The structure and temporal regulation of the 21 nt *let-7* regulatory RNA. a, Northern blot of total RNA from mixed stage wild-type (lane 1), *let-7(n2853)* (lane 2), *lin-28(n719)* (lane 3), and *lin-28(n719); let-7(mn112)* (lane 4) probed with an oligonucleotide probe. *lin-28(n719)* was used to suppress the lethality of *let-7(mn112)* mutants. b. *let-7* is temporally regulated. Shown below are U6 RNA loading controls.

*let-7* is expressed only at late larval and adult stages. This expression profile is consistent with the *let-7* mutant phenotype, which affects development specifically in late larval and adult stages. Expression of *let-7* in late larval stages also coincides with the critical period for *let-7* function in viability, as determined by temperature-shift experiments.

Given the genetic interactions observed between *let-7* and other heterochronic genes, and the precedence for direct interaction between the *lin-4* regulatory RNA and genes in the heterochronic pathway, we searched for complementary regions between the *let-7* RNA and the mRNAs of these heterochronic genes that would suggest a direct regulatory interaction. Five heterochronic genes contain sequences complementary to *let-7* in their 3' UTRs.

The results from genetic epistasis experiments are consistent with a model in which the *let-7* complementary sites in the 3' UTRs of heterochronic genes mediate direct regulatory interactions with the *let-7* regulatory RNA. While the informatic analysis does not distinguish which of these potential targets are key to the *let-7* gene function, genetic analysis shows that it is dysregulation of *lin-41* in a *let-7* mutant that causes most of the lethal and heterochronic phenotypes: 1) loss of function mutations in *lin-41* constitute the majority of genetic suppressors identified in unbiased screens for mutations that suppress the lethal and

heterochronic phenotypes of *let-7*, and 2) increasing the gene dose of *lin-41* causes the distinctive lethal and heterochronic phenotypes of a *let-7* loss of function mutant. *lin-41* also acts in late larval development and is down-regulated at the time of *let-7* expression.

**Genetic and molecular identification of *lin-41* which encodes an RBCC possible RNA binding protein and is down-regulated at the time of *let-7* regulatory RNA expression.**

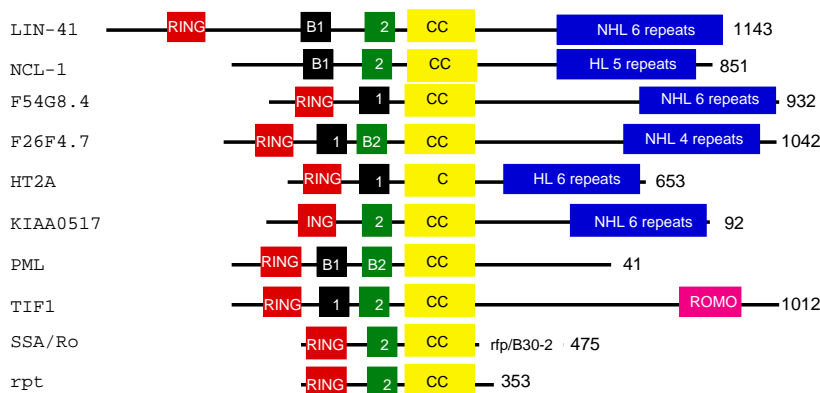
**Frank Slack (now on the faculty at Yale), in collaboration with Michael Basson and Bob Horvitz at MIT**

We sought genes regulated by the *let-7* regulatory RNA by genetic suppression of the *let-7* lethal phenotype. One class of mutants, isolated independently in seven lines, was distinguished because they semi-dominantly suppress the lethality and the heterochronic defects of *let-7* mutations. Genetic mapping and complementation tests showed that these mutations are all alleles of the same gene, *lin-41*.

*lin-41* null mutations cause precocious expression of late stage cell lineages in the hypodermis of the animal. Animals carrying a higher than normal dosage of *lin-41* show a retarded phenotype in the hypodermis. In some cases, the animals die by bursting through the vulva in a manner that resembles the *let-7* lethal phenotype. Thus, overexpression of *lin-41* causes the opposite heterochronic defect to animals missing *lin-41* gene activity, showing that LIN-41 is necessary and sufficient for the repression of adult-specific fates during earlier larval stages. These data also argue that *lin-41* is the major *let-7* regulated output for heterochronic patterning.

*lin-41* encodes a member of a large family of RBCC (Ring finger-B box- Coiled coil) proteins, which contain an N-terminal RING finger (a zinc-chelating domain thought to be involved in protein:protein and/or protein:nucleic acid interactions), a pair of B boxes (additional zinc chelating domains), and a coiled-coil domain. LIN-41 and other RBCC proteins also contain a C-terminal domain that consists of six copies of a 44 amino acid repeat. Most significantly for the rather orphaned heterochronic gene pathway, there is a possible *Drosophila* LIN-41 homologue with similarity across almost the entire protein (accession number AC004280: 48% identity; 67% similarity; TBLASTN P value  $5 \times 10^{-62}$ ), as well as a possible mouse homologue gene defined by 2 ESTs (accession numbers AA930787 and AA919390: 54% identity; 78% similarity; P value  $4 \times 10^{-25}$  over a smaller region of the protein). These genes suggest that the function of LIN-41 has been conserved across animal phylogeny. The *lin-41* homologs may serve an analogous function in temporal patterning, or in the *lin-41* germline function (all null mutant alleles cause sterility due to failure in oocyte development), or both. Significantly, many members of the RBCC family associate with RNA, or are implicated in RNA regulation, including the *C. elegans* NCL-1 RBCC protein, a cytoplasmic protein that regulates ribosomal RNA and rRNA synthesis through an unknown mechanism. The RING finger zinc finger domain of LIN-41 is a potential nucleic acid binding domain that could mediate its regulation of *lin-29* mRNA.

Figure A. LIN-41 is a distantly related member of the RBCC family.

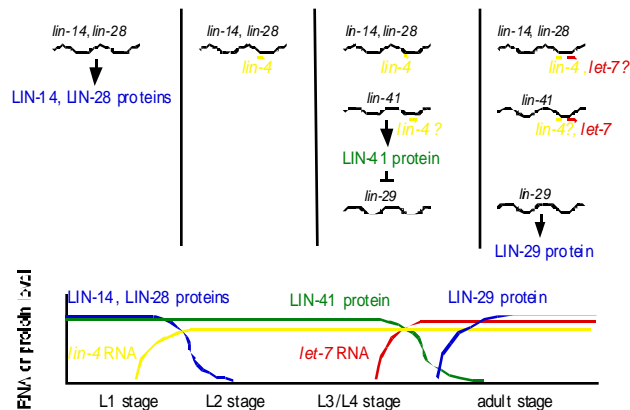


Our genetic and molecular epistasis studies support the model that *lin-41* negatively regulates *lin-29* to control the switch to adult hypodermal cell fates. *lin-41* is expressed in the same cells as *lin-29* (as well as other cells) and its expression is down-regulated at the time that *lin-29* translation is up-regulated. A fusion of the green fluorescent protein (GFP) gene to the very N-terminus of the full-length *lin-41* gene (which rescues a *lin-41* mutant and thus is a

functional fusion protein) is widely expressed, predominantly cytoplasmic, and is down-regulated in hypodermal cells during the L4 stage. This down regulation is mediated by the sequences in the *lin-41* 3' UTR that are complementary to the *let-7* regulatory RNA. Thus *lin-41* transduces temporal information from the *let-7* regulatory RNA to the *lin-29* transcriptional output.

The heterochronic gene pathway contains at least two regulatory RNAs; the *lin-4* RNA acts early, and the *let-7* RNA acts later in development. This observation suggests that a cascade of stage-specific regulatory RNAs controls the temporal sequence of cell lineages and differentiation in *C. elegans*. Expression of *lin-4* RNA before the second larval molt negatively regulates LIN-14 and LIN-28 levels to signal a transition from L1 and L2 to later larval patterns of cell lineage and differentiation. And *let-7* RNA expression during the L4 and later stages negatively regulates *lin-41* translation to signal the transition to late larval and adult stages. Strikingly, the *lin-4* and *let-7* RNAs show no sequence similarity in their RNA products and only very limited homology in their proximal transcriptional regulatory elements, leaving no evidence of a common ancestor to these similarly functioning regulatory RNA genes. It will be interesting to explore if other regulatory RNAs mediate other temporal transitions, for example at the L2 or L3 stages.

Figure A model for the successive regulation of heterochronic gene activities by the *lin-4* and *let-7* RNAs. LIN-14 and LIN-28 expression levels are decreased by *lin-4* at the end of the first larval stage to allow progression to late larval stages. In late larval stages, the expression of LIN-41 and perhaps other genes may be similarly down-regulated by the *let-7* RNA, relieving their repression of LIN-29 translation and allowing progression to the adult stage.



The involvement of two antisense RNAs in this pathway is very striking. Regulatory RNAs like *lin-4* and *let-7* may also be more general in animal species and may perform the same biochemical function in translational control. Our

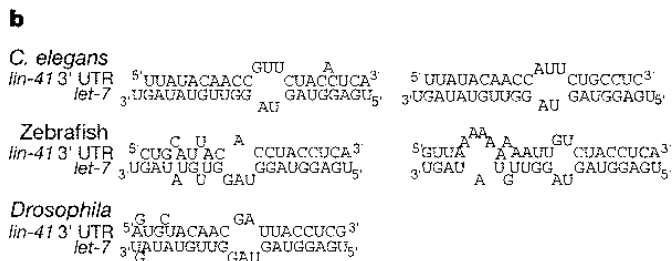
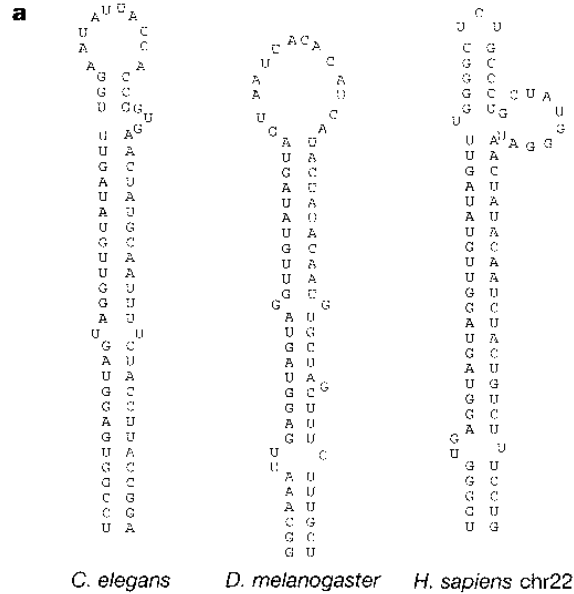
preliminary observation of possible mammalian and other invertebrate *let-7* homologues favor this view. Noncoding RNAs in eukaryotic organisms are involved in translation (the ribosomal and tRNAs), RNA processing (the Un RNAs) and RNA methylation/pseudouridination (the snoRNAs). Regulatory RNAs such as *let-7* tend to be missed by genome sequence analysis and this is a major problem for genomics. Both *lin-4* and *let-7* were identified genetically, but the mutants are rare and their detection depended on large genetic screens or selections. Our analysis of the key features of these regulatory RNAs, their promoter elements, the key loops and bulges for function, the regions complementary to target mRNAs, may enable the informatic detection of analogous but not homologous regulatory RNAs in genome databases.

### Conservation across animal phylogeny of the sequence and temporal expression of the 21 nucleotide *let-7* heterochronic regulatory RNA

Amy Pasquinelli and a host of collaborators from around the world

The sequence and function of both the *lin-4* and *let-7* small RNAs are conserved in the nematode *C. briggsae*. BLASTN searches reveal one DNA segment from the *Drosophila melanogaster* genome sequence, three segments from the human genome sequence on chromosomes 9, 11, and 22 bearing exact sequence matches, and two other human segments on chromosomes 9 and 21 with 20/21 matches to the *let-7* RNA. Database searches did not detect potential *lin-4* homologues in any genus except the *Caenorhabditae*. Similar stem-loop secondary structures are predicted for precursor transcripts of *Caenorhabditae*, *Drosophila* and human *let-7* RNAs. The mature 21 nt. *let-7* RNA may be efficiently processed from this putative precursor because only the mature RNA is detected in *C. elegans* and most other species (see

below). A similarly structured larger transcript is also predicted for the *lin-4* RNA, and rare transcripts that may correspond to it have been detected in *C. elegans*.



Potential *let-7* complementary sites are present in the 3' UTRs of both the *Drosophila* and zebrafish *lin-41* cDNAs.

*let-7* RNA transcripts ~21 nt. are expressed in human tissues and *Drosophila*.. As in *C. elegans*, only RNAs in the ~21 nt. range were detected. Also consistent with the database searches, no *lin-4* RNA was detected in other non-nematode species.

The expression and function of the *let-7* RNA in *C. elegans* begins during the third larval stage, when the gene specifies a transition from late larval to adult cell fates, and continues at all subsequent stages. Expression of *Drosophila let-7* is also temporally regulated: *let-7* RNA is absent until the late third instar, just prior to metamorphosis. At the early pupal stage there is at least a 10 fold increase in *let-7* RNA expression that is sustained to adulthood. Thus temporal regulation of *let-7* is similar in *Drosophila* and *C. elegans*, distantly related members of the ecdysozoan clade.

The size and temporal regulation of *let-7* are also conserved in the more distantly related lophotrochozoan clade, members of which do not moult but have larval and adult stages. For example, in two mollusc species, and in a polychaete annelid, *let-7* RNA is expressed at the adult stage, but not at larval stages. Thus, *let-7* may function at later stages of these species to regulate developmental progression to the adult.

Vertebrates do not develop through classically defined larval stages. Nevertheless, in zebrafish expression of *let-7* RNA is also temporally regulated: expression commences in the developing embryo between 24 to 48 hours post-fertilization and continues with strong expression at the adult stage. Analyses of other deuterostomes shows expression of a 21 nt. *let-7* RNA in other vertebrates, as well as in urochordate ascidians, and in a hemichordate. In the echinoderm *Strongylocentrotus purpuratus*, *let-7* may be regulated at the level of precursor RNA processing rather than transcription; a ~100 nt. *let-7* RNA is detected during embryonic and early larval development, but at the adult stage the 21 nt. *let-7* RNA as well as possible processing intermediates appear.

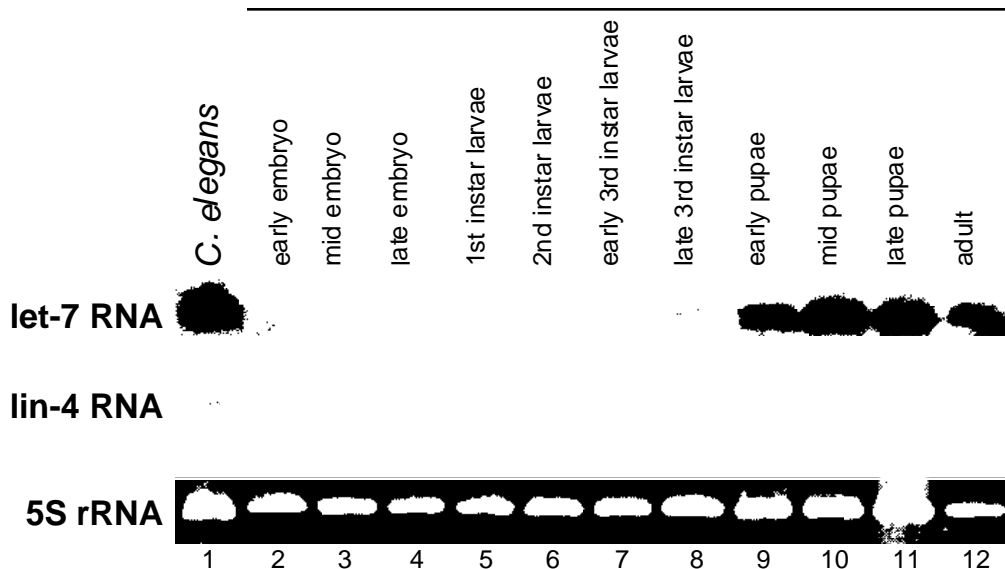
We did not detect convincing *let-7* RNA in anthozoan and hydrozoan cnidarian species nor in two poriferan species. Plant and unicellular organisms also failed to show *let-7* RNA expression, consistent with database searches for those species that have been completely sequenced.

Because all three major clades of bilaterian animals express a *let-7* RNA that is temporally regulated, but cnidarian, poriferan, and all non-animal species that we analyzed do not express a detectable *let-7* RNA, we favor the hypothesis that the gene evolved after the divergence of diploblastic and bilaterian animals.

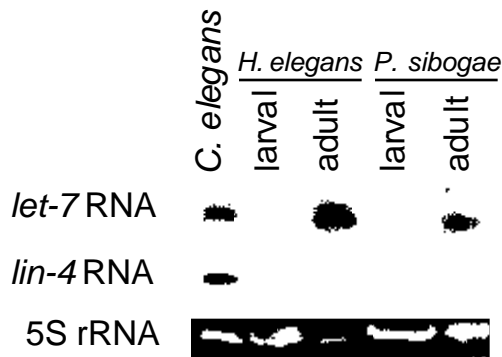
While the observation that *let-7* homologous RNAs across phylogeny are temporally regulated does not prove that these RNAs actually specify temporal patterning in this broad range of species, the conservation of sequence, of a longer structured precursor, of the 21 nucleotide length, of temporal regulation, and of complementary sites in the *lin-41* target in two ecdysozoan and one chordate species are strong evidence of a conserved function. There are candidate mutations that map to the location of the *Drosophila let-7* and *lin-41* genes which may provide an avenue to test the function of these genes in another ecdysozoan. The *let-7* RNA is conserved across bilaterian phylogeny but the earlier acting *lin-4* RNA is not. Consistent with this, the *let-7* target *lin-41* is conserved in *Drosophila* and vertebrates, whereas the *lin-4* target *lin-14* appears to be unique to nematodes. The *let-7*-regulated late larval transition in *C. elegans* may be ancestrally related to late transitions in other species, for example from larval to reproductive forms, whereas the earlier *lin-4/lin-14*-regulated transition may be a recent invention of the nematode phylum. Alternatively, *lin-4* and *lin-14* may evolve more quickly.

The 21 nt. length of the *let-7* RNA is highly conserved, indicating that this size is central to its function. It may be significant that this length is similar to the 21-25 nt. RNAs observed during RNAi-directed down-regulation of target mRNAs.

*D. melanogaster*

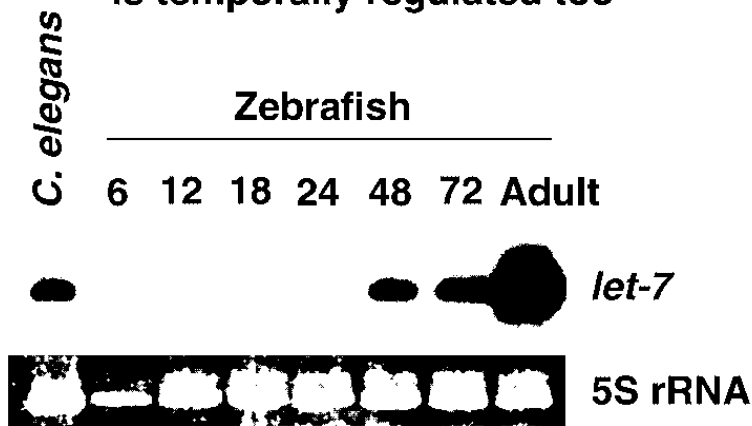


# Temporal regulation of *let-7* RNA in indirectly developing lophotrochozoans



developmental transitions.

## The vertebrate *let-7* is temporally regulated too



Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing  
 Amy Pasquinelli and Alla Grishok from the Mello lab

A paradigm to emerge from the study of RNAi in plants and animals is that of the small RNA guide that can direct an RNA-Protein complex to a complementary target sequence. Guide RNA molecules approximately 22 nucleotides, recently termed “small interfering RNAs” (siRNAs) have been identified in *Drosophila* that copurify with and provide sequence specificity to an RNase complex that degrades the target mRNA. Furthermore, small synthetic dsRNAs of 22-26 nts are sufficient to direct destruction of complementary RNAs both *in-vitro* and *in-vivo* and duplexes of small 21nt RNAs suppress gene expression in cultured mammalian cells.

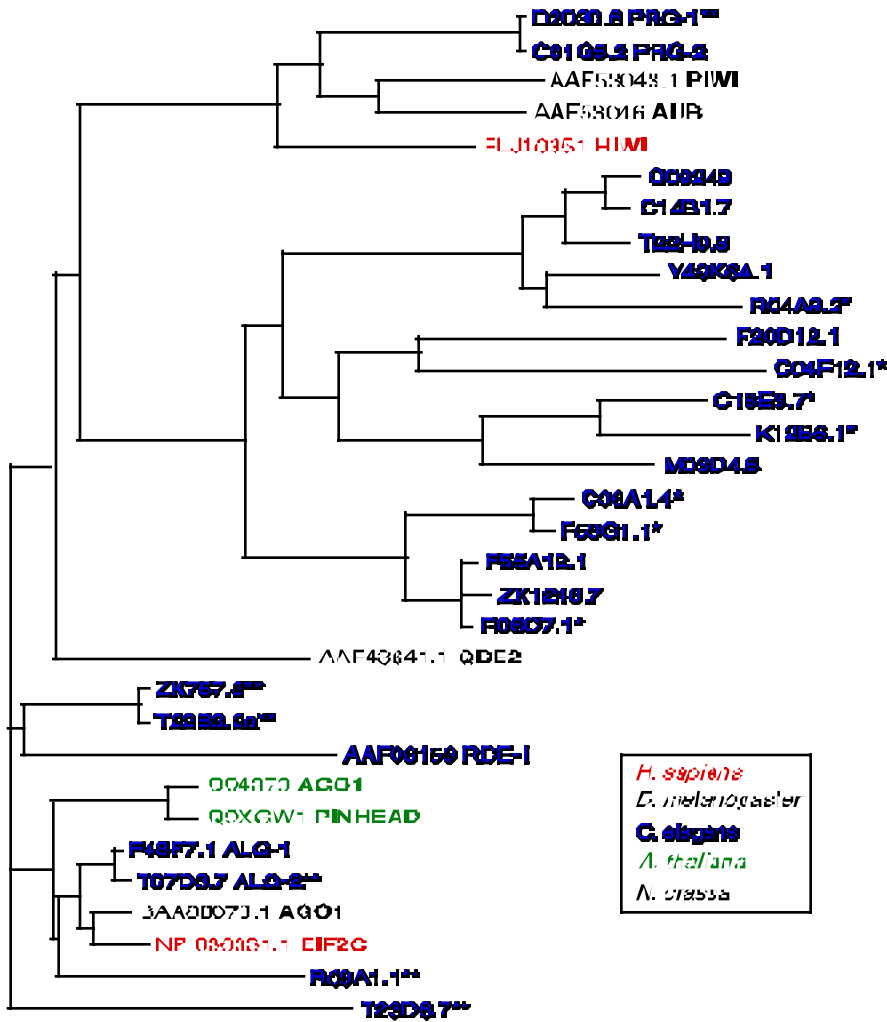
It is tantalizing that the size of the siRNAs implicated in RNAi is similar to the approximately 22 nucleotide size of the *lin-4* and *let-7* small temporal RNAs (stRNAs) that regulate *C. elegans* developmental timing. Although the 22 nucleotide forms of the *let-7* and *lin-4* RNAs are more abundant, low levels of larger transcripts of approximately 70 nucleotides can also be detected and are predicted to fold into similar stem-loop structures. Human and *Drosophila let-7* orthologs are also expressed as larger forms that, likewise, have the potential to fold into stable stem-loop structures, suggesting that this potential secondary structure may have functional importance.

The Dicer protein has been implicated in RNAi in *Drosophila* where it functions in the processing of longer dsRNAs into the siRNAs which subsequently guide mRNA destruction. Dicer belongs to a conserved family of proteins, whose members contain a helicase domain, one or two dsRNA-binding domains, and two RNase type III domains. Also present in Dicer family members is a PAZ domain, which was identified in the Piwi/Argonaute/Zwille/RDE-1 family of proteins introduced below..

Genetic studies in *C. elegans* have identified *rde-1*, (for RNAi defective); null mutations in it cause a complete lack of RNAi but no other discernible phenotypes. *rde-1* encodes a 1020 amino acid protein that is a member of a large family of proteins found in a wide range of eukaryotes. Members of the RDE-1 family have two conserved domains of unknown biochemical function. The 300 amino acid PIWI domain located in the C terminal region of these homologs shows the highest degree of sequence conservation. The 110 amino acid PAZ domain is located N terminal to the PIWI domain and is also found in the Dicer family of proteins. RDE-1 homologues in the fungus, *Neurospora*, and the plant, *Arabidopsis*, have also been implicated in PTGS mechanisms, suggesting that RDE-1 family members not only share conserved structures but also have conserved functions in gene silencing in three kingdoms of eukaryotic organisms.

Mutations in *rde-1* homologs have also been shown to have developmental consequences. For example, in *Drosophila*, the *agol* gene is required for embryogenesis, the *piwi* gene is required for the maintenance of the germline stem-cell population, and *aubergine* is required for the proper expression of the germline determinant Oskar. In *Arabidopsis* two very similar genes, *argonaute* (*agol*) and *pinhead/zwille*, are required for stem-cell patterning of the plant meristem. *argonaute* is also necessary for PTGS in *Arabidopsis*. The *C. elegans* genome contains 23 homologs of *rde-1* including orthologs of both *piwi* and *agol*. The pleiotropic nature of the defects associated with loss of function mutations in members of this family could reflect discrete regulatory functions in numerous developmental events or alternatively might reflect a more general misregulation of silencing mechanisms that are necessary to insure proper stem-cell maintenance and differentiation.

There are 23 *C. elegans* homologs of *rde-1*. dsRNAs derived from two closely related genes F48F7.1 and T07D3.7, which we have named *alg-1* and *alg-2* (for argonaute like genes), induced developmental phenotypes in the progeny of injected animals, including a phenotype of bursting at the vulva, and a lack of the adult specific alae, longitudinal stripes that run the length of the cuticle on both sides of the adult animal.



We also assayed *alg-1* and *alg-2* for possible roles in RNAi. Inhibition of *alg-1* or *alg-2* by RNAi did not suppress RNAi targeting a second gene. These findings suggest that *alg-1* and *alg-2* are not necessary for RNAi. Nevertheless, it remains possible that these genes might have some redundant function in RNAi with *rde-1* or with other members of this gene family.

*C. elegans dcr-1* functions in development and RNAi. The *C. elegans* gene K12H4.8, which we have named *dcr-1*, is predicted to encode a protein related to the *Drosophila* Dicer and the *Arabidopsis* Carpel Factory, proteins implicated in RNAi and regulation of development, respectively. *dcr-1(RNAi)* induced developmental abnormalities during larval growth that were very similar to those induced by *alg-1/alg-2(RNAi)*. These included a protruding and non-functional vulva, and a tendency to burst at the vulva shortly after the molt from the larval to the adult stage. In addition, *dcr-1(RNAi)* animals frequently exhibited faint or missing adult-specific alae.

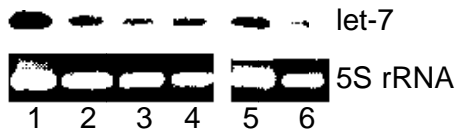
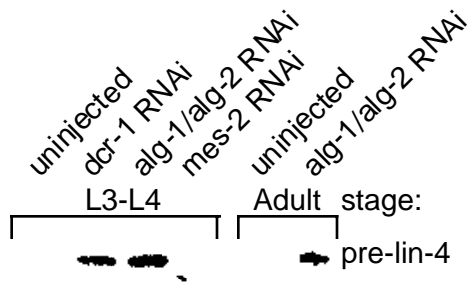
We next asked if *dcr-1(RNAi)* might sufficiently reduce *dcr-1* activity to cause an RNAi deficient phenotype. For this assay, we injected *dcr-1* dsRNA into adult hermaphrodites and then assayed for sensitivity to RNAi targeting a second gene. In experiments targeting two different genes we observed a significant reduction of RNAi among the progeny of *dcr-1(RNAi)* animals but not among control animals injected with unrelated dsRNAs. These results support the findings that implicate *Drosophila* Dicer in RNAi and suggest that DCR-1 may have a similar activity in *C. elegans*.

*dcr-1(RNAi)* and *alg-1/alg-2(RNAi)* cause retarded heterochronic defects. RNAi of either *dcr-1* or *alg-1/alg-2* resulted in adults with extra seam cells that arise from reiterated L2 type divisions. We consistently observe inappropriate seam cell division patterns in L3 through later stages in *dcr-1(RNAi)* and *alg-1/alg-2(RNAi)* animals. The similarity of phenotypes described above to those of the heterochronic genes *lin-4* and *let-7* raised the possibility that *alg-1*, *alg-2* and *dcr-1* might act upstream of the *lin-4* or *let-7* stRNAs or might be necessary for their

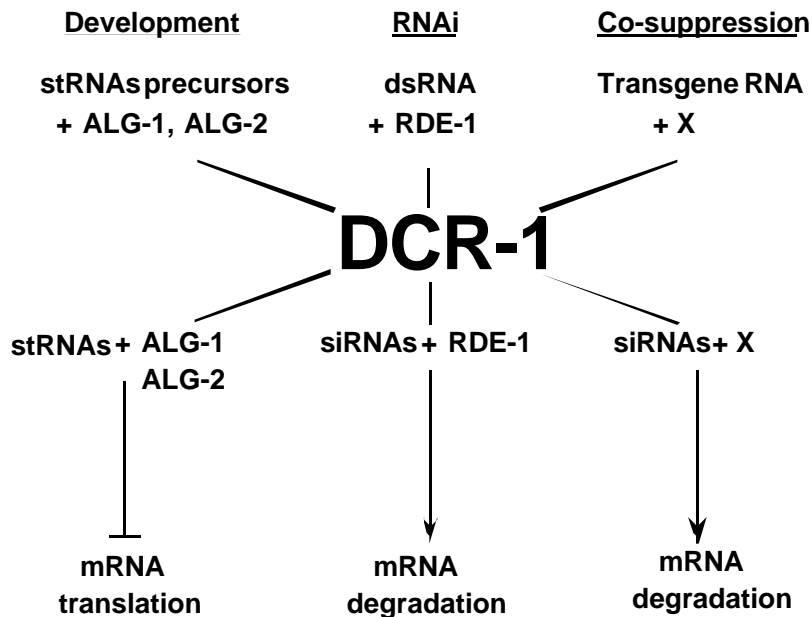
regulatory activities. Because the retarded phenotypes of *lin-4* and *let-7* are caused in part by failure to downregulate their target genes, mutations in *lin-14* and *lin-41* partially suppress the *lin-4* and *let-7* mutant phenotypes. Indeed, we found significant suppression of the RNAi-induced *alg-1/alg-2* and *dcr-1* heterochronic phenotypes including alae and vulval defects by the *lin-14(n179)* and *lin-41(ma104)* non-null mutations. These findings are consistent with the idea that the retarded heterochronic phenotypes induced by *alg-1/alg-2* and *dcr-1 (RNAi)* are caused, at least in part, by misregulation of *lin-14* and *lin-41*.

Elements in the 3'UTRs of *lin-14* and *lin-41* mRNAs are responsible for negative regulation mediated by the *lin-4* and *let-7* stRNAs. If *alg-1*, *alg-2* and *dcr-1* are necessary for *lin-4* and *let-7* function, then we would expect misregulation of reporter genes that carry the *lin-14* and *lin-41* 3'UTR elements. In fact, *lin-14* 3' UTR and *lin-41* 3'UTR reporter genes are up regulated in *dcr-1* and *alg-1/2* RNAi experiments. The findings that reporter genes bearing the *lin-14* and *lin-41* 3'UTRs are up-regulated by *dcr-1* and *alg-1/alg-2* inhibition, together with the observation that *lin-14* and *lin-41* mutations suppress the retarded heterochronic phenotypes caused by *dcr-1* and *alg-1/alg-2* RNAi are consistent with the model that *dcr-1*, *alg-1* and *alg-2* function in the *lin-4* and *let-7* pathway to regulate larval development.

*dcr-1* and *alg-1/alg-2(RNAi)* animals exhibit defects in stRNA processing. We collected progeny from mothers subjected to *dcr-1(RNAi)* or *alg-1/alg-2 (RNAi)* and performed Northern blot analyses to monitor the size and abundance of the *lin-4* and *let-7* RNAs. Both *dcr-1* and *alg-1/alg-2(RNAi)* animals exhibited a marked accumulation of the *lin-4* long-form at both L3-L4 and adult stages. The same RNA preparations from the *dcr-1* or *alg-1/alg-2 (RNAi)* animals were probed for the expression of *let-7*. We found that, as with *lin-4*, *let-7* processing depends on *dcr-1* activity but, in contrast, did not depend on *alg-1/alg-2* activity. The quantity of mature *lin-4* and *let-7* stRNAs consistently was reduced in RNA populations prepared from *alg-1/alg-2(RNAi)*, *dcr-1(RNAi)* and *dcr-1(ok247)* animals. These findings suggest *alg-1/alg-2* activities may be more important for the stability or function of *let-7* stRNA than for its processing from the larger form.



RNAi and natural tiny RNAs. These data show that ALG-1/ALG-2 act upstream of the ribonuclease DCR-1 in the processing of the stRNAs, and suggests that the other RDE-1/Argonaute orthologs may also act upstream of DCR-1 in the processing of siRNAs and other tiny RNAs. Thus, RNAi and stRNA expression share a requirement for DCR-1 activity for the production of small guide RNAs, while RDE-1 and its homologues provide parallel functions in the RNAi and other tiny RNA pathways, respectively (Figure 6). Our findings are consistent with a model in which members of the RDE-1 and DCR-1 families act not only in gene silencing but also with naturally expressed dsRNAs to execute cellular and developmental gene regulatory events.



Although there are compelling similarities between RNAi and developmental regulation by *lin-4* and *let-7* there are also several important differences. In RNAi, the dsRNAs utilized, typically contain long stretches of perfect base pairing, whereas stRNA precursors, however, are predicted to be bulged and looped. Whereas, cleavage of the perfectly base-paired RNAs that initiate RNAi yields both sense and antisense, or potentially double-stranded siRNAs, only one strand of the *lin-4* and *let-7* stRNAs is detected. Thus, after generation of the mature stRNA the remaining sequences must undergo rapid degradation. The RNAi and stRNA pathways also induce distinct outcomes: RNA destruction versus translation inhibition.

There are 24 members of the RDE-1/AGO1/PIWI family in *C. elegans*. The degree of conservation between certain members of this family is striking. For example, ALG-1 and ALG-2 exhibit 41% identity with AGO1 from *Arabidopsis* and 67-69% identity with AGO1 relatives in animals. The fact that divergent members of this family including *rde-1*, *qde-2* and *ago-1* all function in gene silencing suggests that PTGS mechanisms represent an important ancestral function of genes within this family. We speculate that the *Drosophila* genes *piwi*, *aubergine* and *ago1*, the *Arabidopsis* gene *ago1*, and perhaps many other members of this family in *C. elegans* and other organisms may similarly have small endogenous RNA co-factors with which they function to regulate specific target mRNAs.

While there are 24 members of the RDE-1/Argonaute gene family in *C. elegans*, based on a current census from the nearly complete genome sequences, there are 7 in *Arabidopsis*, 4 in *Drosophila*, and 2 in humans. Only the Piwi and Argonaute subtypes are conserved in many species; RDE-1 as well as most of the other *C. elegans* family members are divergent from the family members of other eukaryotes. Whether the ancestral function of RDE-1 related genes was in developmental control or RNA interference, it is clear that a great expansion of the family has taken place in the phylogenetic lineage to *C. elegans*. There may have also been an associated expansion in the family of tiny RNAs that may act with these proteins. Whether the ancestral function of RDE-1 related genes was in developmental control or sequence directed immunity, it is clear that a great potential exists for exploiting these proteins, along with small RNAs as guides, to direct the regulation of specific gene targets in the cell.

RDE-1 plays an upstream role in the initiation of interference in response to dsRNA in *C. elegans*. ALG-1 and ALG-2, may play a similar upstream role in the *lin-4* and *let-7* stRNA pathways. One attractive possibility is that these diversified factors provide specificity to their respective pathways. This might involve a role in the recognition of the distinct trigger sequences or in insuring that the processed small RNAs are assembled into distinct downstream complexes. Perhaps members of the RDE-1 family remain associated with the RNA sequences throughout processing and provide specificity needed to insure that the small RNAs produced are targeted to the appropriate downstream complex, for example to mediate mRNA destruction vs translation inhibition.

DCR-1 has several motifs that might be expected in a dsRNA processing enzyme, including a helicase, a dsRNA binding domain and two RNase III type dsRNA exonuclease domains. Thus, DCR-1 functions in multiple pathways important for developmental and PTGS mechanisms, and may be guided in its processing of distinct substrates by members of the RDE-1 family. Perhaps the embryonic larval lethal phenotypes associated with *dcr-1* inhibition and the developmental phenotypes associated with the *Arabidopsis* homolog, *caf 1*, reflect a role for members of this gene family in the processing of other as yet unidentified small regulatory RNAs. Thus tiny RNAs may regulate a broader range of gene regulatory and developmental events than the temporal transitions mediated by the founding members of the class, the *lin-4* and *let-7* stRNAs. It is likely that the ramified family of RDE-1/AGO1/PIWI related proteins has co-evolved with numerous small RNA encoding genes analogous to *lin-4* and *let-7*, and that many such genes await discovery in plant and animal genomes.

### **Characterization of the effect of early expression of the small RNA *let-7* on developmental timing in *C. elegans*.**

**Gabriel Hayes**

Timing of development in *C. elegans* is controlled by the expression of the small RNAs *lin-4* and *let-7*. Expression of *lin-4* during the L1 stage downregulates the expression of *lin-14* and *lin-28* to allow the execution of L2 developmental programs. Similarly, expression of *let-7* during the L3 stage downregulates expression of *lin-41*, relieving repression of *lin-29* and allowing the execution of adult developmental programs. *lin-4* and *let-7* encode RNA precursors that are processed to 22- and 21-nucleotide mature transcripts, respectively, that function by base-pairing to the 3' UTR of their targets.

In order to answer the question of whether expression of *let-7* is sufficient to drive the execution of adult developmental programs, we have designed a construct to drive the early expression of *let-7* during the L1 phase when *lin-4* is normally expressed. We have fused the putative *lin-4* promoter and *lin-4* precursor to the 5' end of the *let-7* precursor. This construct was injected into wild-type (N2) animals and the resulting transgenic lines examined for heterochronic phenotypes. We expected that early expression of *let-7* might result in a *lin-41* loss of function phenotype (precocious development of adult alae, and sterility), since early downregulation of *lin-41* expression should occur. Our transgenic animals develop a protruding vulva as young adults and exhibit egg-laying defects but do not exhibit the classic *lin-41* phenotype. Therefore, we will determine whether a construct including more of the genomic region upstream of the *lin-4* precursor provides stronger expression of *let-7*, resulting in the expected phenotype for early *let-7* expression. We would then screen for mutations that suppress the phenotype caused by early expression of *let-7*.

We are currently determining whether the injected arrays provide *lin-4* and *let-7* activity sufficient to rescue mutations in these genes. If the *let-7* transgene is expressed early, a putative negative regulator of *let-7* gene action could be responsible for the lack of a precocious phenotype. We would then perform a screen for mutations that allow early expression of *let-7* to result in a phenotype suggestive of early execution of adult developmental programs.

### **Genetics and genomics of heterochronic regulatory RNAs**

**Brenda Reinhart, Gabe Hayes, and John Aach and Yonaton Grad and George Church**

One tissue affected by the heterochronic pathway is the lateral hypodermis, or seam. A wild-type animal hatches with 10 seam cells that divide once in the first, third and fourth larval stages, with the anterior daughter fusing with *hyp7* and the posterior daughter remaining in the seam to keep the seam cell number constant. However, at the second larval stage, the V cells undergo a double division that increase the total number so that when the seam cells cease divisions in the fourth larval stage, the final number of seam cells is 16. Heterochronic mutants can result in a change in the number of seam cells at the late L4/young adult stage in two ways, either by altering the timing or expression of the L1 and L2 stage or by inappropriately reiterating cell divisions as young adults. *lin-4(lf)* animals reiterate L1 divisions and fail to execute L2 patterns of division, and the number of seam cells remains at 10 in late larval stages.

let-7(lf) animals reiterate seam cell divisions as young adults, resulting in a temporary increase in seam cell number before the anterior daughters fuse with hyp7. We screened directly for mutants which have an altered number of seam cells in the late L4/young adult stage in a background carrying a seam cell-specific GFP marker (J. Rothman, personal communication). Such a screen could isolate mutations in genes involved in the production or function of the small RNAs and their regulatory targets but also allows isolation of mutations that may be tissue specific rather than global regulators of temporal development. From a pilot screen of 1500 haploid genomes, we recovered three independent alleles that alter the number of seam cell nuclei in late L4s/young adult. One has up to 22 seam cell nuclei as an adult, and the other two both have 10-11 seam cell nuclei. We are currently analyzing the mutant phenotypes to determine which developmental stage is affected.

We are also using a computational screen to search for other small RNAs in the *C. elegans* genome. Both the *let-7* and *lin-4* genes produce >60nt precursors that fold into stem-loop structures, and the smaller RNAs are processed from the 5' end of the precursor. Our screen looks for ~100nt regions of the genome that could fold into low free energy structures and then checks for conservation of sequence and structure in the *Drosophila* genome. A test run of the program on the X chromosome identifies hundreds of such regions, but the only region which folds into a long stem loop structure with sequence conservation at the 5' end is in the *let-7* gene. We are currently screening the remaining chromosomes.

### **Future directions**

How do the *lin-4/lin-14* and the *let-7/lin-41* RNAs generate temporal transitions in the levels of LIN-14 and LIN-41 proteins?

We have already proven that the *lin-4* regulatory RNA and *lin-14* mRNA interact in vitro and in vivo to control the translation of the *lin-14* mRNA. We are testing genetically and biochemically whether the *let-7* regulatory RNA and *lin-41*RNA form an RNA duplex *in vivo*, and whether this antisense interaction also operates at a post transcriptional level. We are testing for similar in vitro binding of the *let-7* RNA to its possible targets *lin-41*, *lin-14*, *lin-42*, *lin-28*, and *daf-12*.

### **Structure/function studies of *let-7* using an *in vivo* assay of temporal gradient formation**

i. Transcriptional regulatory sequences in *let-7*:

We do not yet know if *let-7* and *lin-4* expression utilize RNA polymerase I, II, or III. In the comparison of the 2.2 kb *C. briggsae* and 2.5 kb *C. elegans let-7* gene sequences that mediate potent rescuing activity, four 60 to 150 nt long regions of greater than 90% nucleotide sequence identity are obvious upstream of the *let-7* RNA transcript. We expect that deletion of elements crucial to the expression of *let-7* will disrupt or partially disrupt the ability of a derivative to rescue. In addition, some elements may keep *let-7* off at times earlier than the late L3 stage.

Informatic and genomic identification of other *lin-4* and *let-7* RNA target genes.

Given the training set we now know: *lin-14*, *lin-41*, *lin-42*, *daf-12*, and *lin-28*, we expect that bona fide targets will include multiple examples of both *let-7* and *lin-4* complementary sequences. We are now querying the *C. elegans* genome sequence for transcribed regions bearing sequences complementary to the *lin-4* and *let-7* regulatory RNAs. The *let-7* and *lin-4* complementary regions are sufficiently long in each of these genes that bona fide target genes are expected to emerge from such simple analyses, in distinction for example to the relatively low information content to transcriptional regulatory binding sites.

### **Genetic exploration of genes that act upstream or in concert with the heterochronic regulatory RNAs**

i. *let-7* like phenotype genetics:

At present little is known of how *let-7* causes the down-regulation of mRNAs with 3'UTRs harboring sequences that can base pair with *let-7* RNA. Specific factors may recognize the duplex formed by *let-7* RNA base paired to the target RNA and then inhibit expression of the mRNA, possibly at the translational level. Recruitment of such factors by *let-7* RNA could make the mRNA incompetent for translation by steric hindrance or else could lead to the modification

of the mRNA so that it cannot properly associate with the translational machinery. Clearly, identification by genetics or biochemical analyses of the presumed *let-7* RNA co-factors will be necessary to determine the mechanism by which specific genes are targeted for down-regulation by this small RNA.

We are following up on our genetic characterization of *llh* (*let-7* like heterochronic) mutants in our collection. These mutants identify candidate genes that may act upstream of, in concert with or be regulated by *let-7*, and are, thus, implicated in temporal control of cell fates. We want to know whether the mutations act upstream of *let-7* or act in concert with *let-7*. We expect that some of the genes identified in this screen will regulate the activity of the *let-7/lin-41* RNA duplex in the regulation of *lin-41* mRNA translation.

### **Genetic exploration of transcriptional regulation of *let-7* at the L4 and later stages**

The necessary down-regulation of LIN-14 protein levels at the L1 to L2 molt, and LIN-41 protein levels at the L4 to adult molt is triggered by the timed up-regulation of the regulatory *lin-4* and *let-7* RNAs at each specific molt. Thus the promoters of these genes are key to their temporal regulation and action in the pathway. We are genetically screening for mutations in genes that may encode the transcriptional regulatory factors of *let-7* or *lin-4* by genetic enhancer and suppression genetics. We expect that our molecular analysis will reveal a transcription factor gene for example involved in the genetic cascades of molting control.

### **Analysis of the mechanism by which the *let-7* and the *lin-4* regulatory RNAs regulate expression of target genes**

We need to establish whether *let-7* regulation of *lin-41* occurs at the level of translation or some other RNA modification. Observations that RNA levels of genes, including reporter constructs, that undergo *lin-4* RNA mediated down-regulation are unaffected, indicate that the control is likely to be at the level of translation. Because the RNAs are so short and are complementary to their targets, the universe of possible mechanisms by which the *lin-4* and *let-7* regulatory RNAs actually down-regulate translation is actually quite delimited. Thus small RNAs with complementarity to their targets have been implicated in just a few types of functions. These are precedents for how the *let-7* and *lin-4* RNAs could modify the translation of the *lin-41* and *lin-14* as well as other target mRNAs.