Endogenous small interfering RNAs in animals (2008)

RNA interference (RNAi), the process by which double-stranded RNA (dsRNA) is processed

into small interfering RNAs (siRNAs) that silence homologous transcripts.

Hmnn, seems like the paper classifies first by the **final protein mechanism** which the siRNA goes through, either AGO or Piwi. Yup, ultimately its either AGO or Piwi.

**From Alvin Paper:**

In Drosophila, small RNAs are sorted to AGO1 and AGO2 according to their duplex structures, and these two small RNA loading pathways compete with each other. There is a difference in the chemical structure of the 3’ nucleotide between AGO1- and AGO2-loaded species. Small RNAs in the AGO1 and AGO2 complexes have 2’-OH and 2’-O-methyl groups at their 3’ ends, respectively.

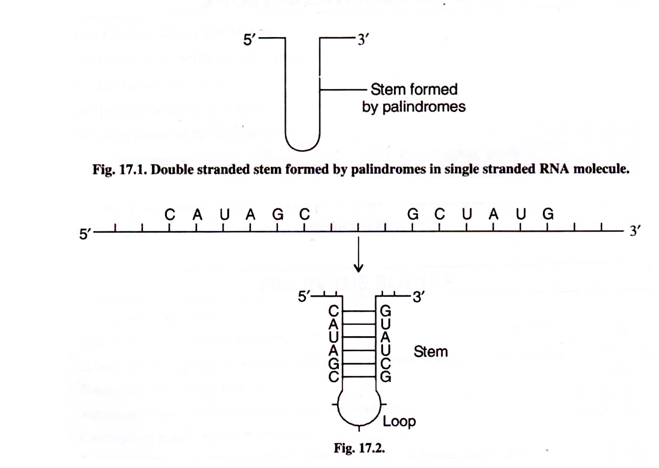
1. Argonaute-bound

Argo complex binds to siRNAs -> affect their regulatory targets.

2 types of Argo complexes, AGO and Piwi.

AGO1 binds to single-stranded processed-miRNAs ~22nt, which are processed primarily from <100nt IRs -> hairpin structure -> processed by (Drosha + Pasha) - > (**Dicer** + LOQs)-> processed- miRNAs

AGO2bins to single-stranded processed-siRNAs ~21nts. Firrst processed by (Drosha + Pasha) - > (**Dicer2** + LOQs)-> processed- siRNAs



1. Piwi bound

Piwi-interacting RNAs (piRNAs) are longer (~24–32 nt) and their biogeneis not exactly known.

Might want to look into this in the future, because of how their biogenesis seems to involve sense/ antisense processing. Interesting.

**Ok weird, so now we go into siRNA classification.**

1. **TE-based**

In D. melanogaster, deep sequencing of the small RNAs that directly associate with AGO2 (the Argonaute that mediates RNAi) revealed that TEs are a substantial source of RNAs of precisely ~21 nt.

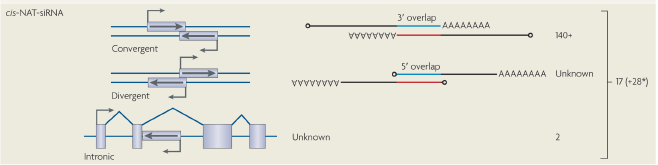
What are the innate function of these transcripts?

The precise structure of the double-stranded (ds)RNA substrates of small interfering (si)RNAs derived from transposable elements (TEs) is unknown, but hundreds or thousands of TEs are inferred to directly generate siRNAs.

1. **Cis-Natural (Nat) antisense transcripts**

*Cis*-natural antisense transcript (*cis*-NAT) arrangements are genomic regions that encode

exons on both DNA strands, and can involve 5′, 3 or internal exons.



K how do we understand this? Seems like the anti-sense strand

If its Convergent transcription, it’s the 3’-ends of the transcripts overlapping each other, while

If its divergent, it’s the 5’-ends

Drosophila melanogaster cis-NAT-siRNAs derive almost exclusively from 3′ untranslated region (UTR) overlaps, **BUT** two highly-expressed siRNA loci include annotated introns.

Then how do we explain intronic? Its exon + intron overlap?

Notable mentions:

1. CG7739/Ago2 gene pair, AGOs own transcript.
2. *Klar / thickveins* gene pair, 5’ exon overlap, internal transcript exons and/or annotated introns

The regulation of dsRNA formation is another mystery. For example, the *cis*-NAT-siRNA pathway accepts many substrates — at least 17 in mouse oocytes6,7 and at least 140 in *D. melanogaster*8,9,12. However, *cis*-NAT-siRNA loci constitute only 25% of co-express\d *cis*-NATs in *D. melanogaster*. NOTE, not all Cis-NAT genes are Cis-NAT siRNA

1. **Trans-NAT/Pseudogene**

Ahhh difference? Cis is right next to each other, while trans-nat is formed between transcripts that are produced from distinct genomic locations, and usually comprise an mRNA and an antisense-transcribed pseudogene



Could have potential for further exploration. Note papers from, Tarn *et al*.6 and Watanabe *et al*.7,15

Watanabe and colleagues describe 17 *cis* NAT siRNA loci7, but their precise categorization is ambiguous as many of them lack an annotated overlapping transcript. Tarn and colleagues describe another 28 mRNAs (\*) with siRNAs whose complementary transcript was not specifically described6. These might be *cis*- NATs, but some might represent *trans* natural antisense transcript (*trans* NAT) pairs.

Based on the cumulative data of Tarn *et al*.6 and Watanabe *et al*.7,15 *trans*-NAT gene-pseudogene pairs generate siRNAs (\*\*; some of the 28 mRNAs listed in the *cis*-NAT-siRNA category might have antisense

pseudogenes that were not reported

Multiple genes with antisense transcribed pseudogenes were inferred to anneal with their complementary progenitors (as trans-NATs) and be diced into siRNAs.

**##NOTE**, that when it comes to both cis/trans nat siRNAs, It is unclear whether the dicing of targets during *trans*-NAT-siRNA biogenesis accounts for target regulation, or whether pseudogene-derived antisense siRNAs actively slice sense-strand mRNAs.

1. **Hp RNA**

revealed a number of candidate loci that produce

small RNAs from extended inverted repeats that are termed hairpin RNAs (hpRNAs), the stems

of which were up to 400 base pairs in length10. At least seven distinct loci generate siRNAs,

and the hp-CG4068 locus alone encodes 20 tandem hairpins10–12. Despite their structural

similarity to miRNAs, hpRNAs are processed by DCR2 instead of DCR1, and generate 3′

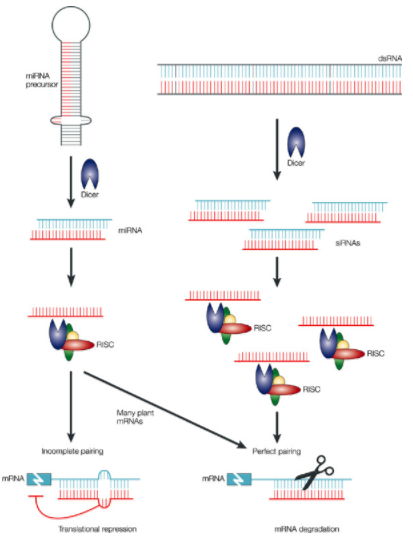
blocked siRNAs that load AGO2

AHHH, note that our siRNAs are always double-stranded, BUT internally produce hairpin structures.

They are effectively sorted, however, as long miRNAs make only a single small-

RNA duplex (as is typical for DCR1 substrates), whereas short hpRNAs produce multiple

duplexes (as is typical for DCR2 substrates).



**Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.**

By means of deep sequencing, we identify a largenumber of both 25–27-nucleotide Piwi-interacting RNAs

(piRNAs) and 21-nucleotide siRNAs corresponding to messenger RNAs or retrotransposons in growing oocytes.

siRNAs are exclusively mapped to retrotransposons or other genomic regions that produce transcripts capable of forming dsRNA structures. Inverted repeat structures, bidirectional transcription and antisense transcripts from various loci are sources of the dsRNAs. Some precursor transcripts of siRNAs are derived from expressed pseudogenes, indicating that one role of pseudogenes is in RNAi.

**Library construction: Total RNA seq from mouse oocytes**

**Interesting points:**

One peak was observed at **21 nucleotides, corresponding to the length of microRNAs (miRNAs) and siRNAs**, and the other at 25–26 nucleotides, corresponding to the length of piRNAs,

Annotation of the small RNAs revealed that both the 21-nucleotide and 25–26-nucleotide small RNAs were mainly derived from repeat sequences, most of which were retrotransposons.

Genomic mapping of the oocyte small RNAs (**excluding those with more than ten hits to the genome**, which **are likely to be repeat sequences**)

Within the non-piRNA cluster of 21nt, **the lack of a short stem-loop structure** (Supplementary Fig. 4), which is a characteristic of miRNA precursors, in most of the genomic sequences encompassing the 21-nucleotide small RNAs indicated that they are not miRNAs.

Clusters within mapped sequences were identified, found that a large cluster mapped to a particular pseudogene locus, which has an inverted repeat region, AND that these reads were specifically mapped to this inverted repeat region. They then concluded that since these transcripts are dependant on dicer and ago2, they are then termed as ‘hairpin siRNA (hp-siRNA) cluster.

**They also tried to find “cis-nat” and “trans-nat” clusters! WHAT ARE THE CRITERIA FOR THIS.**

**clusters using our in-house program designed to detect the siRNA mapping patterns expected for these (Supplementary Methods).**

At this locus, 135 clones of small RNAs comprising 117 different sequences were mapped to the first exon of the Kif4 gene (Fig. 3). Of these small RNAs, 93% were 19–22 nucleotides in length. A 59 rapid amplification of cDNA ends (59RACE) analysis of the transcripts from growing oocytes revealed that the first exons of Pdzd11 and Kif4 overlapped (Fig. 3). Notably, almost all of the small RNAs mapped to this locus were derived from this overlapping region, suggesting that these small RNAs were produced from an intermolecular dsRNA formed between the oppositely oriented transcripts.

A search predicted seven sets of trans-nat-siRNA clusters (Supplementary Table 3), all of which were pairs of a mRNA and its pseudogene.