

Supplementary Figures for “Diversity of miRNAs, siRNAs and piRNAs across 25 Drosophila cell lines”

Jiayu Wen¹, Jaaved Mohammed^{2,3}, Diane Bortolamiol-Becet¹, Harrison Tsai¹, Nicolas Robine^{1,4}, Jakub O. Westholm¹, Erik Ladewig¹, Qi Dai¹, Katsutomo Okamura^{1,5}, Alex S. Flynt¹, Dayu Zhang⁶, Justen Andrews⁶, Lucy Cherbas⁶, Thomas C. Kaufman⁶, Peter Cherbas⁶, Adam Siepel² and Eric C. Lai^{1,7}

¹Department of Developmental Biology, Sloan-Kettering Institute, New York, NY 10065, USA

²Department of Biological Statistics and Computational Biology, Cornell University, Ithaca NY 14853, USA

³Tri-Institutional Training Program in Computational Biology and Medicine

⁴New York Genome Center, 590 Madison Avenue, 21st Floor, New York, NY 10022, USA

⁵Temasek Life Sciences, Temasek Lifesciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604

⁶Dept. of Biology, Indiana University, 1001 East 3rd Street, Bloomington, IN 47405. USA

⁷ Corresponding author

Eric C. Lai

ph: 212-639-5578

fax: 212-717-3623

email: laie@mskcc.org

Index of Supplementary Text, Figures and Tables

Supplementary Text: modENCODE protocol for RNA extraction.

Figure S1: miRNA replicate correlation. Scatter plots of cell line biological replicates are plotted. Most samples show high reproducibility with high correlation coefficients ($r > 0.89$).

Figure S2: Multi-dimensional scaling (MDS) of cell lines vs. tissues. MDS based on sRNA expression clusters cell lines and various tissues (embryo, testes, ovary, and head) into distinct groups. The three ovarian cell lines (fGS/OSS, OSS, OSC) are separated out from other cell lines and have a smaller between-centroid distance to ovaries ($d=23$; indicated in pink arrow) than to testes ($d=30$; purple), embryos ($d=34$; orange), and heads ($d=78$; blue).

Figure S3: Northern analysis of miR-184 expression in cell lines and selected tissues. We analyzed 5 μ g of total RNA in each lane for the designated cell lines and tissues. The blots were hybridized to miR-184, then stripped and reprobed for 2S rRNA as a loading control. The miR-184 levels were normalized to 2S levels, and then normalized to the level in ovaries as 1. The 20nt and 30nt ladder lanes are marked. Consistent with the sequencing results, most cell lines accumulated much higher levels of miR-184 than the three embryonic stages or ovaries did. Note that these high levels of miR-184 appear to exaggerate the presence of isomiR species (i.e. the appearance of longer and shorter miR-184 species). We also note that S1 and G2 cells, which accumulated fewer miR-184 reads than the other cell lines (see main Figure 2C), also exhibited much weaker expression of miR-184 as assessed by these Northern tests.

Figure S4: *mir-309->6* cluster. This cluster was identified as a highly differentially expressed cluster in early embryos (dark blue bar) versus cell lines (light blue bar). Since the three *mir-6* loci generate identical species from their 3p arms, we also plotted the numbers of reads from their 5p arms, which generate unique sequences. Benjamini-Hochberg adjusted p-values for expression differences between early embryos and the aggregate cell lines are shown in parentheses. Early embryos libraries used: GSM286604 (0-1h), GSM286613 (0-1h), GSM180330 (0-1h), GSM180331 (2-6h), GSM286605 (2-6h), GSM286606 (2-6h).

Figure S5: Cell line-specific miRNAs. 23 miRNAs were much more highly expressed in particular cell lines compared with other cell lines (see Methods for the outlier detection).

Figure S6: *mir-972 ->979* cluster from clustering of miRNA expression across 25 cell lines and various tissues. This cluster is highly expressed in S1 and G2 cell lines, as well as in testes.

Figure S7: Additional miRNA clusters detected in cell lines. In addition to the two clusters in Figures 2E and 2F, shown are another four large genomic clusters exhibiting highly correlated expression across the cell lines (see clustering method in Methods).

Figure S8: Association of miRNA and targets expression in 25 cell lines. For each cell line, the distribution across 20 percentiles of miRNA expression (x-axis) versus mean target gene

expression (y-axis) is plotted. Linear regression lines were fitted to the binned data (correlation coefficient of unbinned data shown in brackets). Most cell lines show the expected trend of miRNA-target anti-correlation.

Figure S9: Secondary structure prediction, read distribution, and read-pileup for 18 novel miRNAs predicted by miRDeep2. Description of table fields:

Read size: Length of read.

Hit Count: Number of genomic locations a read mapped to.

AGO1 and AGO2: Number of AGO1-IP, or AGO2-IP respectively, reads from Ovary, Head, and S2-R+ libraries. All AGO-IP libraries are publicly accessible.

Cellline: Total read count summed over all cell-line libraries analyzed in this study.

Public: Total read count summed over all publicly accessible libraries analyzed in this study.

Total: Sum of read count in all “cell-line” and “public” libraries

Total Norm: Total divided by Hit Count

Figure S10: siRNA differential expression between cell lines and tissues. Comparison of TE siRNA expression (A-D) or cis-NAT siRNA expression (E-F) in cell lines vs tissues (heads, testes, ovaries, embryos). Differentially expressed siRNAs (FDR <1%) are shown in blue (high in cell lines) and red (high in tissues). Cell lines generally show elevated TE siRNA expression compared to various tissues. cis-NAT-siRNAs are expressed more highly in testes relative to cell lines.

Figure S11: 3'-cis-NAT-siRNA loci that were expressed in at least 15 cell lines. Shown are expression of those loci in 25 cell lines. AGO2-IP and AGO1-IP are shown in red and yellow, respectively. An additional plot shows mean of the minimum expression of gene pairs (RNA-seq) producing 3'-cis-NAT-siRNAs in at least 15 cell lines versus other gene pairs producing 3'-cis-NAT-siRNAs.

Figure S12: More cis-NAT siRNA examples are listed in several classes.

- (1) “klar-like” cis-NAT siRNAs produced from gene body
- (2) cis-NAT siRNAs produced from 5' divergent pairs
- (3) cell line specific cis-NAT siRNAs
- (4) cis-NAT siRNAs produced from un-annotated 3' overlapping regions
- (5) cis-NAT siRNAs in most cell lines

Figure S13: Additional examples of miRNAs highly expressed in the mixed ovarian germline/somatic fGS/OSS cell line vs ovarian somatic cell lines OSS and OSC.

Figure S14: piRNA master cluster expression. Y-axis is the fraction of unique 24-30nt reads mapped to a master piRNA cluster loci divided by total number of unique 24-30nt reads in the library. (A) piRNA master clusters expressed higher in the mixed germline/somatic (fGS/OSS) than in somatic cell lines (OSS and OSC). (B) piRNA master clusters expressed higher in OSS and OSC, compared to fGS/OSS.

Figure S15: (A) Genomic distribution of identified piRNA loci. (B) Number of identified piRNA loci in three ovarian cell lines. (C) piRNA loci size distribution (inset shows zoomed view of <3000 bp size distribution). (D) fGS/OSS-enriched piRNA loci.

Figure S16: Representative piRNA loci in several classes.

- (1) Examples of piRNAs in unannotated 3'UTRs
- (2) Examples of fGS/OSS enriched piRNAs
- (3) Examples of exonic piRNAs
- (4) Examples of intronic piRNAs
- (5) Examples of piRNAs spanning both CDS and intron
- (6) Examples of antisense piRNAs
- (7) Examples of bidirectional piRNAs

Table S1: Publicly available *Drosophila melanogaster* small RNA datasets used in this study.

Table S2: miRNA list corresponding to Figure 2A and 2B.

Table S3: GO enrichment analysis for targets of highly vs. lowly expressed miRNAs in cell lines.

Table S4: Novel miRNA loci annotated in this study.

Table S5: cis-NAT siRNA loci annotated in this study.

Table S6: piRNA loci annotated in this study.

Table S7: Cell culture conditions

Protocol Text

RNA extraction: TRIzol procedure for insects

Note: This protocol is used to prepare RNA that will be used for the isolation of microRNAs.

Materials

75% ethanol, made with DNase/RNase-free water.

TRIzol® reagent: Invitrogen, cat. no. 15596-026 (100 ml) or 15596-018 (200 ml).

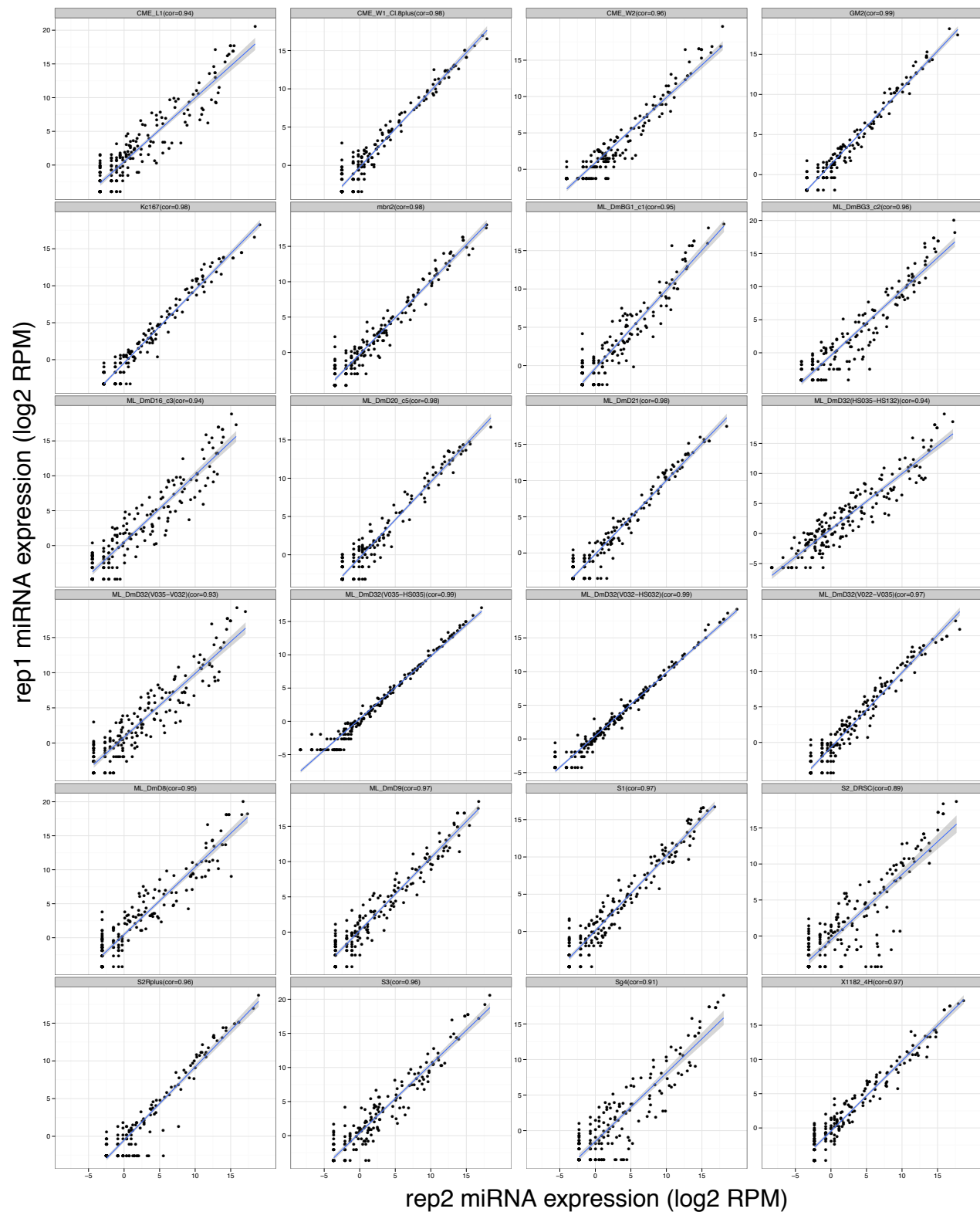
DNase/RNase-free water: Invitrogen cat. no. 10977.

75% ethanol, prepared with DNase/RNase-free water.

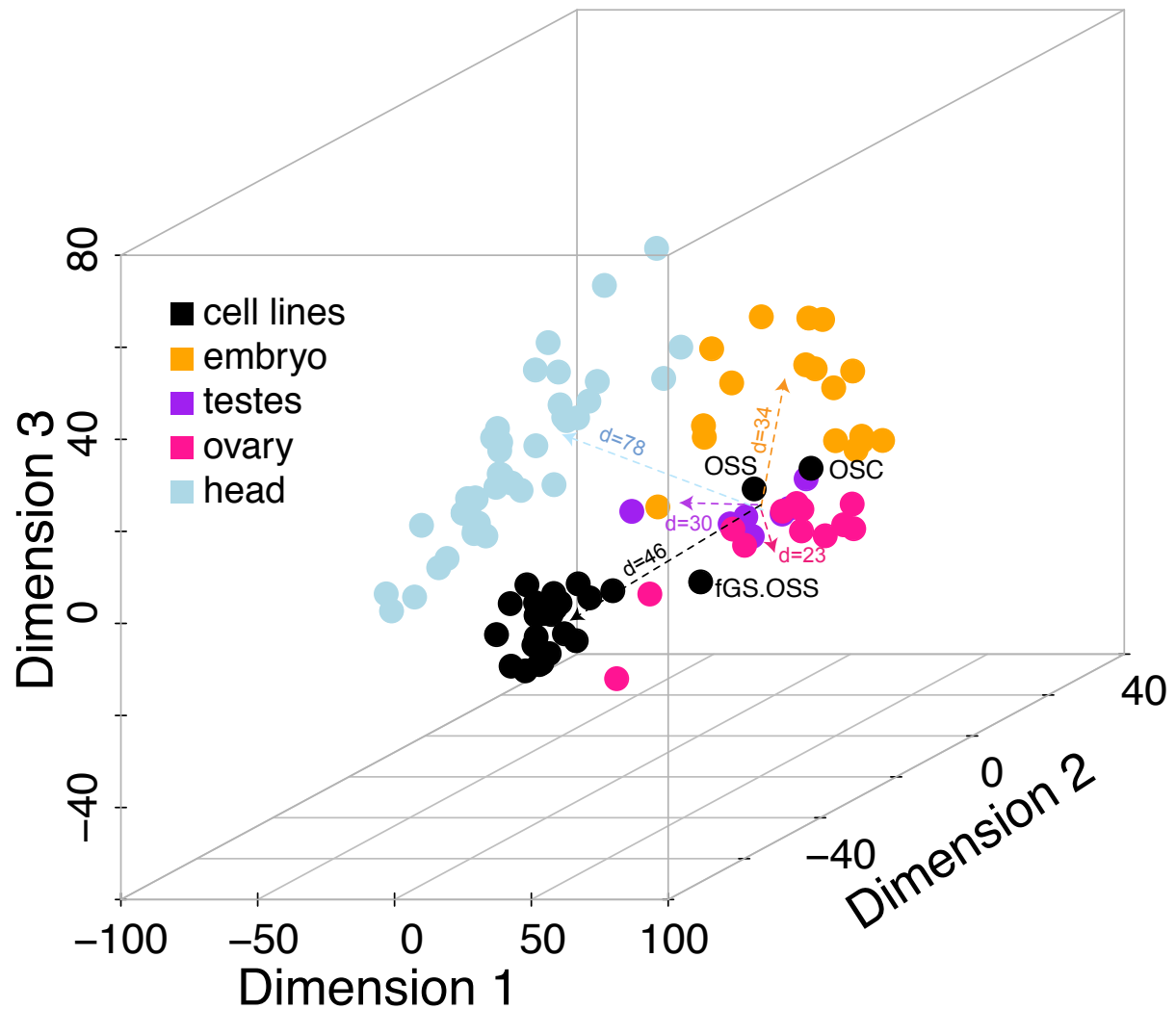
TenBroeck homogenizer: Bellco biotechnology cat no. 1982-10002.

Procedure

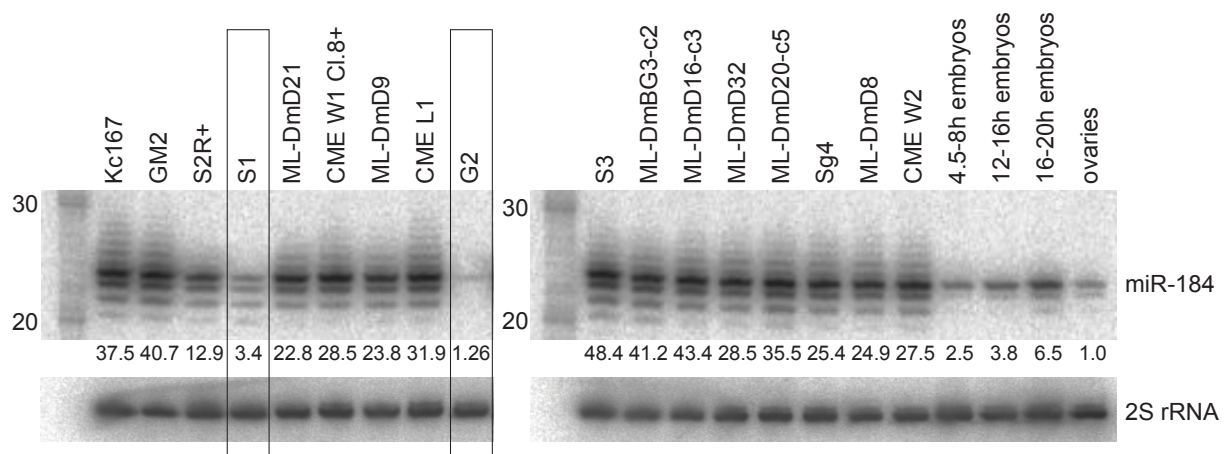
1. Homogenize insect samples in TRIzol reagent using a TenBroeck tissue homogenizer. The sample volume should not exceed 10% of the volume of TRIzol reagent used for homogenization.
2. Incubate at room temperature for 5 minutes.
3. Aliquot samples into 1.5 ml microcentrifuge tubes.
4. Add 0.267 ml chloroform per ml of TRIzol used.
5. Shake tubes vigorously for 15 seconds.
6. Incubate at room temperature for 2 minutes.
7. Centrifuge for 15 minutes at 4°C at 12,000 x g.
8. Transfer the top (aqueous) phases to clean tubes.
9. Precipitate the RNA from the aqueous phase by adding 0.67 ml of isopropanol per ml of TRIzol used in step 5.
10. Invert tubes once to mix gently.
11. Incubate samples at -20° overnight.
12. Centrifuge for 10 minutes at 4° at 12,000 x g.
13. Remove supernatant.
14. Wash RNA pellet once with 75% ethanol, 0.7 ml per microfuge tube.
15. Vortex briefly.
16. Centrifuge at 7,500 x g for 5 minutes at 4°.
17. Let pellet air dry for 10 minutes. Note: DO NOT let the pellet dry completely.
18. Dissolve in RNase-free water.
19. Incubate at 37° overnight to dissolve the RNA. Determine concentration by absorbance, using a Nanodrop® ND-1000 Spectrophotometer. Store at -80°.



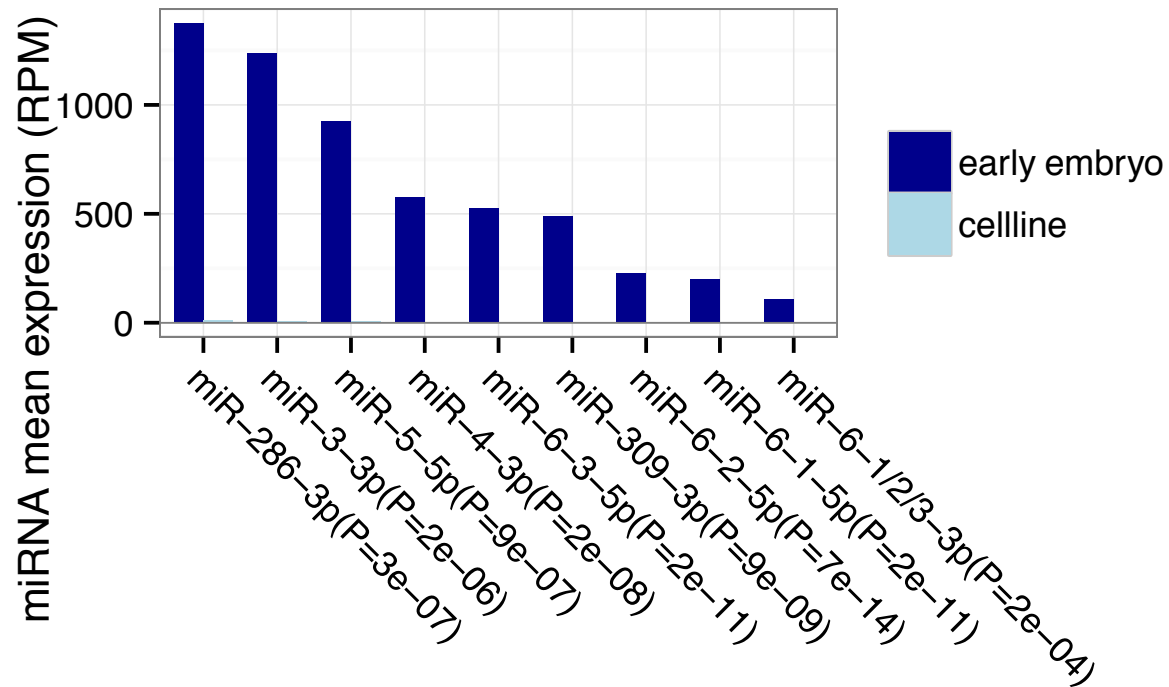
Supplementary Figure 1. miRNA replicate correlation. Scatter plots of cell line biological replicates are plotted. Most samples show high reproducibility with high correlation coefficients ($r > 0.89$).



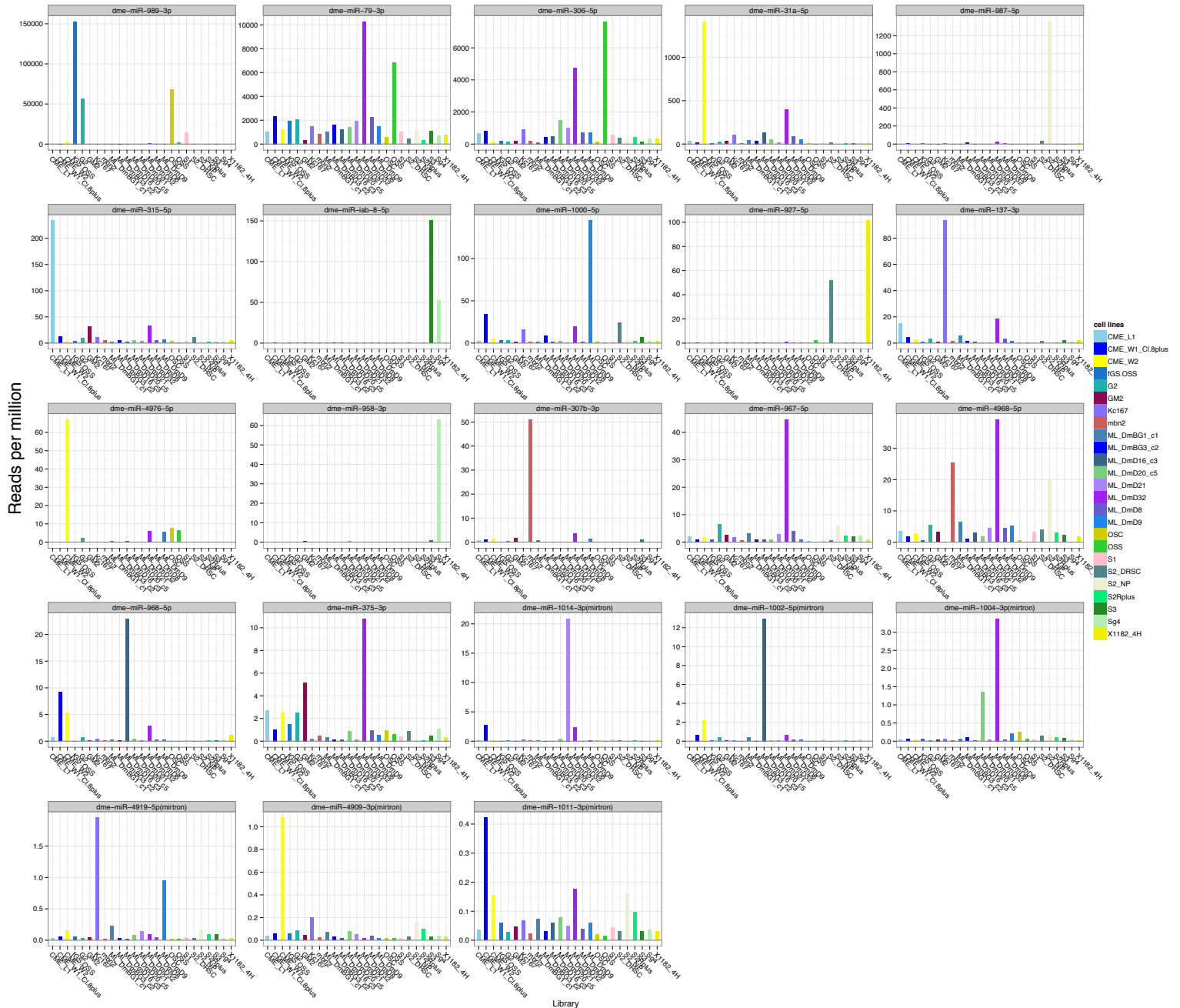
Supplementary Figure 2. Multi-dimensional scaling (MDS) of cell lines vs. tissues. MDS based on sRNA expression clusters cell lines and various tissues (embryo, testes, ovary, and head) into distinct groups. The three ovarian cell lines (fGS/OSS, OSS, OSC) are separated out from other cell lines and have a smaller between-centroid distance to ovaries ($d=23$; indicated in pink arrow) than to testes ($d=30$; purple), embryos ($d=34$; orange), and heads ($d=78$; blue).



Supplementary Figure 3. Northern analysis of miR-184 expression in cell lines and selected tissues. We analyzed 5 g of total RNA in each lane for the designated cell lines and tissues. The blots were hybridized to miR-184, then stripped and reprobbed for 2S rRNA as a loading control. The miR-184 levels were normalized to 2S levels, and then normalized to the level in ovaries as 1. The 20nt and 30nt ladder lanes are marked. Consistent with the sequencing results, most cell lines accumulated much higher levels of miR-184 than the three embryonic stages or ovaries did. Note that these high levels of miR-184 appear to exaggerate the presence of isomiR species (i.e. the appearance of longer and shorter miR-184 species). We also note that S1 and G2 cells, which accumulated fewer miR-184 reads than the other cell lines (see main Figure 2C), also exhibited much weaker expression of miR-184 as assessed by these Northern tests.

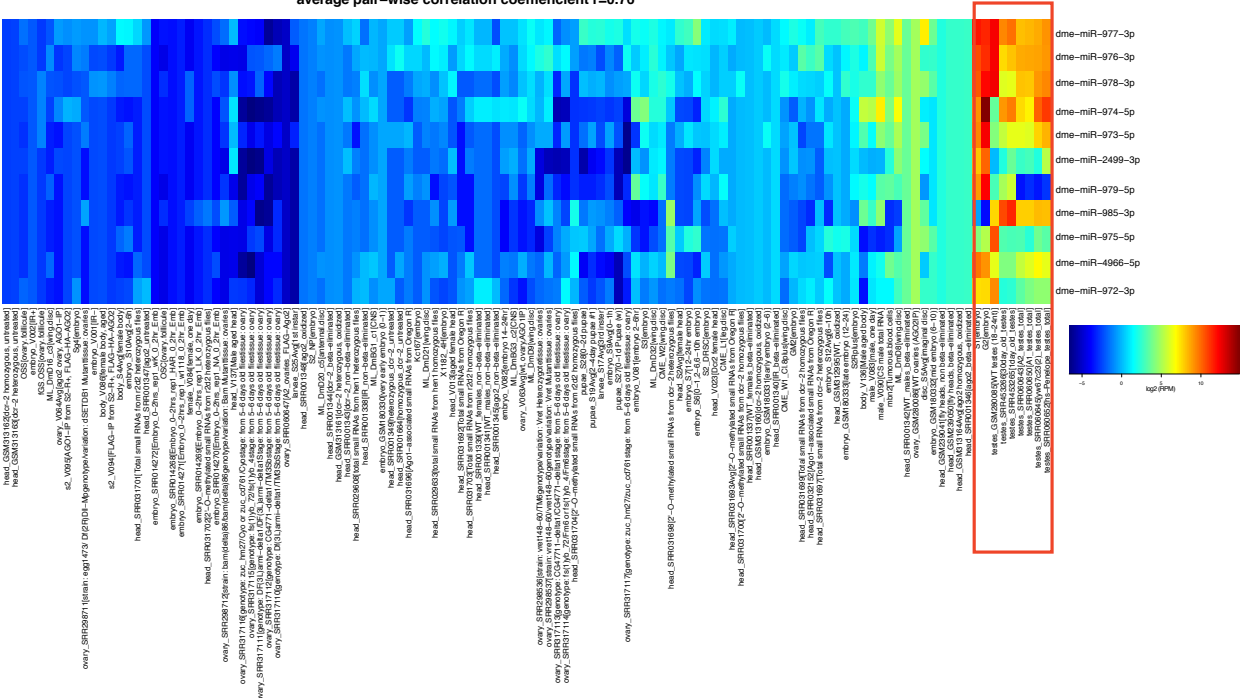


Supplementary Figure 4. mir-309→6 cluster. This cluster was identified as a highly differentially expressed cluster in early embryos (dark blue bar) versus cell lines (light blue bar). Since the three mir-6 loci generate identical species from their 3p arms, we also plotted the numbers of reads from their 5p arms, which generate unique sequences. Benjamini-Hochberg adjusted p-values for expression differences between early embryos and the aggregate cell lines are shown in parentheses. Early embryos libraries used: GSM286604 (0-1h), GSM286613 (0-1h), GSM180330 (0-1h), GSM180331 (2-6h), GSM286605 (2-6h), GSM286606 (2-6h).

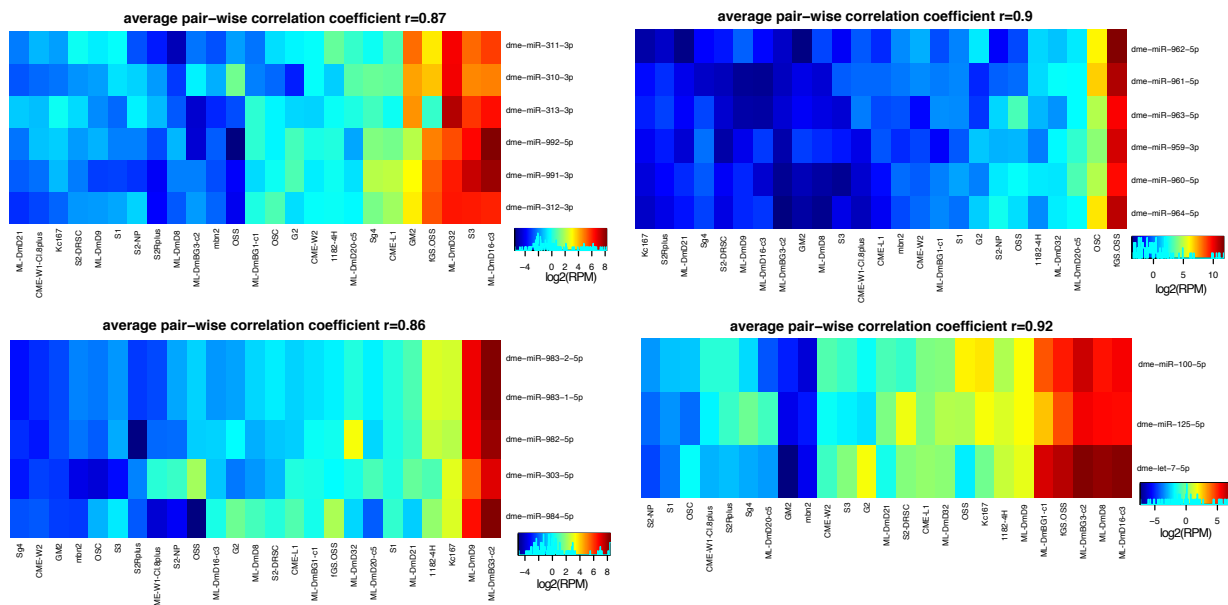


Supplementary Figure 5. Cell line specific miRNAs. 23 miRNAs were identified to be relatively highly expressed in particular cell lines compared with other cell lines (see Methods for the outlier detection).

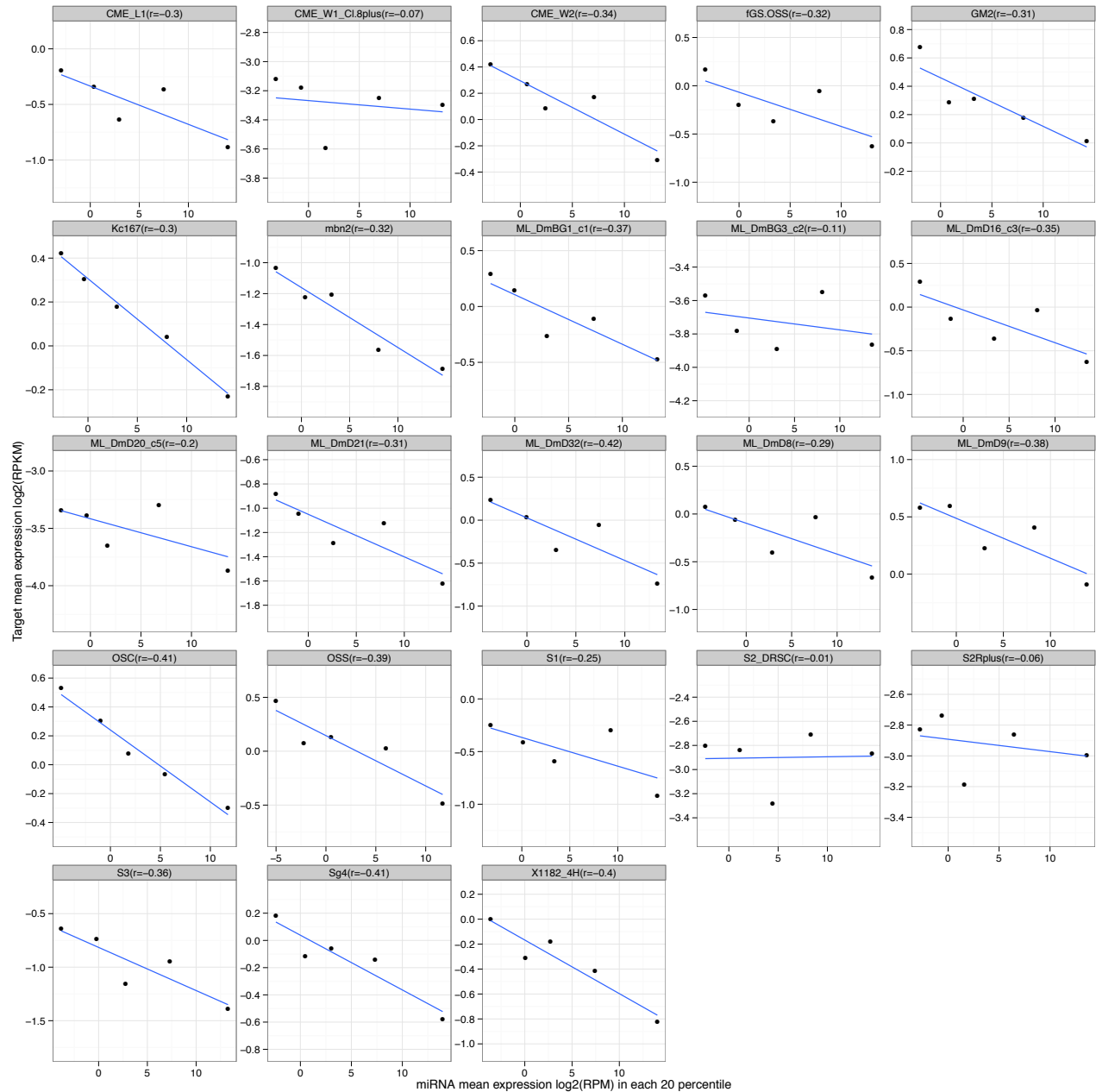
average pair-wise correlation coefficient $r=0.76$



Supplementary Figure 6. mir-972→979 cluster from clustering of miRNA expression across 25 cell lines and various tissues. This cluster is highly expressed in S1 and G2 cell lines, as well as in testes.



Supplementary Figure 7. Additional miRNA clusters detected in cell lines. In addition to the two clusters in Figures 2E and 2F, shown are another four large genomic clusters exhibiting highly correlated expression across the cell lines (see clustering method in Methods).



Supplementary Figure 8. Association of miRNA and targets expression in 25 cell lines. For each cell line, the distribution across 20 percentiles of miRNA expression (x-axis) versus mean target gene expression (y-axis) is plotted. Linear regression lines were fitted to the binned data (correlation coefficient of unbinned data shown in brackets). Most cell lines show the expected trend of miRNA-target anti-correlation.

Supplementary Figure S9: Secondary structure prediction, read distribution and read pileup for 18 novel miRNAs predicted by miRDeep2

Legend

Read size: Length of read.

Hit Count: Number of genomic locations a read mapped to.

AGO1: Number of AGO1-IP reads or AGO2-IP reads, respectively, from Ovary, Head, and S2-R+ small RNA libraries. All AGO1-IP libraries are publicly accessible.

Cellline: Total read count summed over all cell-line libraries analyzed in this study.

Public: Total read count summed over all publicly accessible libraries analyzed in this study.

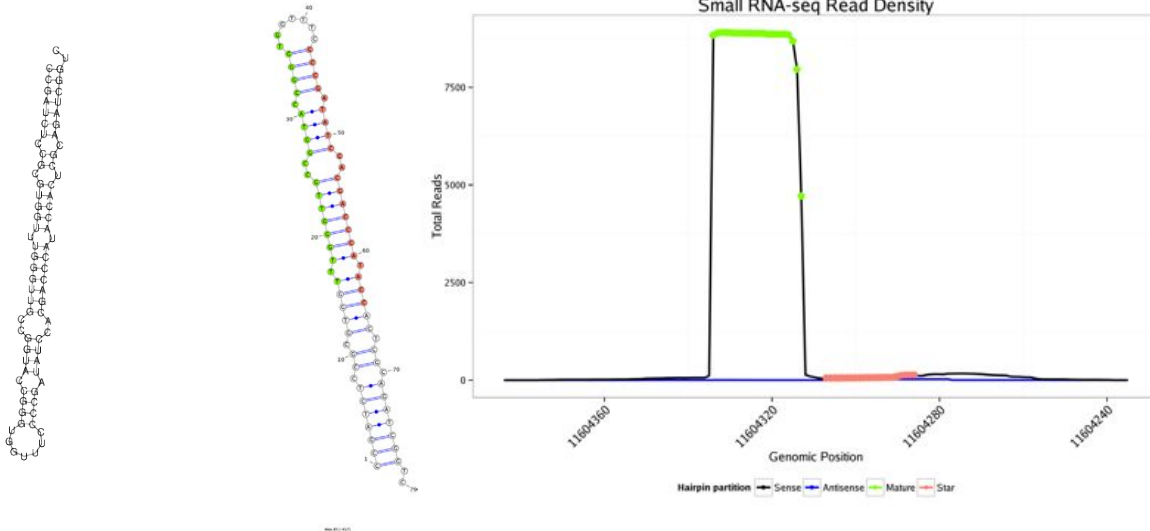
Total: Sum of read count in all “cell-line” and “public” libraries

Total Norm: Total divided by Hit Count

Table Truncated: Reads with low read counts were removed for better visualization. See HTML documents for complete table.

Legend: mature star

Predicted structure



Sense Strand Reads

Read size	# Mismatch	Hit Count	Total Norm	Total	AGO1	AGO2	cellline	public
22	0	1	4553.00	4553	230.00	52.00	3352.00	1201.00
21	0	1	3227.00	3227	237.00	31.00	2448.00	779.00
20	0	1	674.00	674	31.00	5.00	558.00	116.00
19	0	1	159.00	159	2.00	2.00	139.00	20.00
21	0	1	44.00	44	0.00	1.00	35.00	9.00
22	0	1	39.00	39	2.00	0.00	5.00	34.00
27	0	1	29.00	29	0.00	0.00	19.00	10.00
23	0	1	29.00	29	2.00	0.00	19.00	10.00
21	0	1	21.00	21	4.00	0.00	7.00	14.00
25	0	1	17.00	17	0.00	0.00	14.00	3.00
24	0	1	15.00	15	0.00	0.00	3.00	12.00
24	0	1	15.00	15	1.00	0.00	10.00	5.00
22	0	1	13.00	13	0.00	0.00	1.00	12.00
25	0	1	12.00	12	0.00	0.00	0.00	12.00
20	0	1	12.00	12	1.00	0.00	8.00	4.00
26	0	1	12.00	12	0.00	0.00	0.00	12.00
27	0	1	11.00	11	0.00	0.00	8.00	3.00
21	0	1	10.00	10	0.00	0.00	6.00	4.00
22	0	1	10.00	10	1.00	0.00	1.00	9.00
26	0	1	10.00	10	1.00	0.00	4.00	6.00
18	0	1	10.00	10	1.00	0.00	8.00	2.00
21	0	1	9.00	9	0.00	0.00	7.00	2.00
22	0	1	9.00	9	0.00	0.00	3.00	6.00
21	0	1	8.00	8	0.00	0.00	0.00	8.00
26	0	1	8.00	8	0.00	0.00	2.00	6.00
23	0	1	7.00	7	4.00	0.00	0.00	7.00
22	0	1	6.00	6	0.00	0.00	0.00	6.00
28	0	1	6.00	6	0.00	0.00	0.00	6.00
22	0	1	6.00	6	0.00	0.00	3.00	3.00
25	0	1	5.00	5	0.00	0.00	3.00	2.00
24	0	1	5.00	5	0.00	0.00	3.00	2.00
20	0	1	5.00	5	0.00	0.00	5.00	0.00
27	0	1	5.00	5	0.00	0.00	2.00	3.00
29	0	1	4.00	4	0.00	0.00	4.00	0.00
25	0	1	4.00	4	0.00	0.00	0.00	4.00
23	0	1	4.00	4	0.00	0.00	0.00	4.00
26	0	1	4.00	4	0.00	0.00	2.00	2.00
23	0	1	4.00	4	0.00	0.00	1.00	3.00
25	0	1	4.00	4	0.00	0.00	4.00	0.00
20	0	1	4.00	4	0.00	0.00	3.00	1.00
29	0	1	4.00	4	0.00	0.00	4.00	0.00
26	0	1	4.00	4	0.00	0.00	2.00	2.00
25	0	1	4.00	4	0.00	0.00	1.00	3.00
27	0	1	4.00	4	0.00	0.00	4.00	0.00

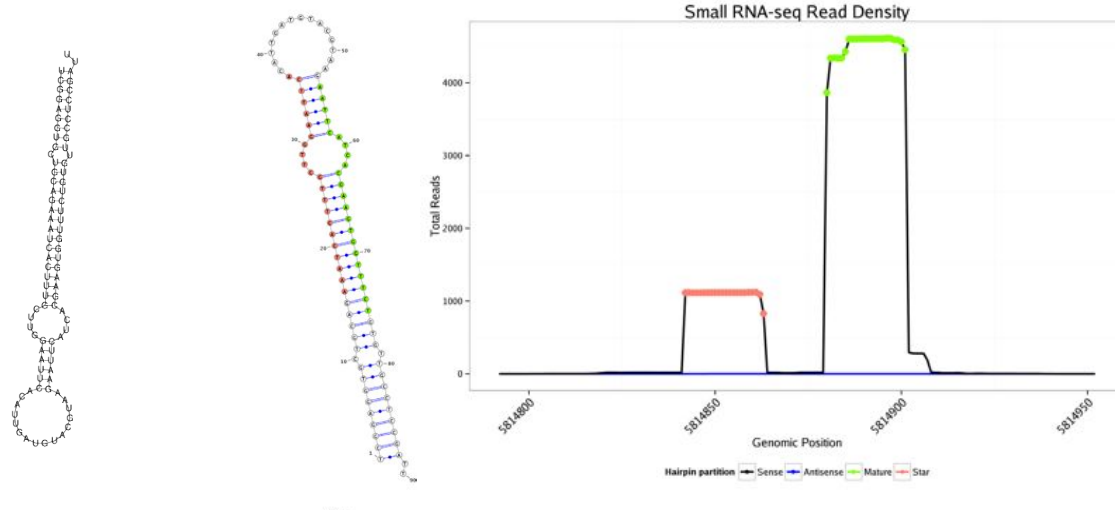
...Table Truncated...

Anti-sense strand reads

Read	# Mismatch	Hit Count	Total Norm	Total	AG1	AG2	cellline	public	
GGCCCGCATCTTAATTGCACCAGAGCGAGTCGTCGGATCTCCGCGTGGTTGGGTTCGCGGATACGGGTGTTTC	22	0	1	21.00	21	0.00	0.00	20.00	1.00
*****(((-----(((									

Legend: mature star

Predicted structure



Sense Strand Reads

[illegible]

Anti-sense strand reads

Read size	# Mismatch	Hit Count	Total Norm	Total	AGO1	AGO2	celline	public
23	0	1	1.00	1	0.00	0.00	1.00	0.00

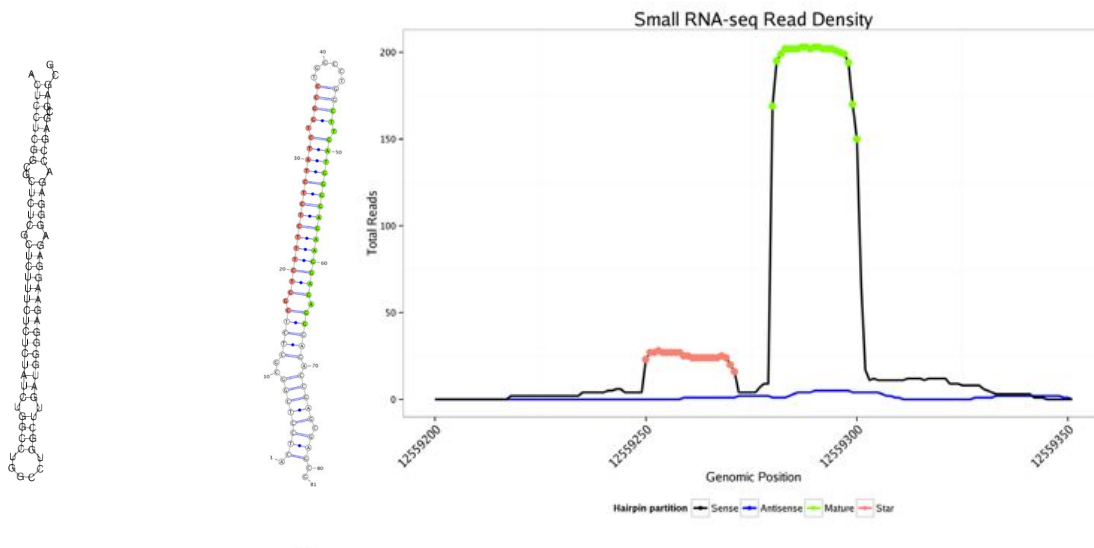
Legend: mature star



[illegible]

Legend: mature star

Predicted structure



Sense Strand Reads

[illegible]

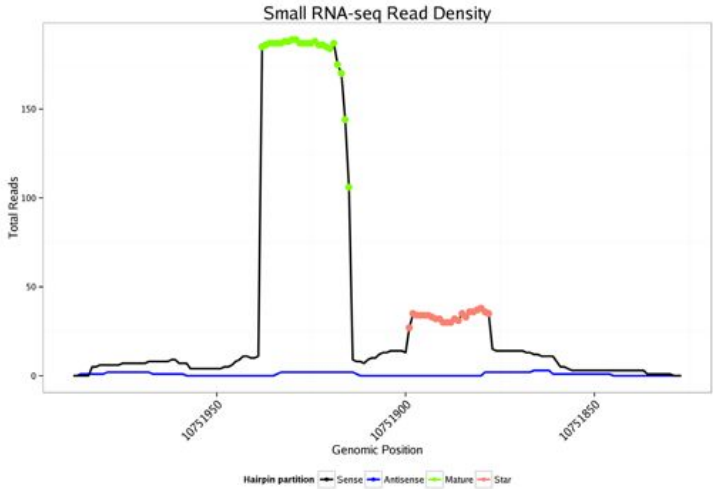
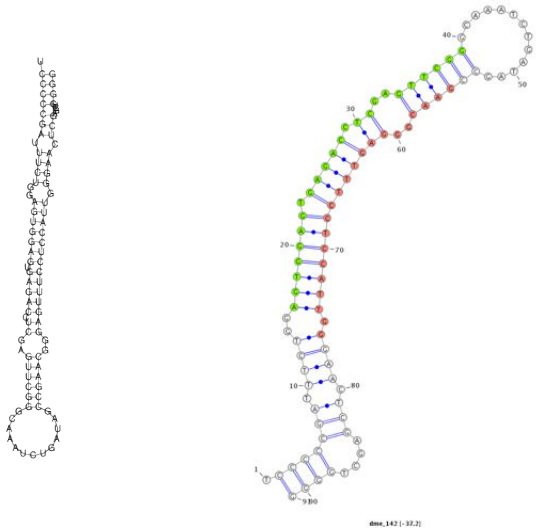
Anti-sense strand reads

[illegible]

ID:	Coordinate:	Confidence:	
dme-mir-9378	chr2L:10751877-10751938 -	Novel miRNA	[View on UCSC Genome Browser]

Legend: mature star

Predicted structure



Sense Strand Reads

[illegible]

Anti-sense strand reads

[illegible]

ID:

dme-mir-9380

Coordinate:

chr3R:4174424-4174500 -

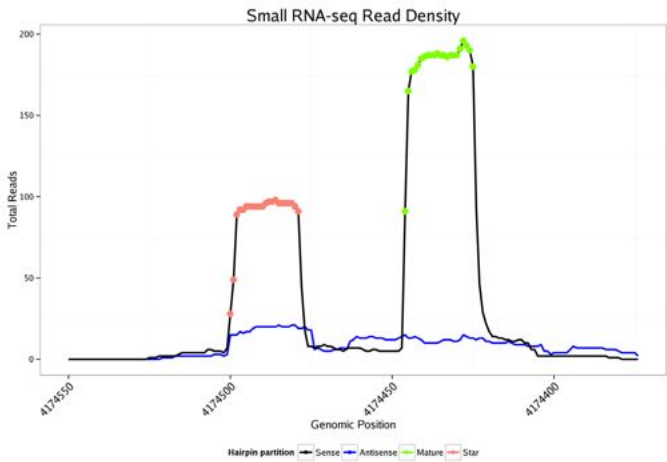
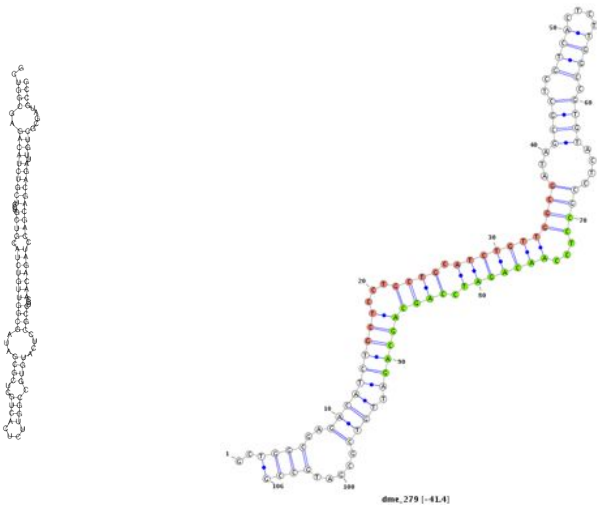
Confidence:

Novel miRNA

[\[View on UCSC Genome Browser\]](#)

Legend: ■ mature ■ star

Predicted structure



Sense Strand Reads

CTCCTGTCGATAGCGATCCAACTCCCTGTCTGGTGCCTGGCGAGACATCT GCCTCTGCTGCATCTGTTGGCG ATAGCGCTGTCACCTCTTGGCCGTGTACTCCG CTCCCAACAGATCCAGCAGCAG ATTGTCGCGATGCCGTTTTTTGTTCTGCAATAAGAGTTAGCTTTATTGATT	Read size	# Mismatch	Hit Count	Total Norm	Total	AG01	AG02	cellline	public
.....CTCCACACAGATCCAGCAGCAG.....	22	0	1	48.00	48	6.00	0.00	39.00	9.00
.....CTCCACACAGATCCAGCAGCAG.....	21	0	1	34.00	34	8.00	0.00	25.00	9.00
.....CTCCACACAGATCCAGCAGCAG.....	23	0	1	26.00	26	1.00	0.00	20.00	6.00
.....CTCCACACAGATCCAGCAGCAG.....	22	0	1	21.00	21	3.00	0.00	14.00	7.00
.....GCTCCTGCTGCATCTGTTGGCG.....	22	0	1	16.00	16	0.00	0.00	13.00	3.00
.....TCCTGCTGCATCTGTTGGCG.....	20	0	1	15.00	15	1.00	0.00	12.00	3.00
.....CTCCTGCTGCATCTGTTGGCG.....	21	0	1	14.00	14	0.00	0.00	13.00	1.00
.....TCCTGCTGCATCTGTTGGCG.....	21	0	1	13.00	13	0.00	0.00	13.00	0.00
.....TCCTGCTGCATCTGTTGGCG.....	22	0	1	9.00	9	0.00	0.00	8.00	1.00
.....CTCCACACAGATCCAGCAGCAG.....	23	0	1	9.00	9	0.00	0.00	9.00	0.00
.....CTCCACACAGATCCAGCAGCAG.....	21	0	1	8.00	8	0.00	0.00	7.00	1.00
.....CTCCTGCTGCATCTGTTGGCG.....	22	0	1	7.00	7	0.00	0.00	3.00	4.00
.....TCCAACAGATCCAGCAGCAGAT.....	22	0	1	7.00	7	0.00	1.00	6.00	1.00
.....GCAGATTGTGCGATGCCGTT.....	21	0	1	5.00	5	0.00	0.00	1.00	4.00
.....CTCCACACAGATCCAGCAGCAGATT.....	24	0	1	5.00	5	0.00	0.00	5.00	0.00
.....TCCAACAGATCCAGCAGCAG.....	20	0	1	4.00	4	0.00	0.00	4.00	0.00
.....AGCAGATTGTGCGATGCCGTT.....	24	0	1	4.00	4	0.00	0.00	4.00	0.00
.....GCTCCTGCTGCATCTGTTGGCG.....	23	0	1	4.00	4	0.00	0.00	4.00	0.00
.....TCCTGCTGCATCTGTTGGCG.....	19	0	1	3.00	3	0.00	0.00	3.00	0.00
.....TGCTCCTGCTGCATCTGTTGG.....	21	0	1	3.00	3	0.00	0.00	2.00	1.00
.....CTCCACACAGATCCAGCAGC.....	20	0	1	3.00	3	0.00	0.00	3.00	0.00
.....CTCCACACAGATCCAGCAG.....	18	0	1	2.00	2	0.00	0.00	2.00	0.00
.....CTGCGCGTGTACTCCG.....	18	0	1	2.00	2	0.00	0.00	2.00	0.00
.....CTCCACACAGATCCAGCAGC.....	20	0	1	2.00	2	0.00	0.00	2.00	0.00
.....AGCAGCAGATTGTGCGATGCCG.....	23	0	1	2.00	2	0.00	0.00	2.00	0.00
.....CAACAGATCCAGCAGCAGATT.....	21	0	1	2.00	2	0.00	0.00	2.00	0.00
.....TGCTGCTGCATCTGTTGGCG.....	18	0	1	2.00	2	0.00	0.00	2.00	0.00
.....GACATCTGCTCCTGCTGATCT.....	22	0	1	2.00	2	0.00	0.00	2.00	0.00
.....AACAGATCCAGCAGCAGATTG.....	21	0	1	2.00	2	0.00	0.00	2.00	0.00
.....AACAGATCCAGCAGCAGATTG.....	22	0	1	2.00	2	0.00	0.00	2.00	0.00

...Table Truncated...

Anti-sense strand reads

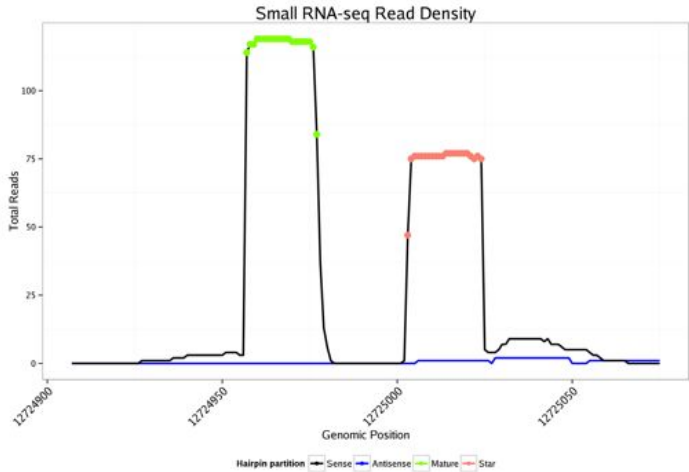
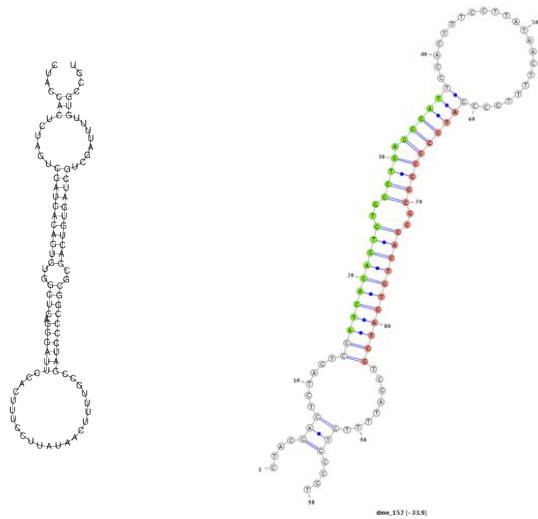
CTCCTGTCGATAGCGATCCAACTCCCTGTCTGGTGCCTGGCGAGACATCT GCCTCTGCTGCATCTGTTGGCG ATAGCGCTGTCACCTCTTGGCCGTGTACTCCG CTCCCAACAGATCCAGCAGCAG ATTGTCGCGATGCCGTTTTTTGTTCTGCAATAAGAGTTAGCTTTATTGATT	Read size	# Mismatch	Hit Count	Total Norm	Total	AG01	AG02	cellline	public
.....COGCTCCACAGATCCAGCAGCAG.....	26	0	1	12.00	12	0.00	0.00	0.00	12.00
.....GCGATAGCGCTGTCACCTCT.....	21	0	1	4.00	4	0.00	0.00	0.00	4.00
.....GCTGCGCTGGCGAGACATCTCC.....	25	0	1	3.00	3	0.00	0.00	0.00	3.00
.....CTGCTCCTGCTGCATCTGTTG.....	21	0	1	3.00	3	0.00	1.00	2.00	1.00
.....CTCGCGCTGGCGAGACATCT.....	21	0	1	2.00	2	0.00	0.00	0.00	2.00
.....TGCGGATAGCGCTGTCACCTC.....	21	0	1	2.00	2	0.00	0.00	1.00	1.00
.....TCCTGTCGATAGCGATCCAACT.....	21	0	1	2.00	2	0.00	0.00	0.00	2.00
.....CTCCTGTCGATAGCGATCCAACT.....	21	0	1	2.00	2	0.00	0.00	0.00	2.00

...Table Truncated...

ID:	Coordinate:	Confidence:	
dme-mir-9382	chr2L:12724957-12725025 +	Novel miRNA	[View on UCSC Genome Browser]

Legend: mature star

Predicted structure



Sense Strand Reads

GCAGTAATAGTACGCCGATTGCATATACCAATTACTACACACTAGTGC	ATCACAGTGTGGCTGAGGGGATT	TCACCTTTCCTTATACACTTTTGC	ATCCCCGGCGCCCACTGTGATCG	TGATTGTGCGCGTTTGTGATTGTGATTATGTTTATTTACTTAAATTG	Read size	# Mismatch	Hit Count	Total Norm	Total	AG01	AG02	cellline	public
GCAGTAATAGTACGCCGATTGCATATACCAATTACTACACACTAGTGC	ATCACAGTGTGGCTGAGGGGATT	TCACCTTTCCTTATACACTTTTGC	ATCCCCGGCGCCCACTGTGATCG	TGATTGTGCGCGTTTGTGATTGTGATTATGTTTATTTACTTAAATTG	21	0	1	47.00	47	0.00	0.00	44.00	3.00
ATCACAGTGTGGCTGAGGGGATT	ATCCCCGGCGCCCACTGTGATCG				22	0	1	42.00	42	13.00	0.00	23.00	19.00
ATCACAGTGTGGCTGAGGGGATT					20	0	1	32.00	32	0.00	0.00	29.00	3.00
ATCACAGTGTGGCTGAGGGGATT	TCGCCGGCGCCCACTGTGATCG				21	0	1	28.00	28	2.00	0.00	23.00	5.00
ATCACAGTGTGGCTGAGGGGATT	ATCACAGTGTGGCTGAGGGGATT				22	0	1	21.00	21	0.00	0.00	20.00	1.00
ATCACAGTGTGGCTGAGGGGATT					23	0	1	7.00	7	0.00	0.00	7.00	0.00
ACAGTGTGGCTGAGGGGATT					21	0	1	3.00	3	0.00	0.00	3.00	0.00
TCACAGTGTGGCTGAGGGGATT					21	0	1	3.00	3	0.00	0.00	3.00	0.00
ATCACAGTGTGGCTGAGGGGATT					19	0	1	2.00	2	0.00	0.00	2.00	0.00
ATCACAGTGTGGCTGAGGGGATT					24	0	1	2.00	2	0.00	0.00	2.00	0.00
GCATGATTGTGCGCGTTT					21	0	1	2.00	2	0.00	0.00	2.00	0.00
GCATGATTGTGCGCGTTT					28	0	1	1.00	1	0.00	0.00	0.00	1.00
TCGCGCGTTTGTGATTGTGATTAT					26	0	1	1.00	1	0.00	0.00	1.00	0.00
TCGCGCGTTTGTGATTGTGATTAT					26	0	1	1.00	1	0.00	0.00	1.00	0.00
TCGCGCGTTTGTGATTGTGATTAT					27	0	1	1.00	1	0.00	0.00	0.00	1.00
TCGCGCGTTTGTGATTGTGATTAT					23	0	1	1.00	1	0.00	0.00	0.00	1.00
TCGCGCGTTTGTGATTGTGATTAT					18	0	1	1.00	1	0.00	0.00	1.00	0.00
CAATTACTACCACTAGTGCAT					24	0	1	1.00	1	0.00	0.00	1.00	0.00
TACTACCACTAGTGCAT					20	0	1	1.00	1	0.00	0.00	1.00	0.00
ACAGTGTGGCTGAGGGATT					28	0	1	1.00	1	0.00	0.00	1.00	0.00
TAGTCGATCAGAGTGGC					22	0	1	1.00	1	0.00	0.00	1.00	0.00
TAGTCGATCAGAGTGGC					19	0	1	1.00	1	0.00	0.00	0.00	1.00
ATCCCCGGCGCCCACTGTGATCG					25	0	1	1.00	1	0.00	0.00	0.00	1.00
ATCCCCGGCGCCCACTGTGATCG					19	0	1	1.00	1	0.00	0.00	0.00	1.00
ATCCCCGGCGCCCACTGTGATCG					11	0	1	1.00	1	0.00	0.00	0.00	1.00
ATCCCCGGCGCCCACTGTGATCG					23	0	1	1.00	1	0.00	0.00	1.00	0.00
ATCCCCGGCGCCCACTGTGATCG					22	0	1	1.00	1	0.00	0.00	0.00	1.00
ATCCCCGGCGCCCACTGTGATCG					21	0	1	1.00	1	0.00	0.00	0.00	1.00
ATCCCCGGCGCCCACTGTGATCG					20	0	1	1.00	1	0.00	0.00	1.00	0.00
GATCCCCGGCGCCCACTGTGA					20	0	1	1.00	1	0.00	0.00	1.00	0.00

Anti-sense strand reads

[illegible]

ID:

dme-mir-9383

Coordinate:

chr3R:11047517-11047577 +

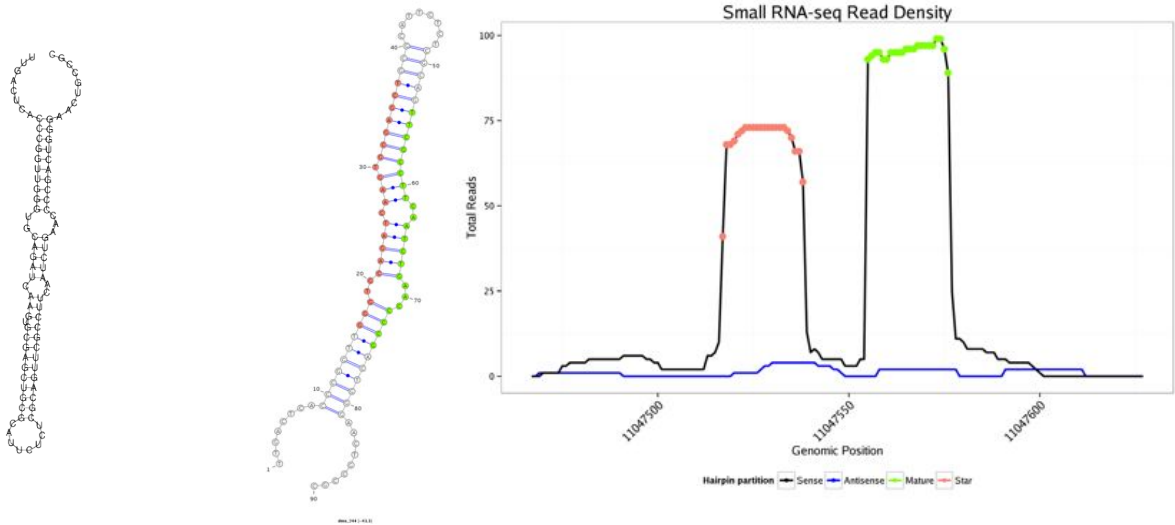
Confidence:

Novel miRNA

[\[View on UCSC Genome Browser\]](#)

Legend: ■ mature ■ star

Predicted structure



Sense Strand Reads

Read size	# Mismatch	Hit Count	Total Norm	Total	AG01	AG02	cellline	public
22	0	1	64.00	64	3.00	0.00	35.00	29.00
22	0	1	27.00	27	0.00	0.00	18.00	9.00
21	0	1	16.00	16	0.00	0.00	14.00	2.00
23	0	1	15.00	15	0.00	0.00	7.00	8.00
21	0	1	6.00	6	0.00	0.00	0.00	6.00
20	0	1	5.00	5	0.00	0.00	5.00	0.00
22	0	1	4.00	4	0.00	0.00	4.00	0.00
21	0	1	4.00	4	0.00	1.00	3.00	1.00
20	0	1	3.00	3	3.00	0.00	0.00	3.00
26	0	1	2.00	2	0.00	0.00	0.00	2.00
19	0	1	2.00	2	0.00	0.00	0.00	2.00
28	0	1	2.00	2	0.00	0.00	2.00	0.00
18	0	1	2.00	2	0.00	0.00	2.00	0.00
23	0	1	2.00	2	0.00	0.00	0.00	2.00
18	0	1	2.00	2	0.00	0.00	0.00	2.00
24	0	1	1.00	1	0.00	0.00	0.00	1.00
21	0	1	1.00	1	0.00	0.00	0.00	1.00
24	0	1	1.00	1	0.00	0.00	0.00	1.00
28	0	1	1.00	1	0.00	0.00	0.00	1.00
27	0	1	1.00	1	0.00	0.00	1.00	0.00
20	0	1	1.00	1	0.00	0.00	0.00	1.00
19	0	1	1.00	1	0.00	0.00	0.00	1.00
20	0	1	1.00	1	0.00	0.00	0.00	1.00
23	0	1	1.00	1	0.00	0.00	0.00	1.00
21	0	1	1.00	1	0.00	0.00	0.00	1.00
21	0	1	1.00	1	0.00	0.00	0.00	1.00
25	0	1	1.00	1	0.00	0.00	0.00	1.00
19	0	1	1.00	1	0.00	0.00	0.00	1.00
20	0	1	1.00	1	0.00	0.00	0.00	1.00
22	0	1	1.00	1	0.00	0.00	0.00	1.00
19	0	1	1.00	1	0.00	0.00	0.00	1.00
21	0	1	1.00	1	0.00	0.00	1.00	0.00
23	0	1	1.00	1	0.00	0.00	0.00	1.00
24	0	1	1.00	1	0.00	0.00	0.00	1.00
26	0	1	1.00	1	0.00	0.00	0.00	1.00
28	0	1	1.00	1	0.00	0.00	0.00	1.00
23	0	1	1.00	1	0.00	0.00	1.00	0.00
27	0	1	1.00	1	0.00	0.00	0.00	1.00
27	0	1	1.00	1	0.00	0.00	0.00	1.00
27	0	1	1.00	1	0.00	0.00	0.00	1.00
22	0	1	1.00	1	0.00	0.00	0.00	1.00
25	0	1	1.00	1	0.00	0.00	0.00	1.00

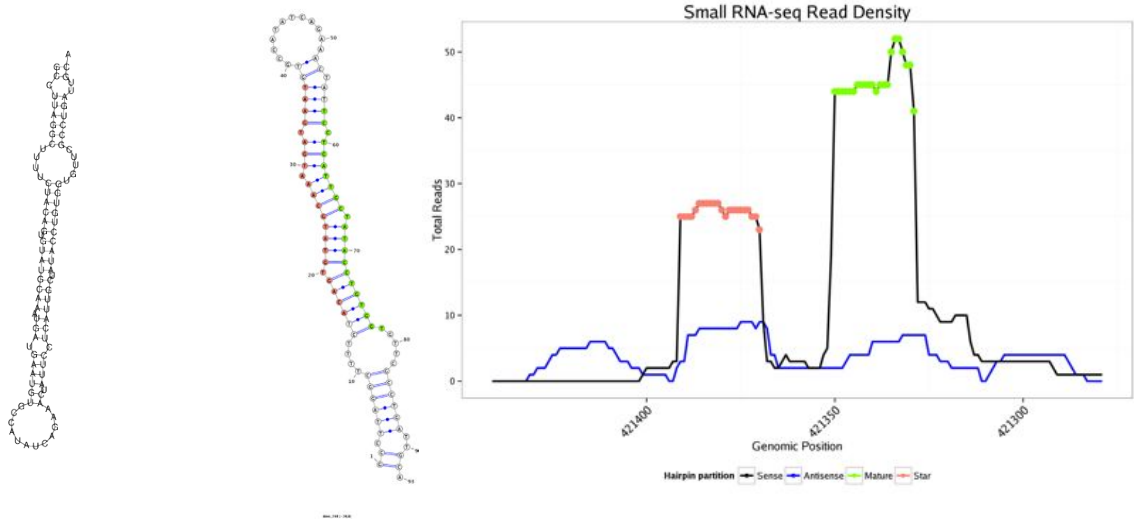
Anti-sense strand reads

Read size	# Mismatch	Hit Count	Total Norm	Total	AG01	AG02	cellline	public
21	0	1	2.00	2	0.00	0.00	2.00	0.00
21	0	1	2.00	2	0.00	0.00	0.00	2.00
22	0	1	1.00	1	0.00	0.00	1.00	0.00
22	0	1	1.00	1	0.00	0.00	0.00	1.00
19	0	1	1.00	1	0.00	0.00	1.00	0.00
21	0	1	1.00	1	0.00	0.00	0.00	1.00
18	0	1	1.00	1	0.00	0.00	0.00	1.00

ID:	Coordinate:	Confidence:	
dme-mir-9385	chr4:421329-421391 -	Novel miRNA	[View on UCSC Genome Browser]

Legend: mature star

Predicted structure



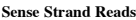
Sense Strand Reads

[illegible]

Anti-sense strand reads

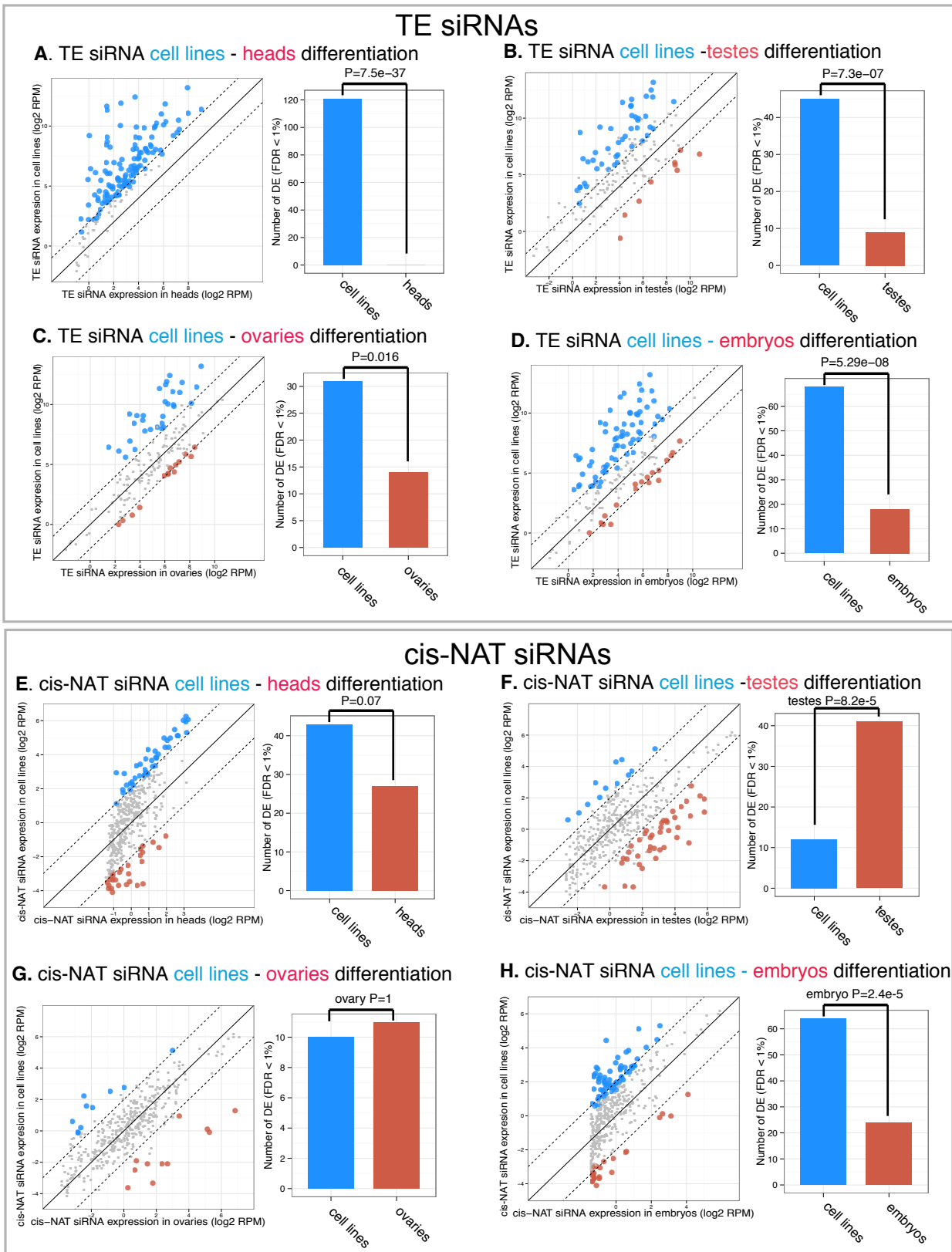
Read size	# Mismatch	Hit Count	Total Norm	Total	AG01	AG02	cellline	public
ACCGACAAGGCGCCTTTAAATTCAAATATTAAGGGCCTTAGGCTTTCTACAGTGTATGCAATATGCAAAATGATGAATGTCGCATATCAGAAATATGCTCATGCTATACCTGTGGTGTTCGCTGATTGCATCTGCCCTATACCATTATGGCTGGGTGATGATTA	22	1	4.00	4	0.00	0.00	0.00	4.00
*****(((((-----(((((((-----(((-----))*****	21	0	1.00	1	0.00	1.00	0.00	1.00
.....ATTCTCTATTGCTATACCTGTG.....	21	0	1.00	1	0.00	1.00	0.00	1.00
.....CTATCTCCCATGCTATACCT.....	21	0	1.00	1	0.00	1.00	0.00	1.00
.....GGCGCTTTAAATTCAAATTT.....	21	0	1.00	1	0.00	0.00	0.00	1.00
.....GGCGCCTTTAAATTCAAAT.....	22	0	1.00	1	0.00	0.00	1.00	0.00
.....CAAGGCGCCTTTAAATTCAAAT.....	18	0	1.00	1	0.00	0.00	0.00	1.00
.....TCTACAGTGTATGCAAAATGATG.....	22	0	1.00	1	0.00	1.00	0.00	1.00
.....CTATTCTCTATTGCTATACCTGTGGT.....	27	0	1.00	1	0.00	0.00	0.00	1.00
.....TTGCACTCGCCCTATACCACT.....	22	0	1.00	1	0.00	0.00	1.00	0.00
.....TGCCCTATACCACTATGGCT.....	21	0	1.00	1	0.00	0.00	0.00	1.00
.....TCTACAGTGTATGCAAAATGCA.....	24	0	1.00	1	0.00	0.00	0.00	1.00
.....TCTACAGTGTATGCAAAATGAT.....	21	0	1.00	1	0.00	0.00	0.00	1.00
.....TTTCTACAGTGTATGCAAAATGATGA.....	21	0	1.00	1	0.00	0.00	0.00	1.00
.....CATCTGCCCTATACCACTATG.....	27	0	1.00	1	0.00	0.00	0.00	1.00
.....GGCTTTCTACAGTGTATGCA.....	22	0	1.00	1	0.00	0.00	0.00	1.00
.....CGCCTTTAAATTCAAATATTA.....	21	0	1.00	1	0.00	1.00	0.00	1.00
.....TTGCCATTGCTATACCTGTGGT.....	21	0	1.00	1	0.00	0.00	0.00	1.00
.....TGTCCTATCAGAAATCTTCCTA.....	23	0	1.00	1	1.00	0.00	0.00	1.00
.....GGGCTTAGGCTTTCTACAGTATGCA.....	26	0	1.00	1	0.00	0.00	0.00	1.00
.....CTCATGCTATACCTGTGGT.....	29	0	1.00	1	0.00	0.00	0.00	1.00
.....TGTCCTATCAGAAATCTATT.....	21	0	1.00	1	0.00	0.00	1.00	0.00
.....GGGCTTAGGCTTTCTACAG.....	21	0	1.00	1	0.00	1.00	0.00	1.00
.....CGCTGATTGCATCTGCCCT.....	21	0	1.00	1	0.00	0.00	1.00	0.00

Predicted structure

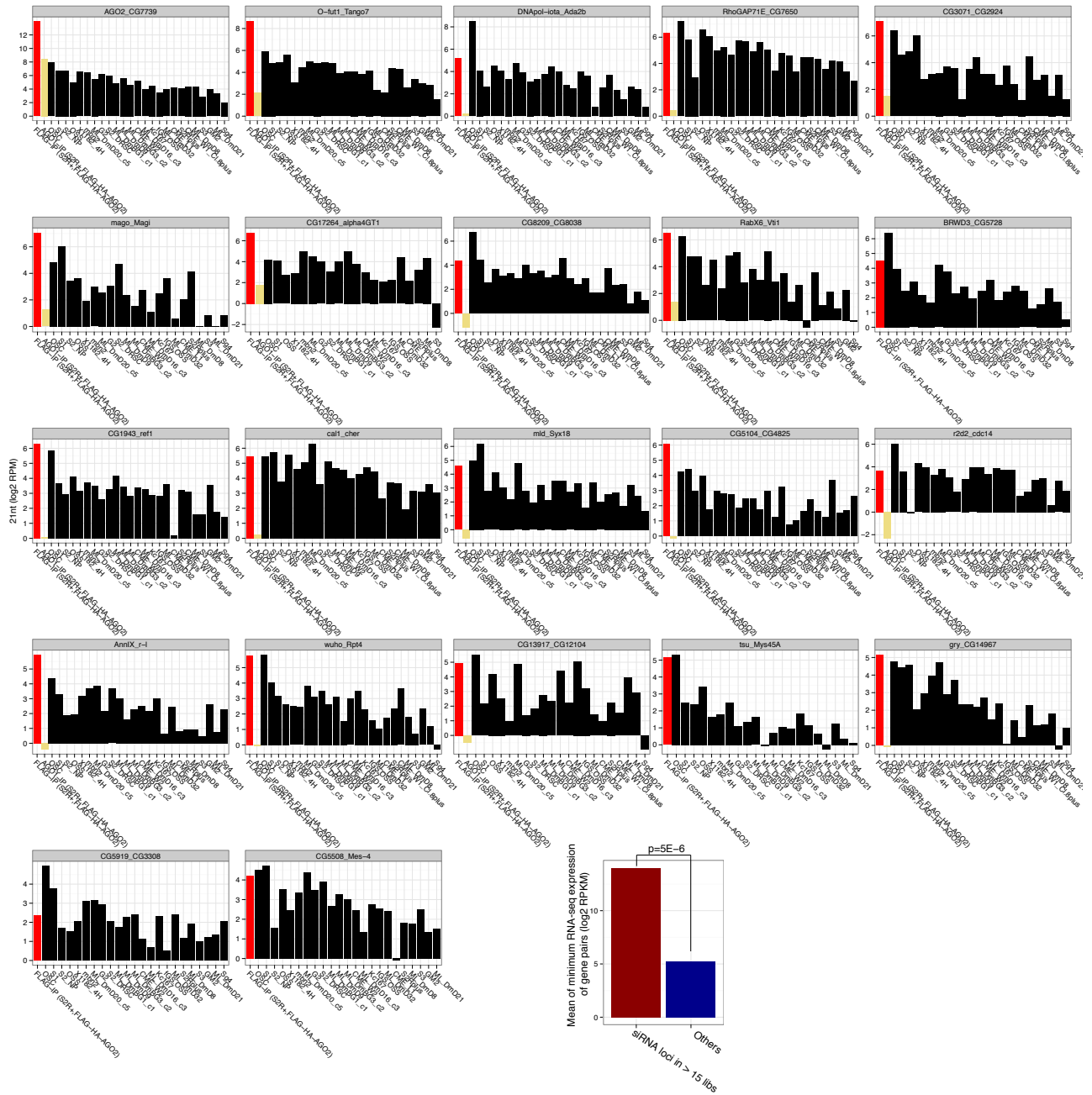
[illegible]

Anti-sense strand reads

[illegible]



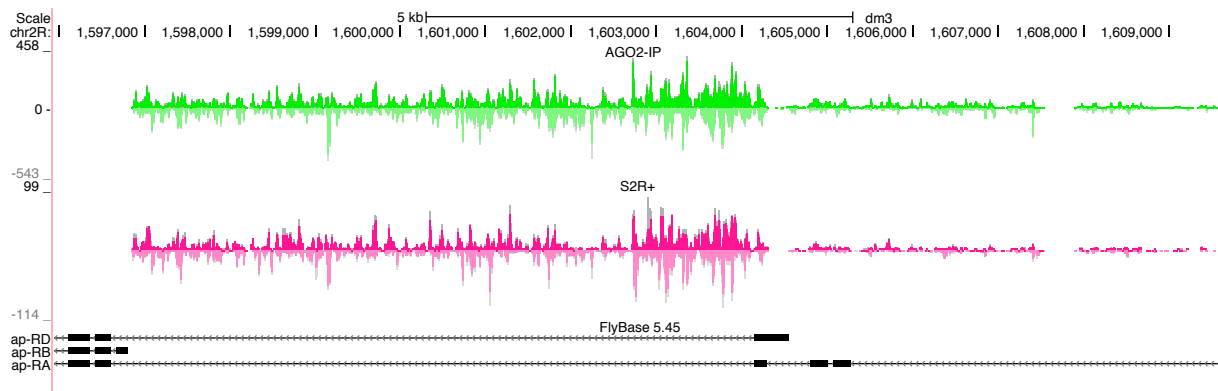
Supplementary Figure 10. siRNA differential expression between cell lines and tissues. Comparison of TE siRNA expression (A-D) or cis-NAT siRNA expression (E-F) in cell lines vs tissues (heads, testes, ovaries, embryos). Differentially expressed siRNAs (FDR < 1%) are shown in blue (high in cell lines) and red (high in tissues). Cell lines generally show elevated TE siRNA expression compared to various tissues. cis-NAT-siRNAs are expressed more highly in testes relative



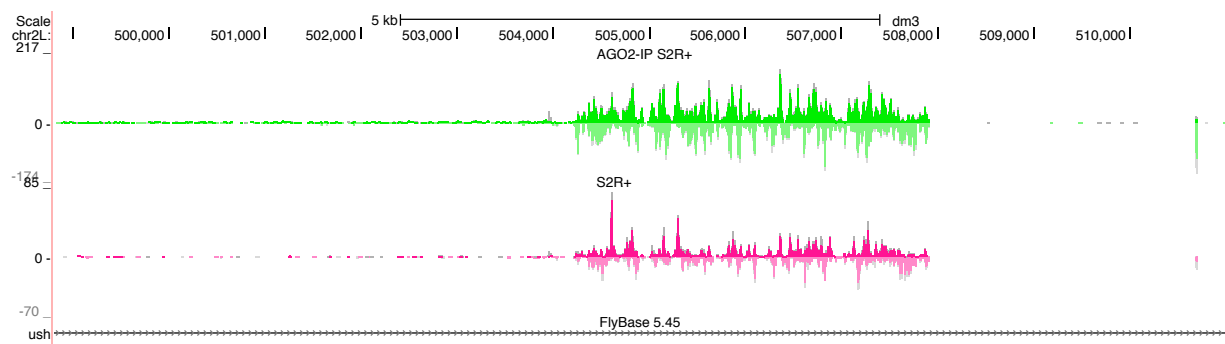
Supplementary Figure 11. 3'-cis-NAT-siRNA loci that were expressed in at least 15 cell lines. Shown are expression of those loci in 25 cell lines. AGO2-IP and AGO1-IP are shown in red and yellow, respectively. An additional plot shows mean of the minimum expression of gene pairs (RNA-seq) producing 3'-cis-NAT-siRNAs in at least 15 cell lines versus other gene pairs producing 3'-cis-NAT-siRNAs.

Supplementary Figure S12: More cis-NAT siRNA examples are listed in several classes.

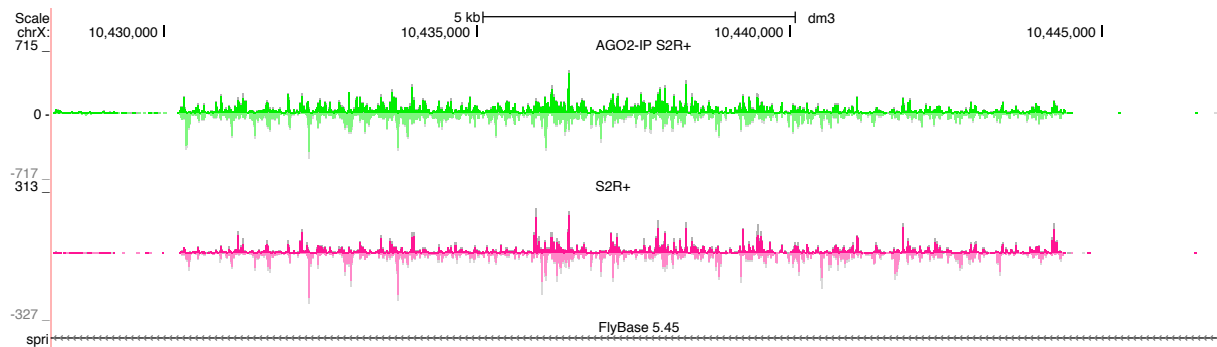
- (1) “klar-like” cis-NAT siRNAs produced from gene body
- (2) cis-NAT siRNAs produced from 5 divergent pairs
- (3) cell line specific cis-NAT siRNAs
- (4) cis-NAT siRNAs produced from un-annotated 3 overlapping regions
- (5) cis-NAT siRNAs in most cell lines



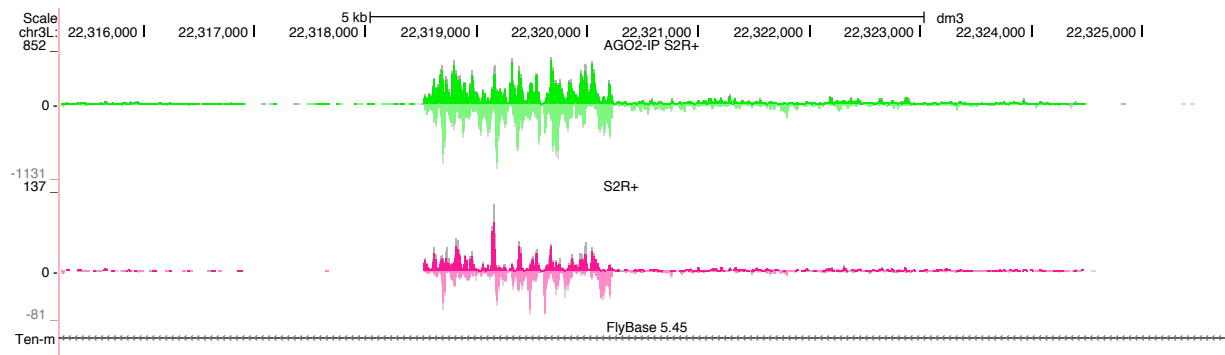
“klar-like” cis-NAT siRNAs produced from gene body



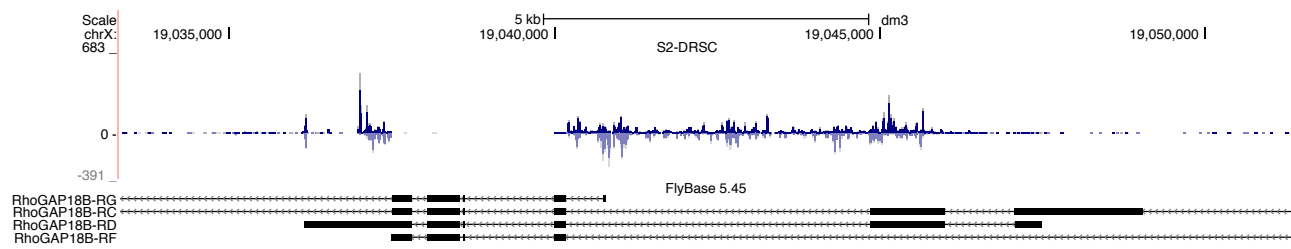
“klar-like” cis-NAT siRNAs produced from gene body

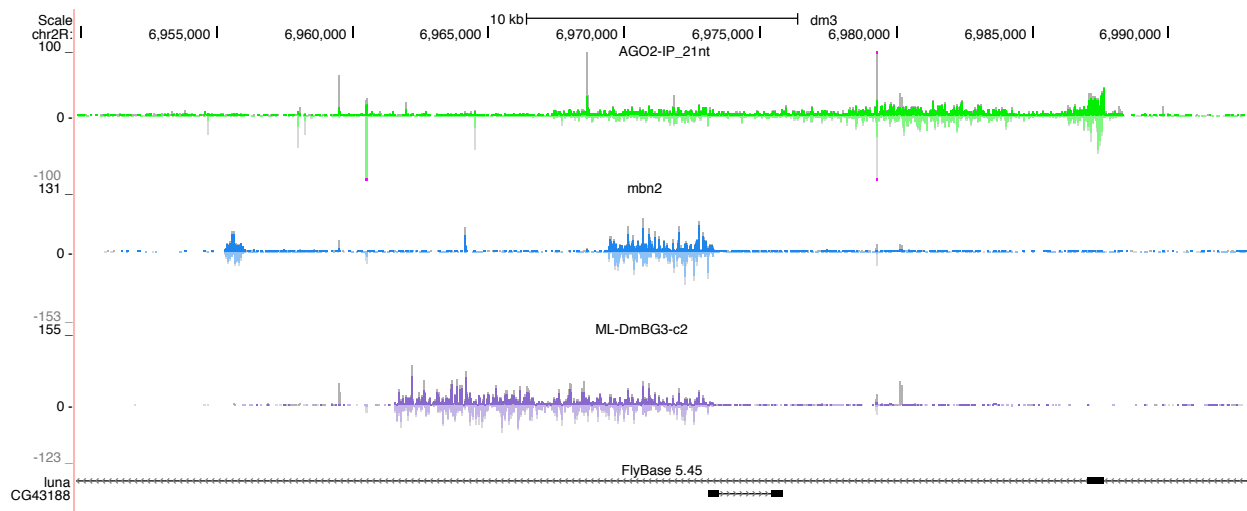


“klar-like” cis-NAT siRNAs produced from gene body

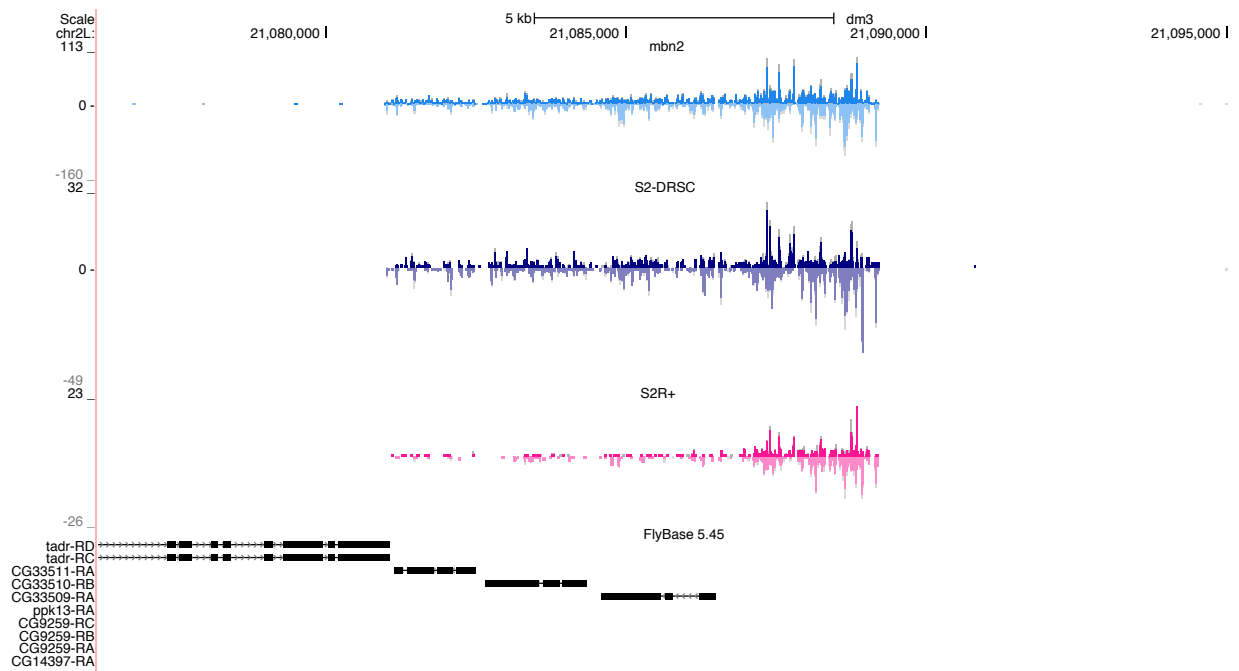


“klar-like” cis-NAT siRNAs produced from gene body

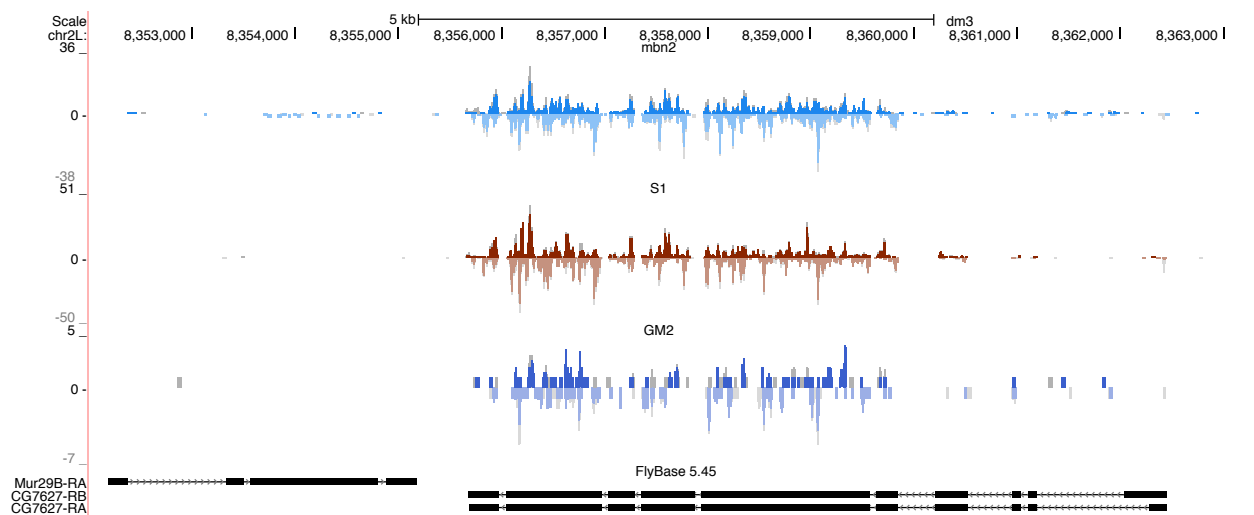




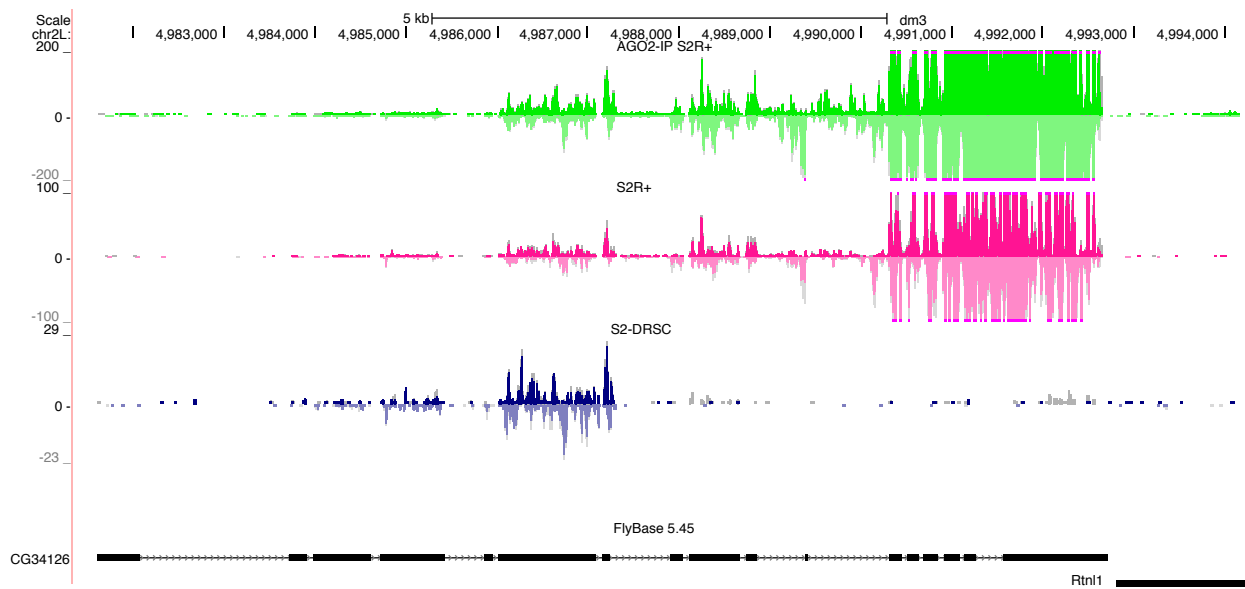
“klar-like” cis-NAT siRNAs produced from gene body



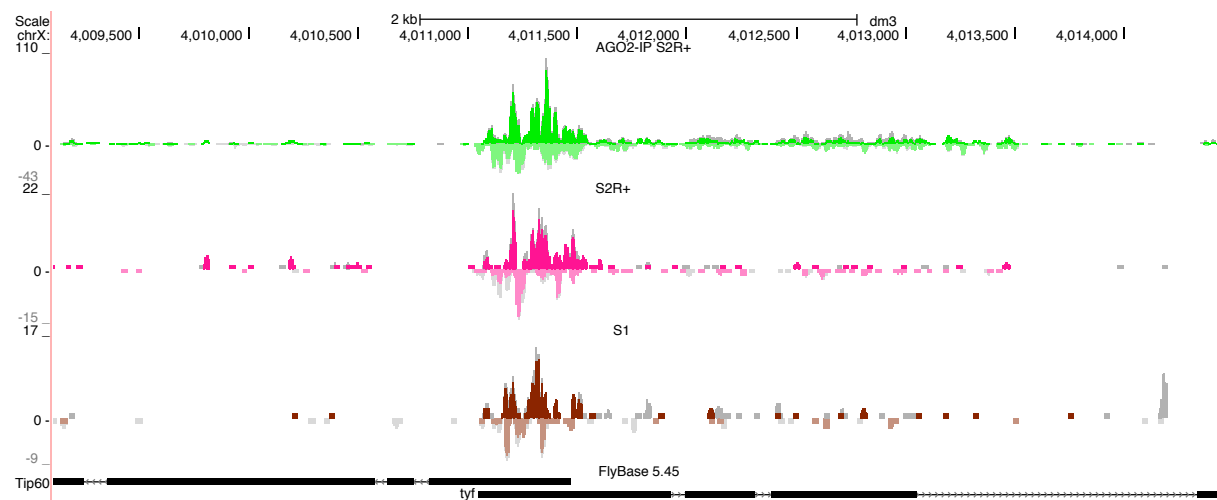
“klar-like” cis-NAT siRNAs produced from gene body

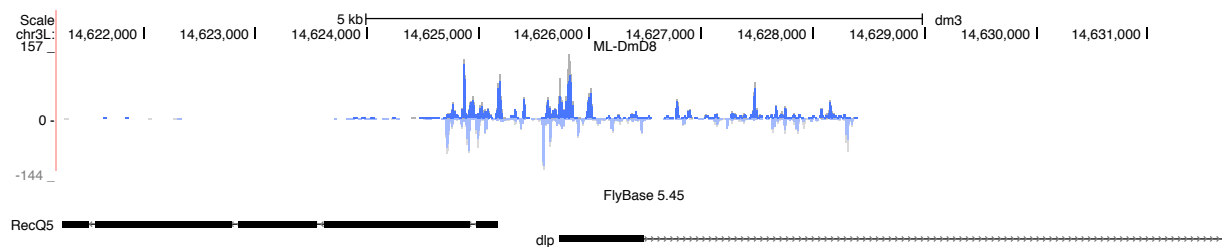


“klar-like” cis-NAT siRNAs produced from gene body

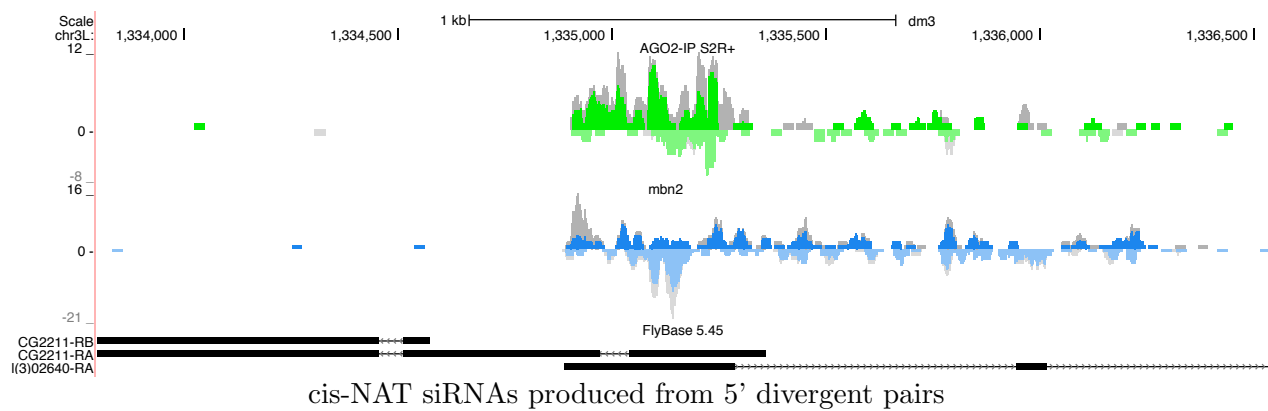


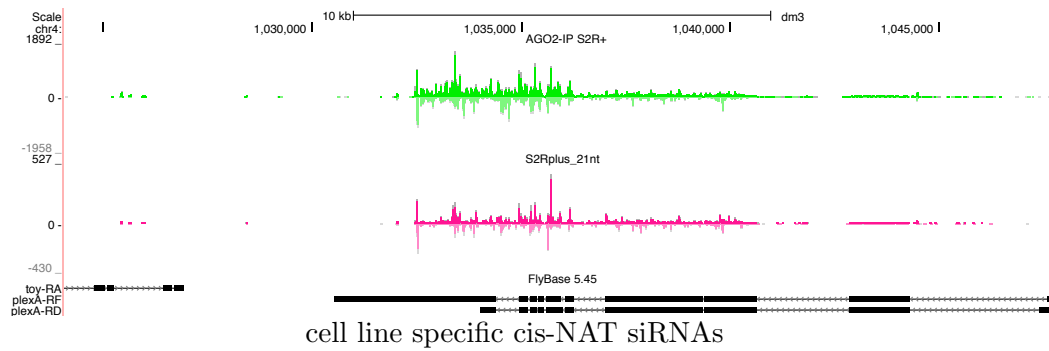
“klar-like” cis-NAT siRNAs produced from gene body

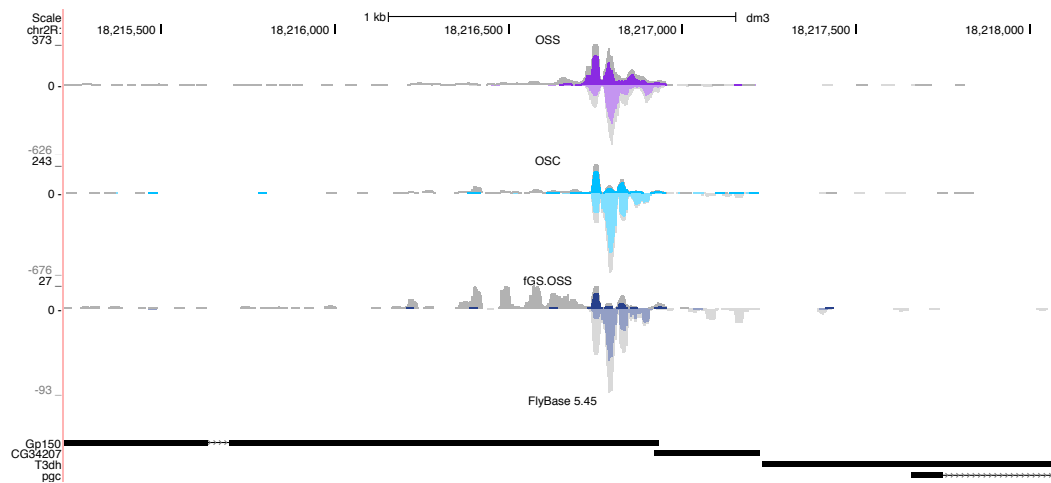




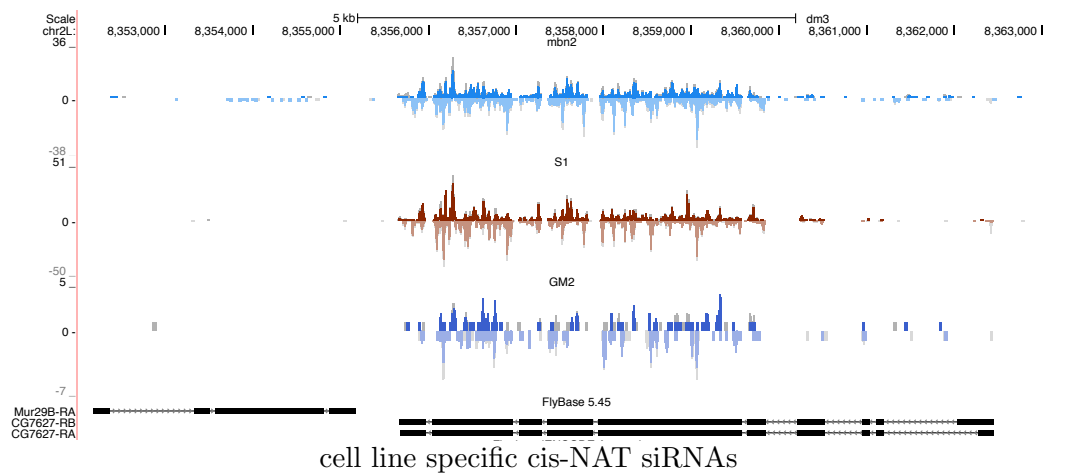
cis-NAT siRNAs produced from 5' divergent pairs

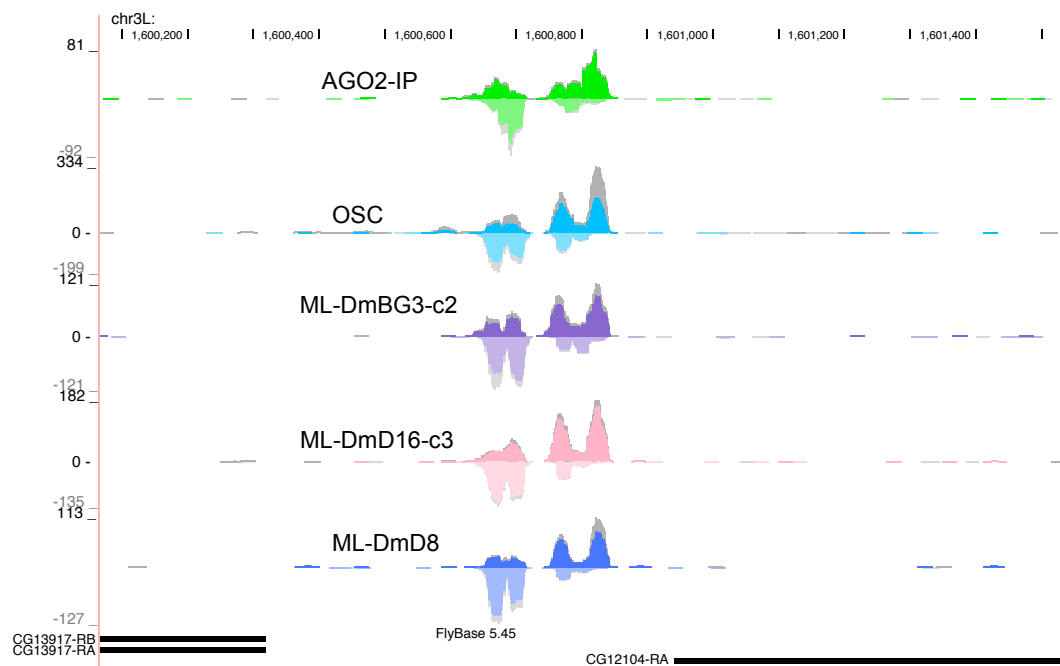




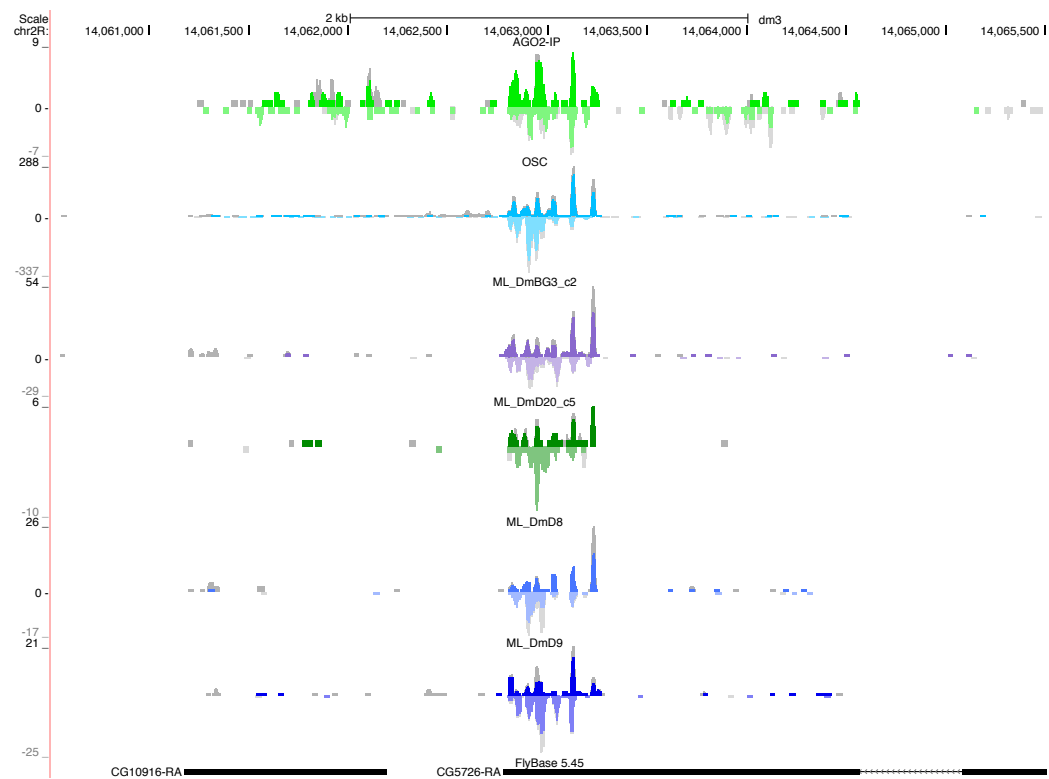


cell line specific cis-NAT siRNAs

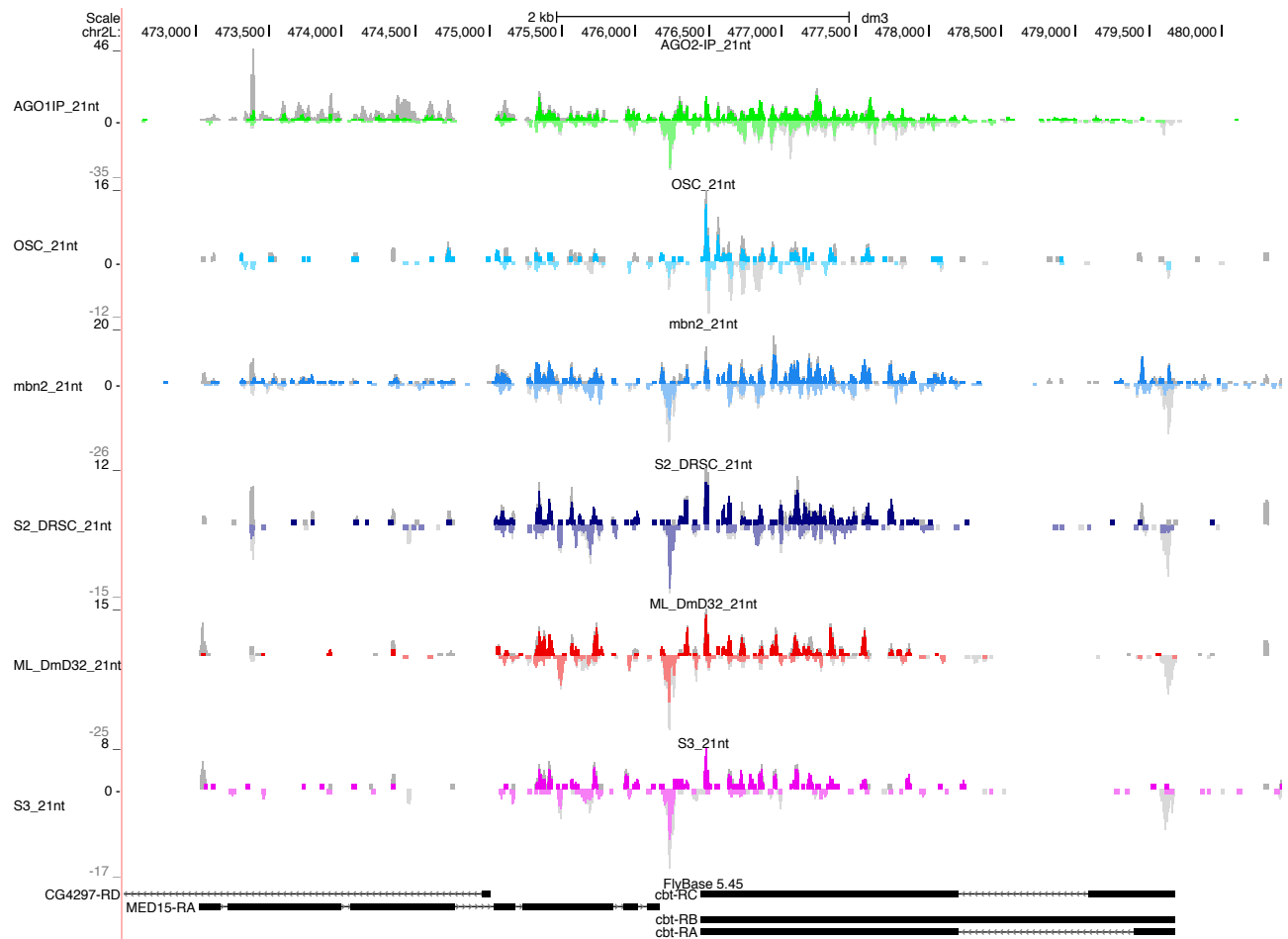




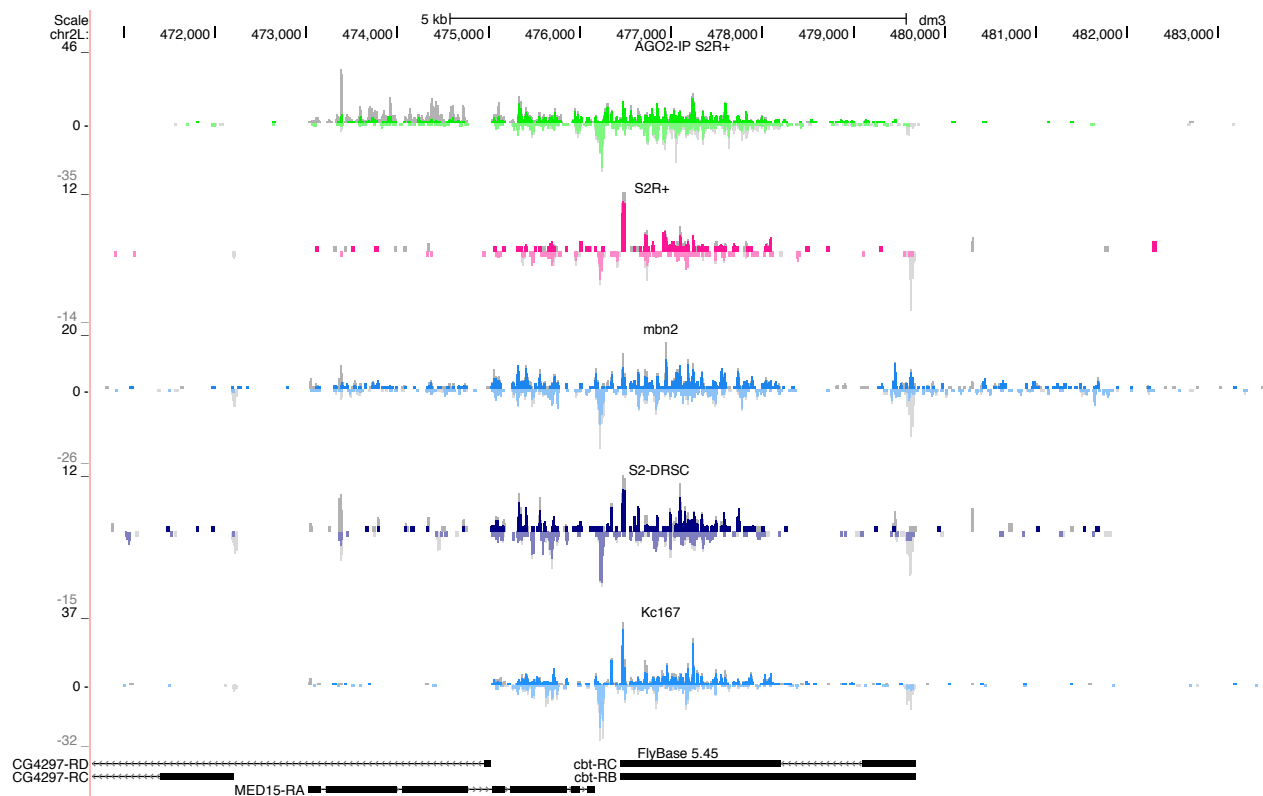
cis-NAT siRNAs produced from un-annotated 3' overlapping regions



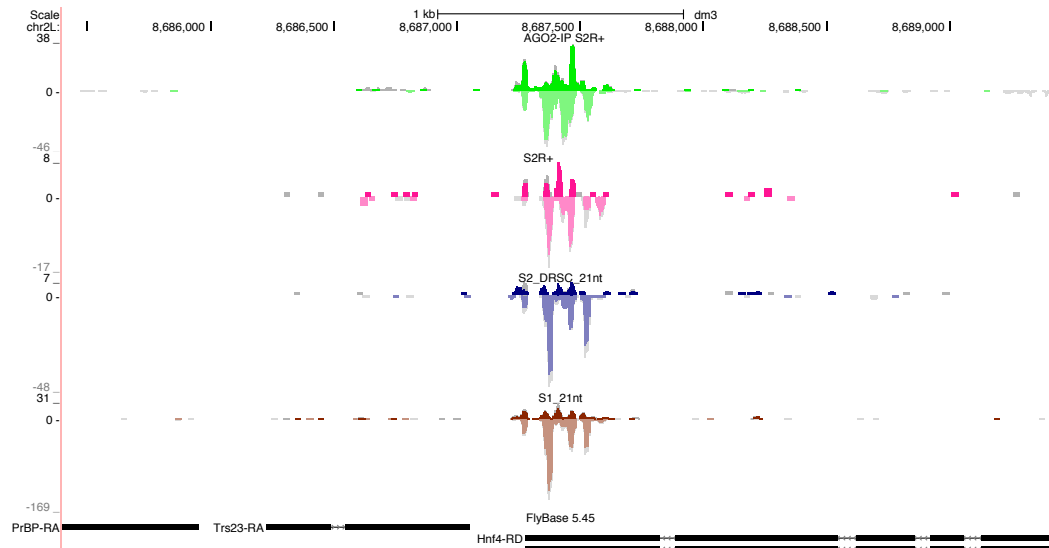
cis-NAT siRNAs produced from un-annotated 3' overlapping regions



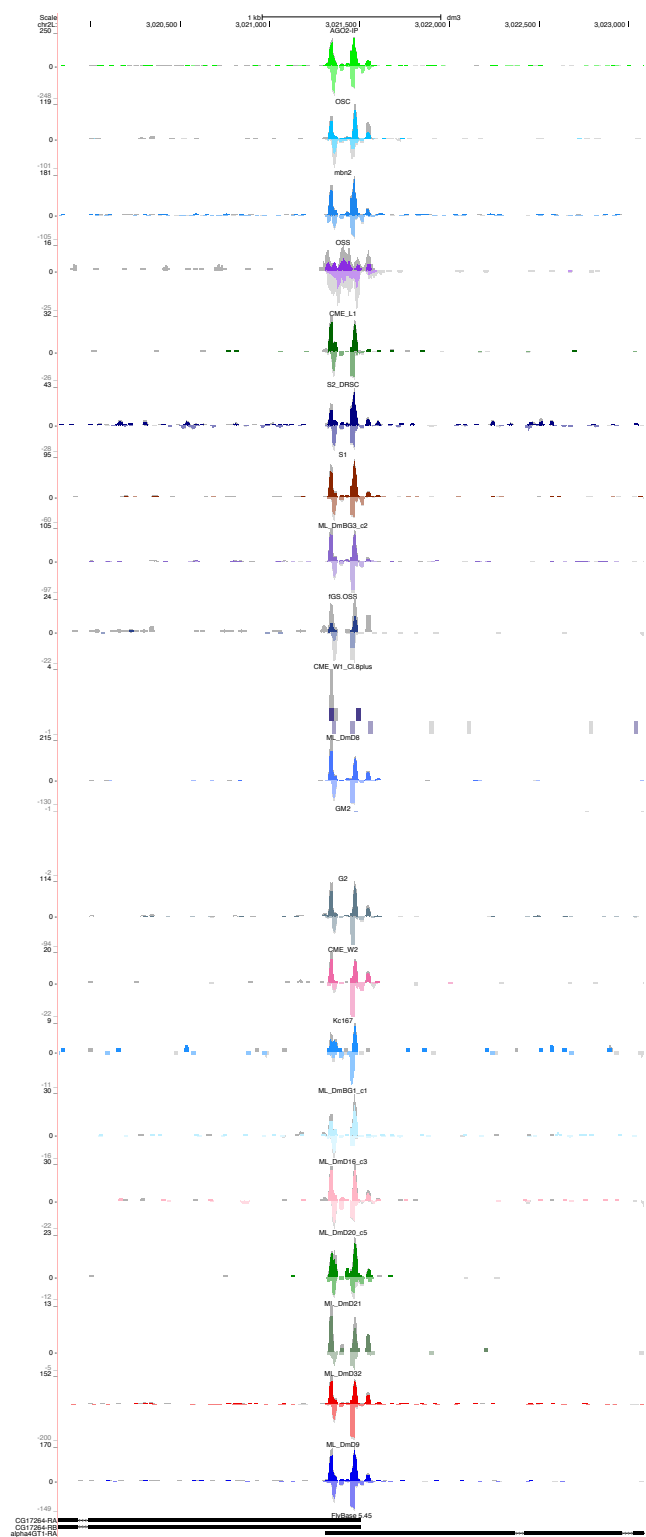
cis-NAT siRNAs produced from un-annotated 3' overlapping regions



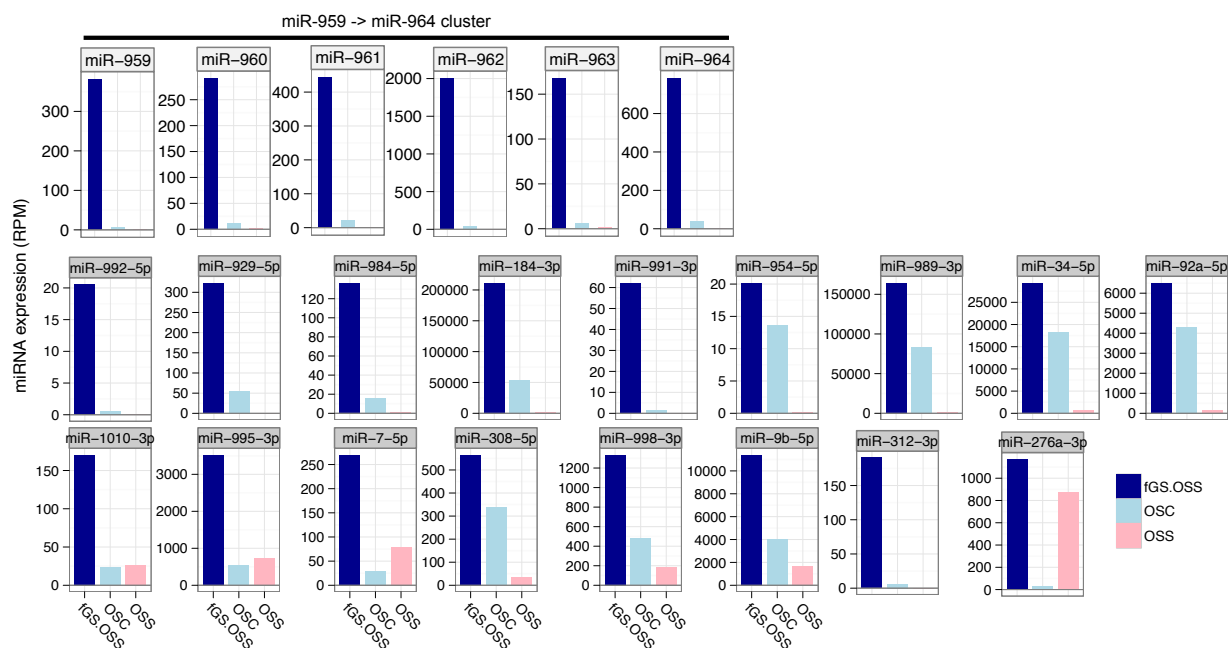
cis-NAT siRNAs produced from un-annotated 3' overlapping regions



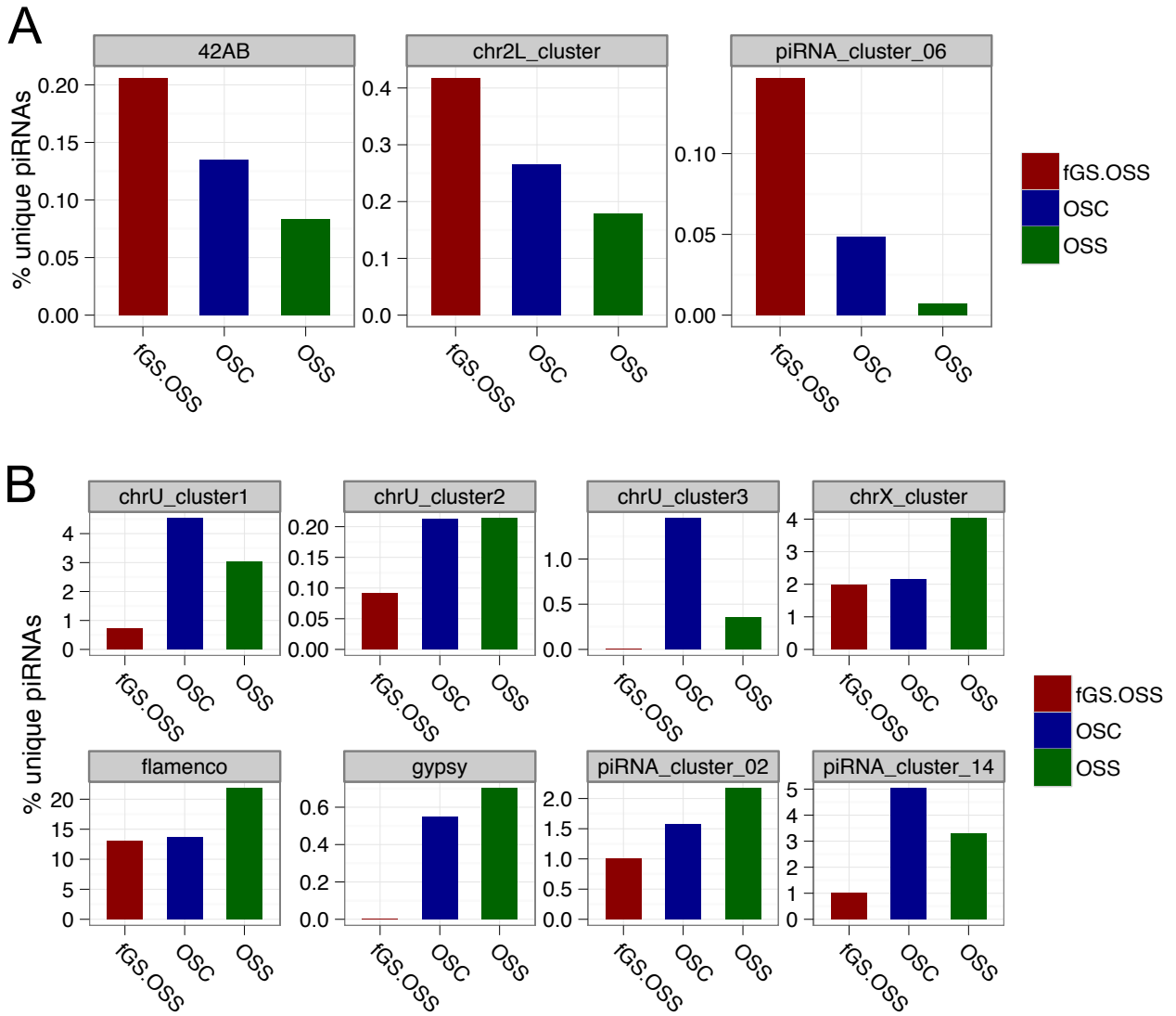
cis-NAT siRNAs produced from un-annotated 3' overlapping regions



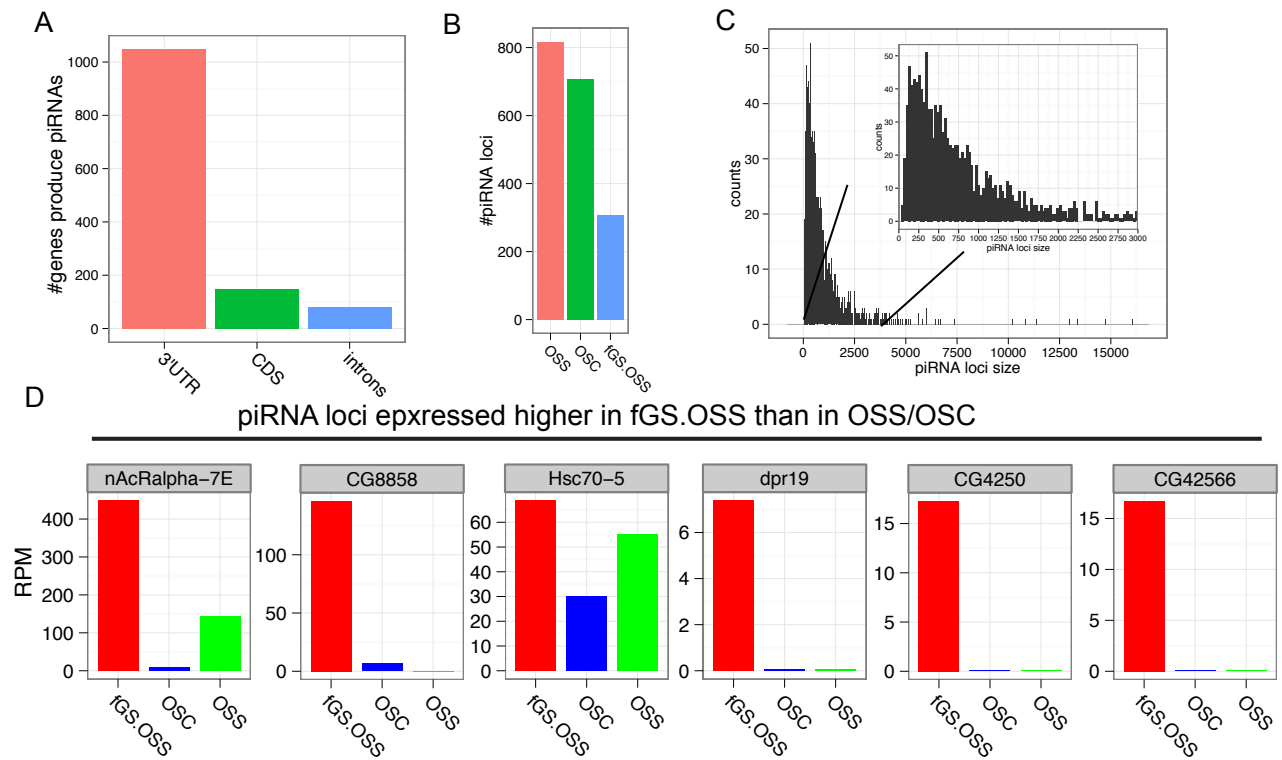
cis-NAT siRNAs in most cell lines



Supplementary Figure 13. Additional examples of miRNAs highly expressed in the mixed ovarian germline/somatic fGS/OSS cell line vs ovarian somatic cell lines OSS and OSC.



Supplementary Figure 14. piRNA master cluster expression. Y-axis is the fraction of unique 24-30nt reads mapped to a master piRNA cluster loci divided by total number of unique 24-30nt reads in the library. (A) piRNA master clusters expressed higher in the mixed germline/somatic (fGS/OSS) than in somatic cell lines (OSS and OSC). (B) piRNA master clusters expressed higher in OSS and OSC, compared to fGS/OSS.

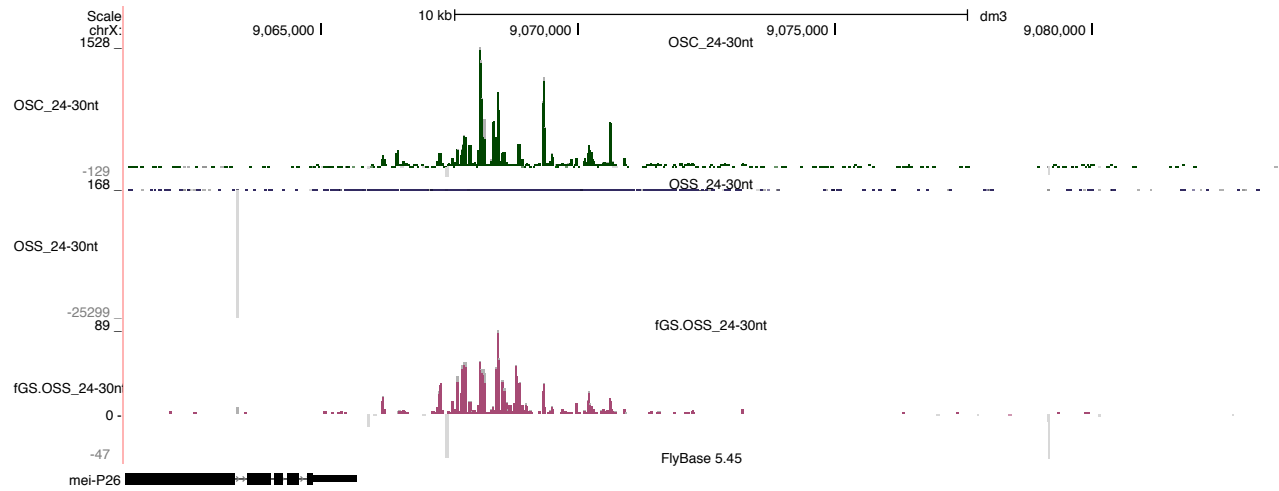


Supplementary Figure 15. (A) Genomic distribution of identified piRNA loci. (B) Number of identified piRNA loci in three ovarian cell lines. (C) piRNA loci size distribution (inset shows zoomed view of <3000 bp size distribution). (D) fGS/OSS-enriched piRNA loci.

Supplementary Figure S16: Representative piRNA loci in several classes.

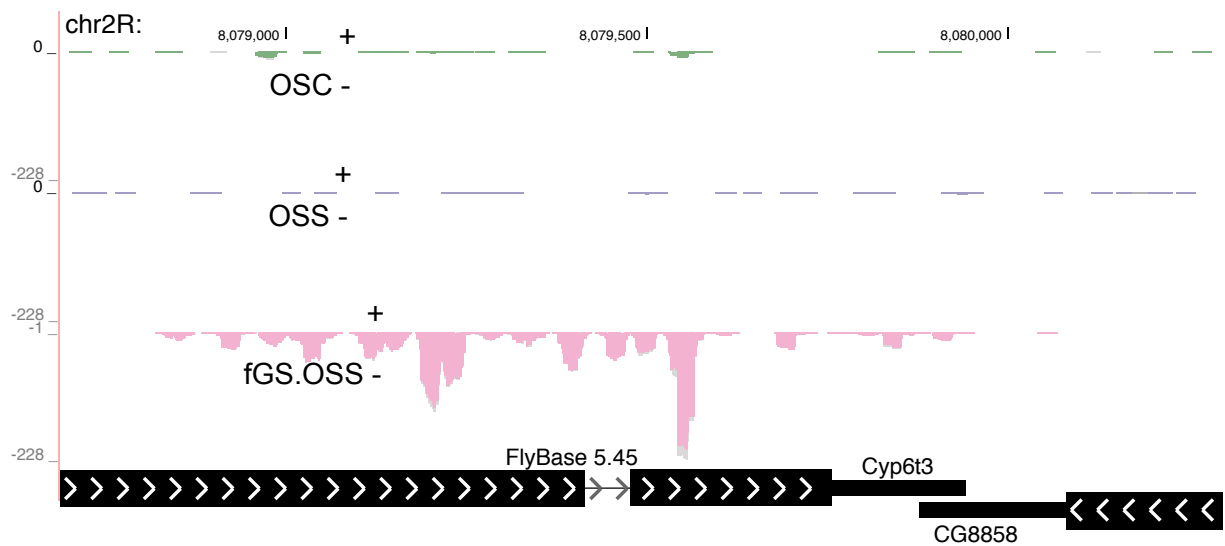
- (1) Examples of piRNAs in unannotated 3'UTRs
- (2) Examples of fGS/OSS enriched piRNAs
- (3) Examples of exonic piRNAs
- (4) Examples of intronic piRNAs
- (5) Examples of piRNAs spanning both CDS and intron
- (6) Examples of antisense piRNAs
- (7) Examples of bidirectional piRNAs

– Examples of piRNAs in unannotated 3'UTRs



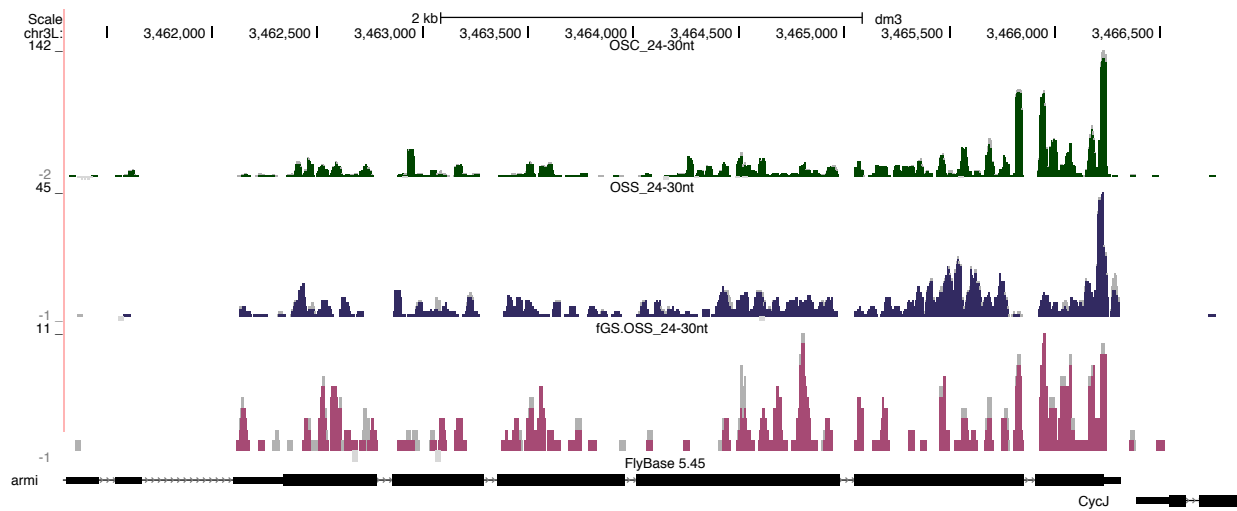
mei-P26 piRNAs reside on the unannotated 3'UTR

- Examples of fGS.OSS enriched piRNA loci

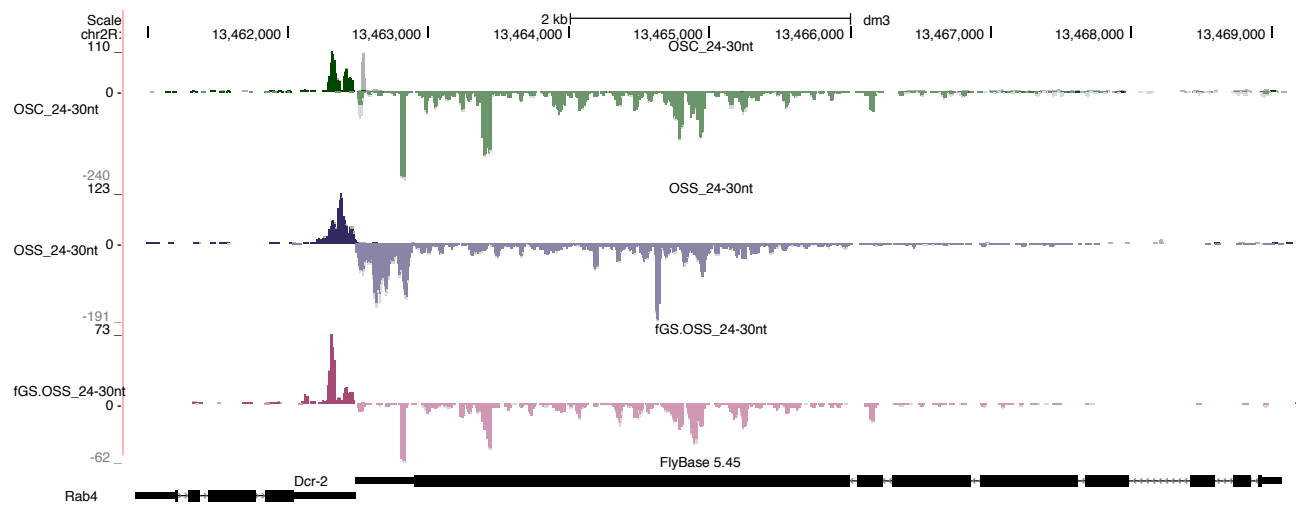


CG8858 piRNAs are fGS.OSS enriched and reside on the unannotated 3'UTR

– Examples of exonic piRNAs

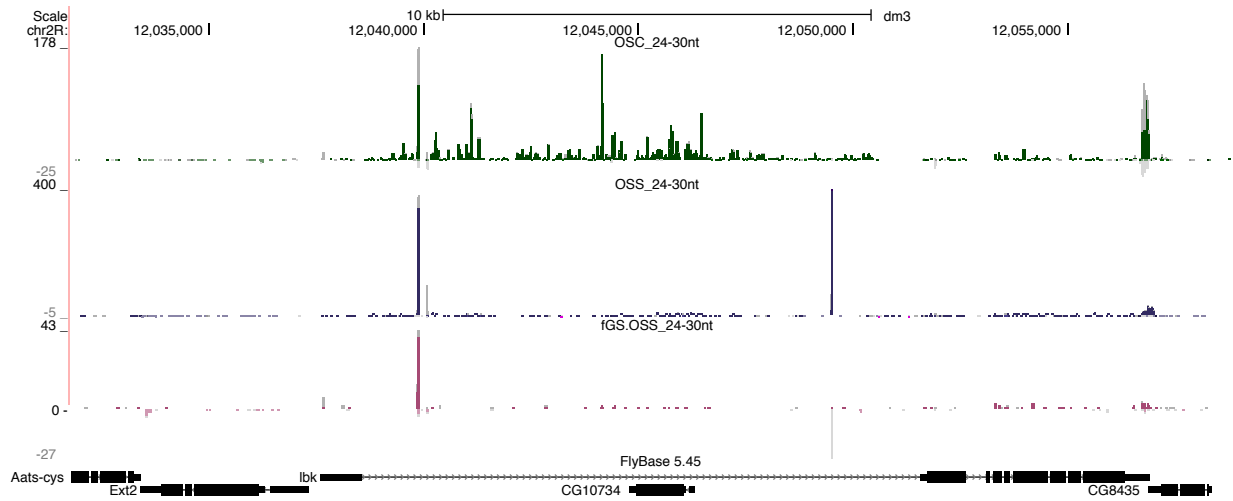


armi piRNAs are produced from CDS

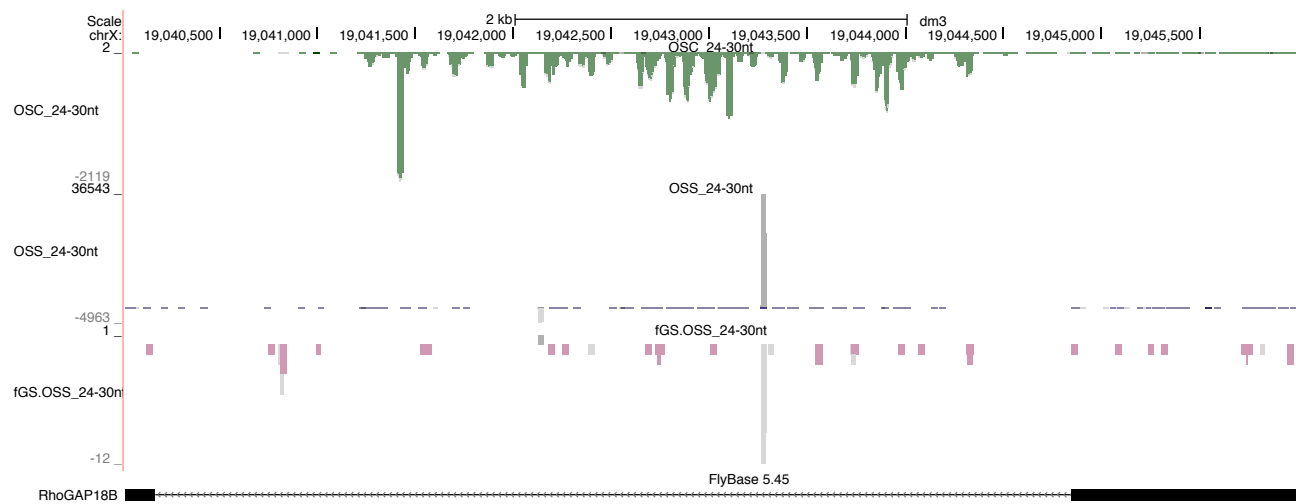


Dcr-2 produces exonic piRNAs

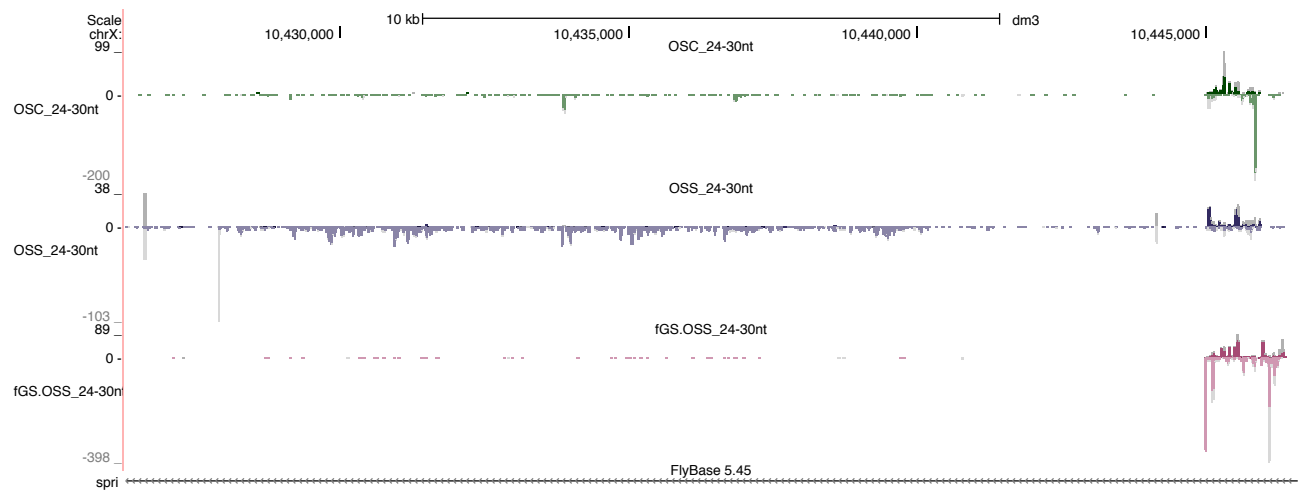
– Examples of intronic piRNAs



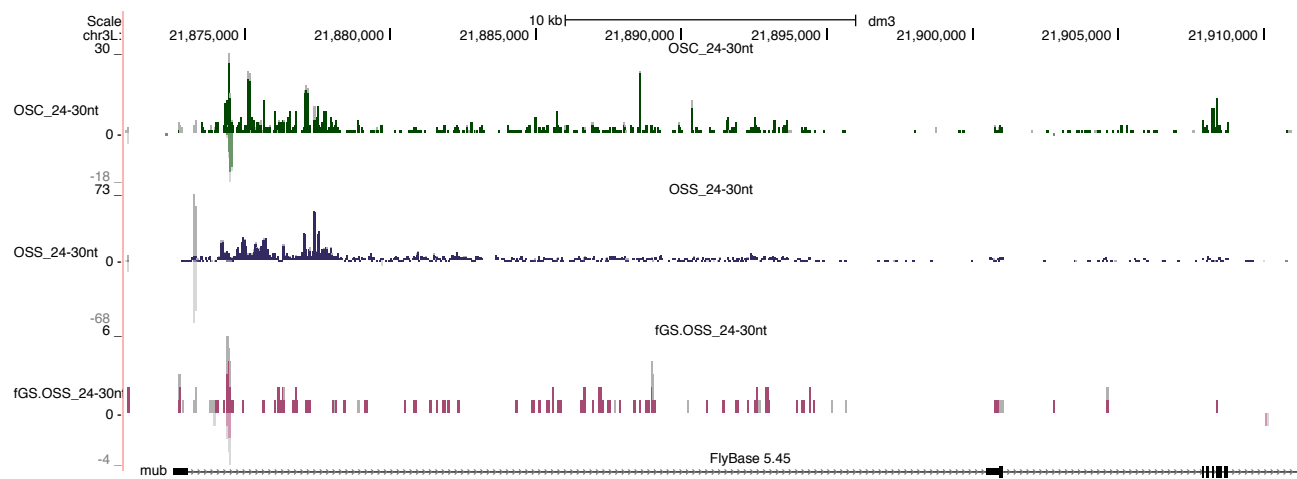
lbk piRNAs are produced from introns of 5'UTR in OSC cells



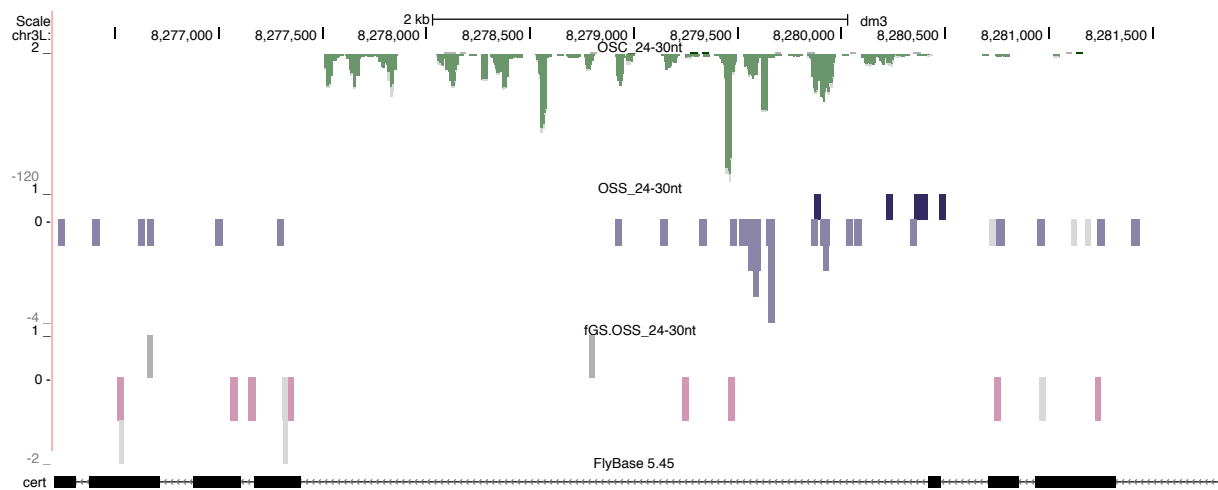
RhoGAP18B has abundant piRNAs in introns



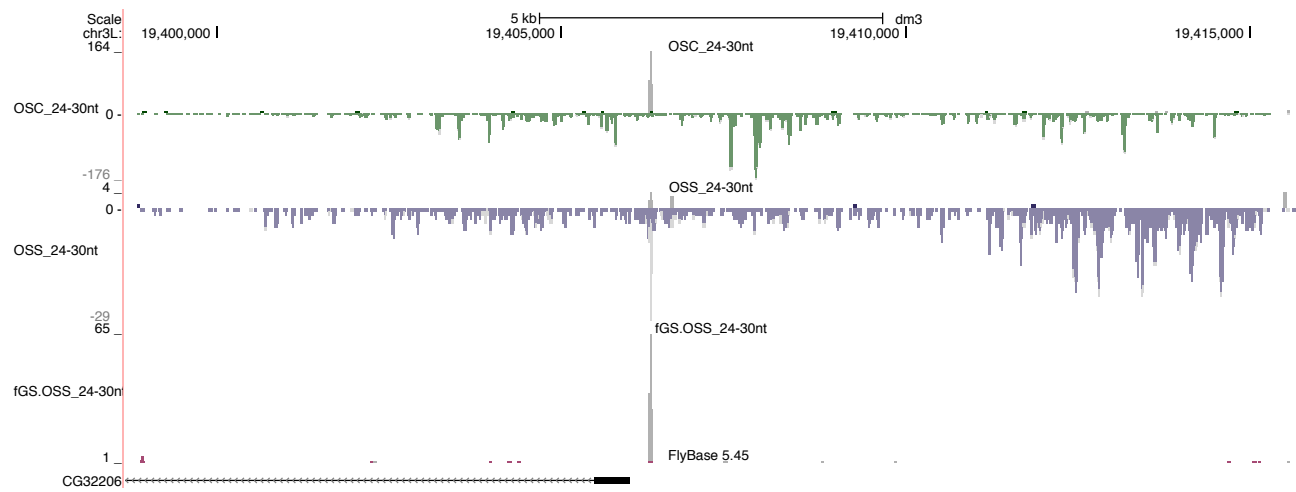
spr has 16kb piRNAs in introns



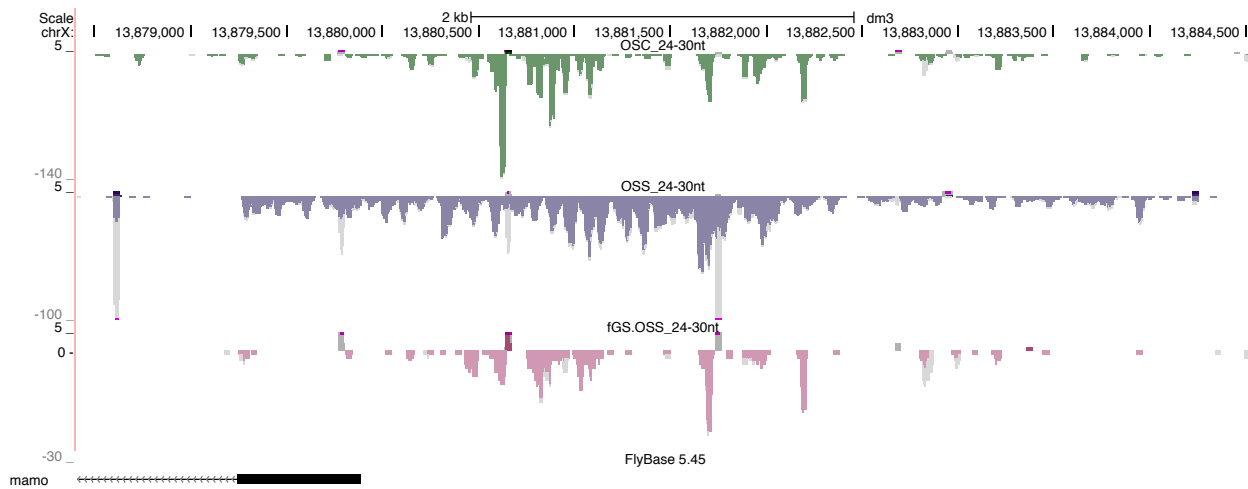
mub has 11kb piRNAs in introns



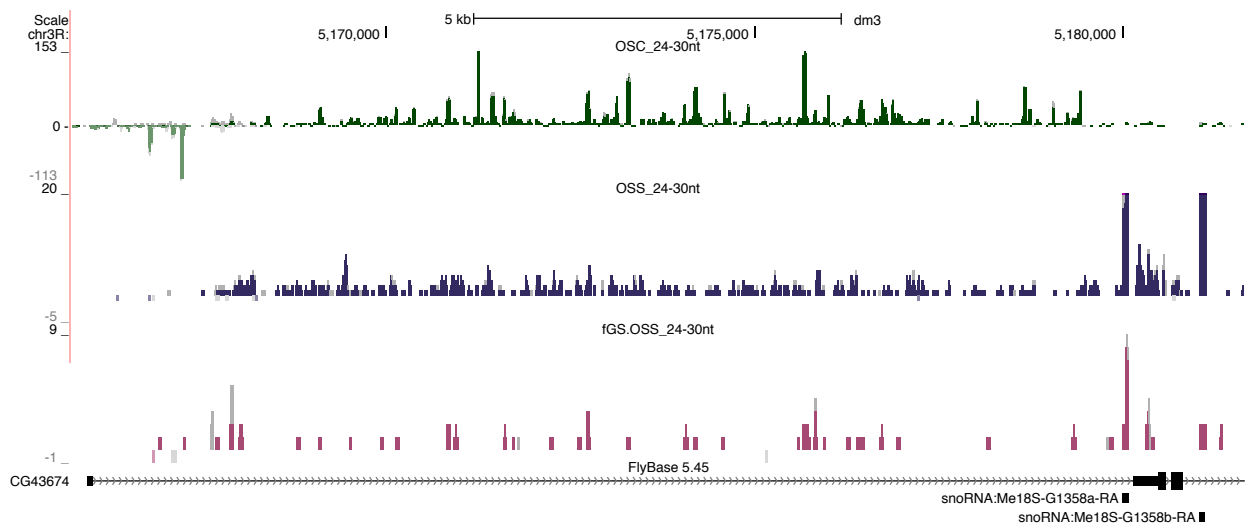
cert has intronic piRNAs in OSC cell



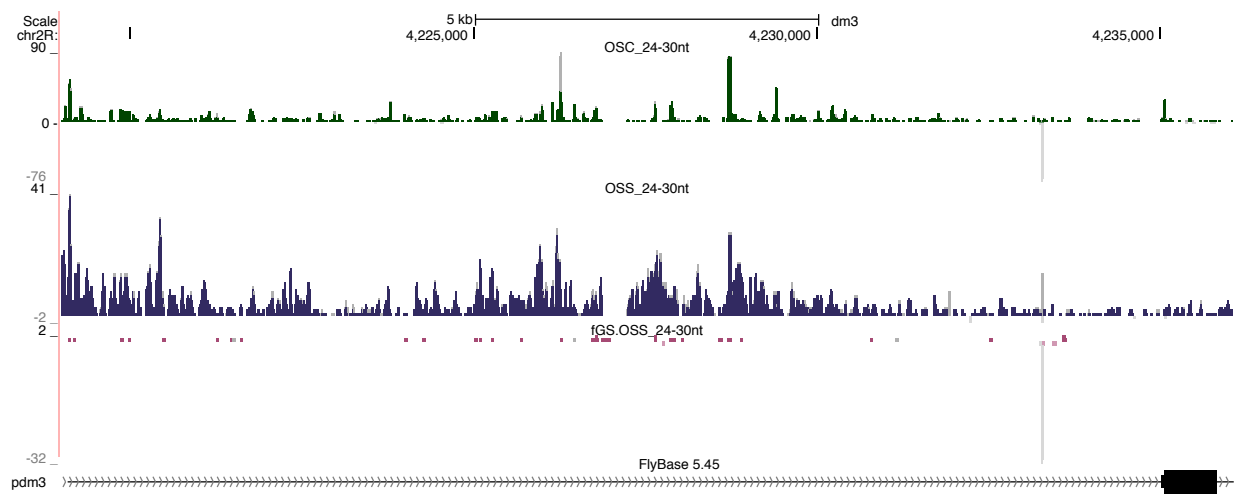
piRNAs span the upstream and 5' UTR of CG32206



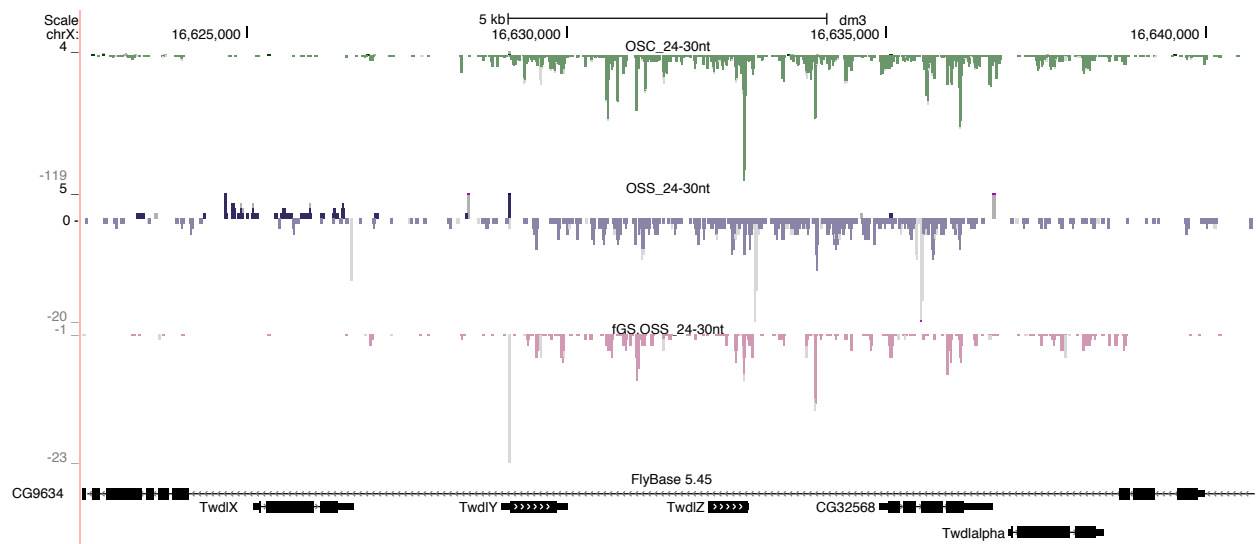
piRNAs located in the upstream region of *mamo*



CG43674 has intronic piRNAs

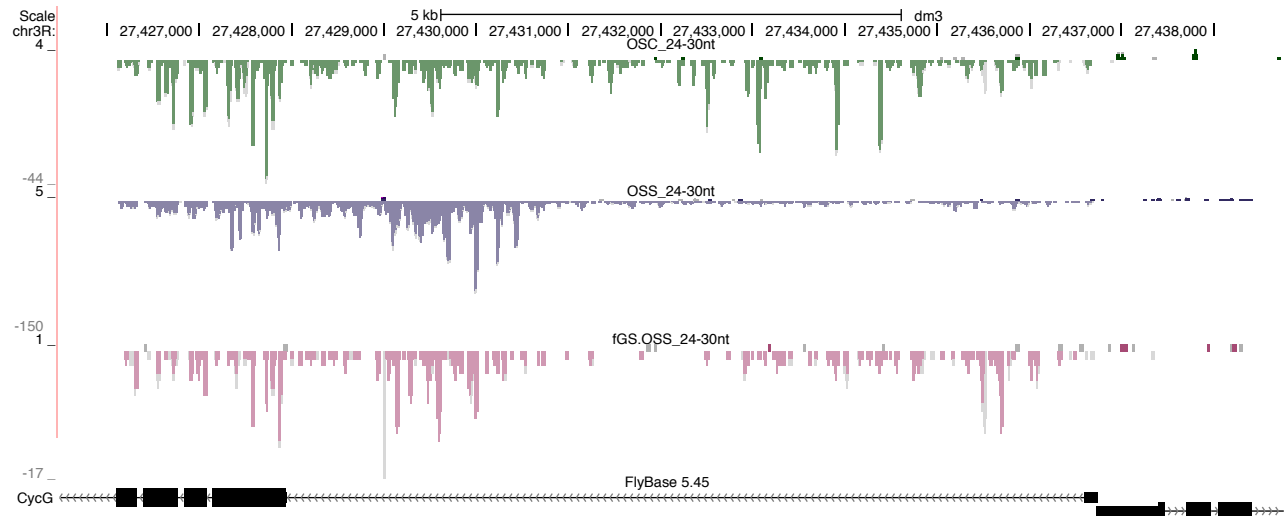


pdm3 has intronic piRNAs

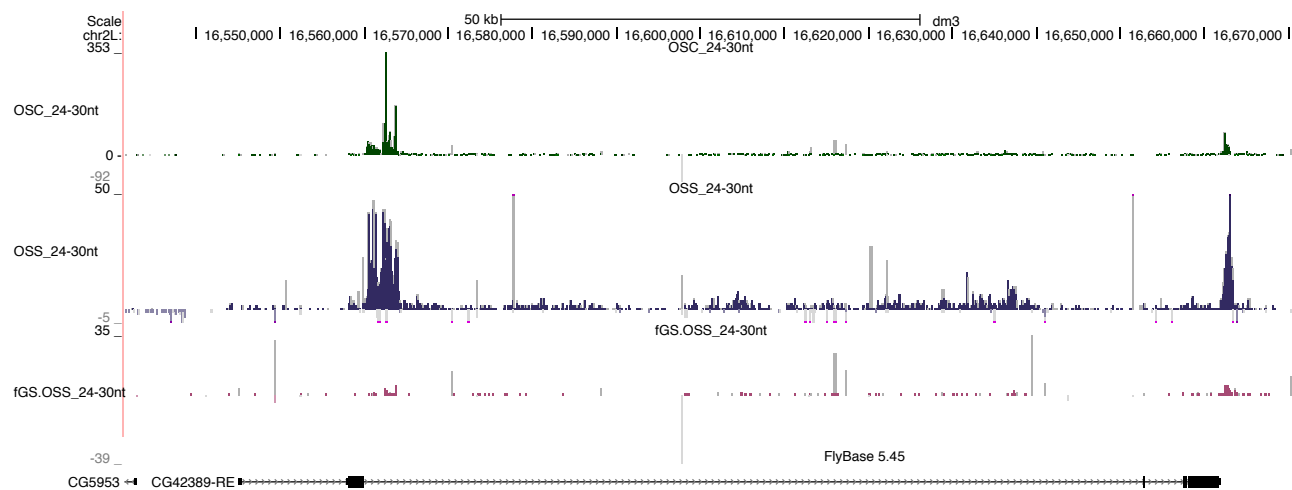


CG9634 has intronic piRNAs

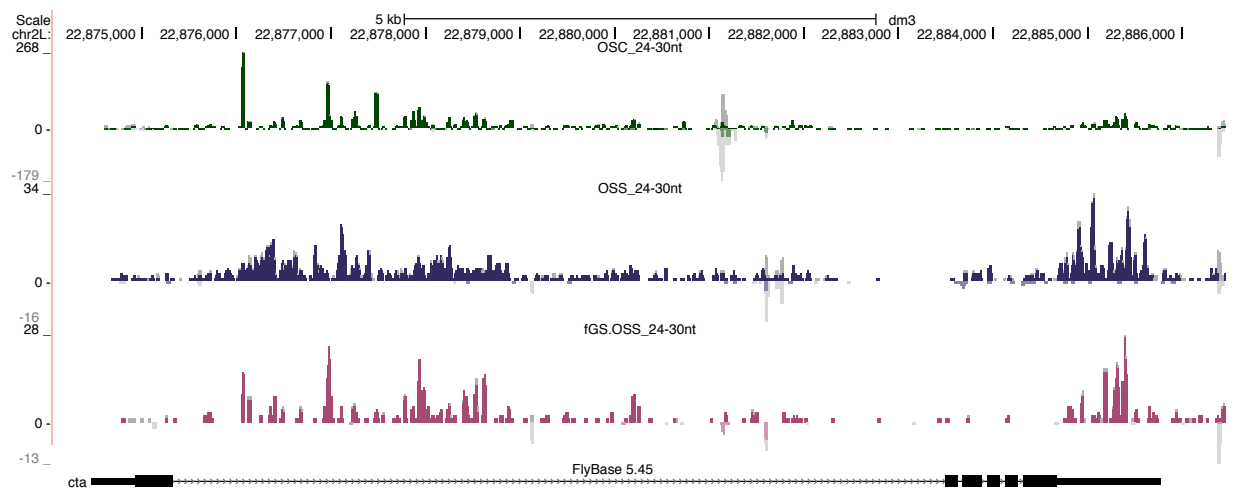
– Examples of piRNAs spanning CDS/introns



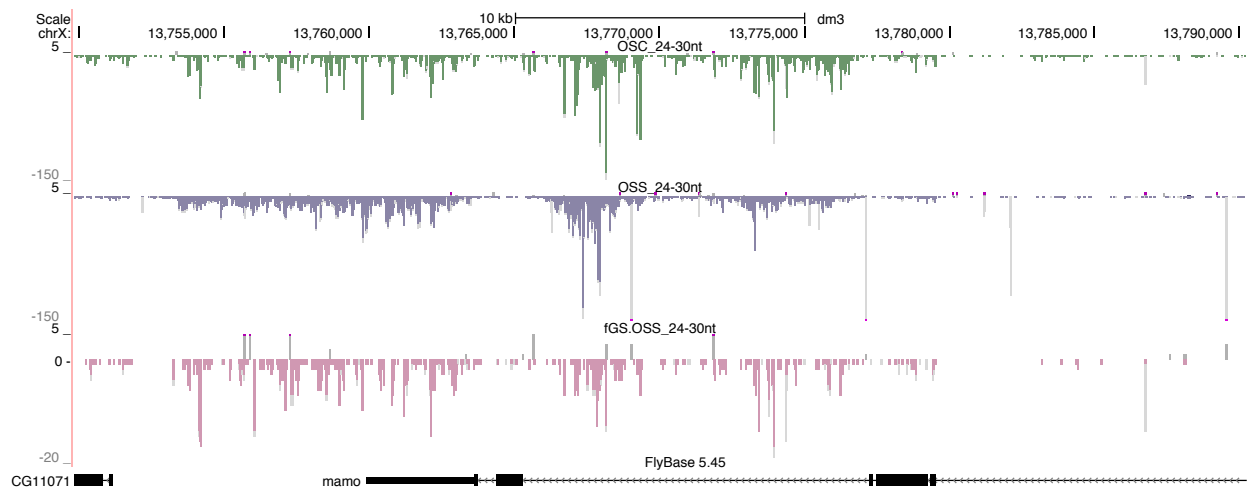
CycG produces piRNAs spanning CDS/introns



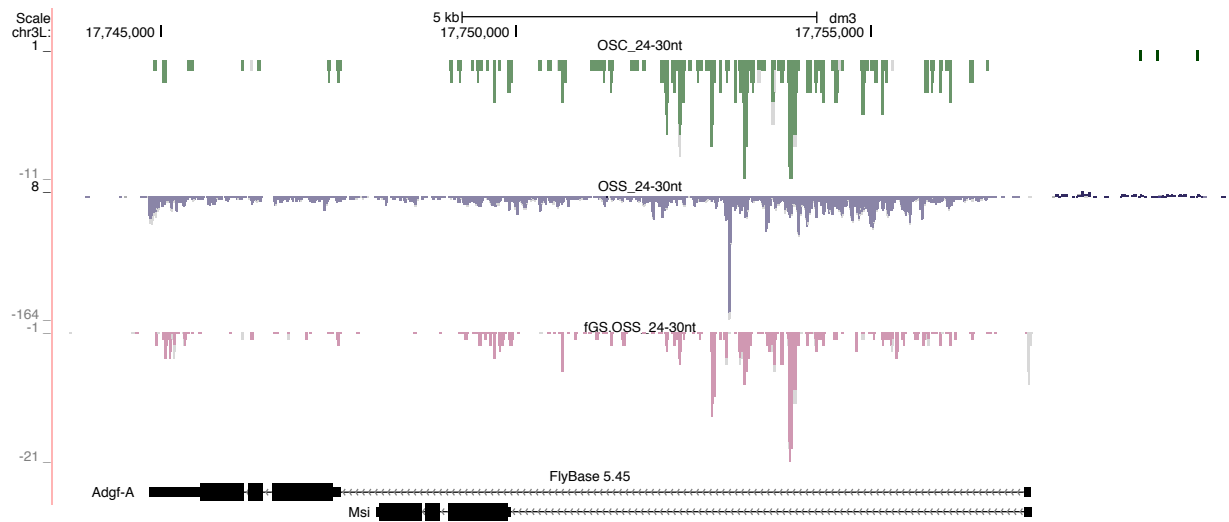
CG42389 produces piRNAs spanning CDS/introns



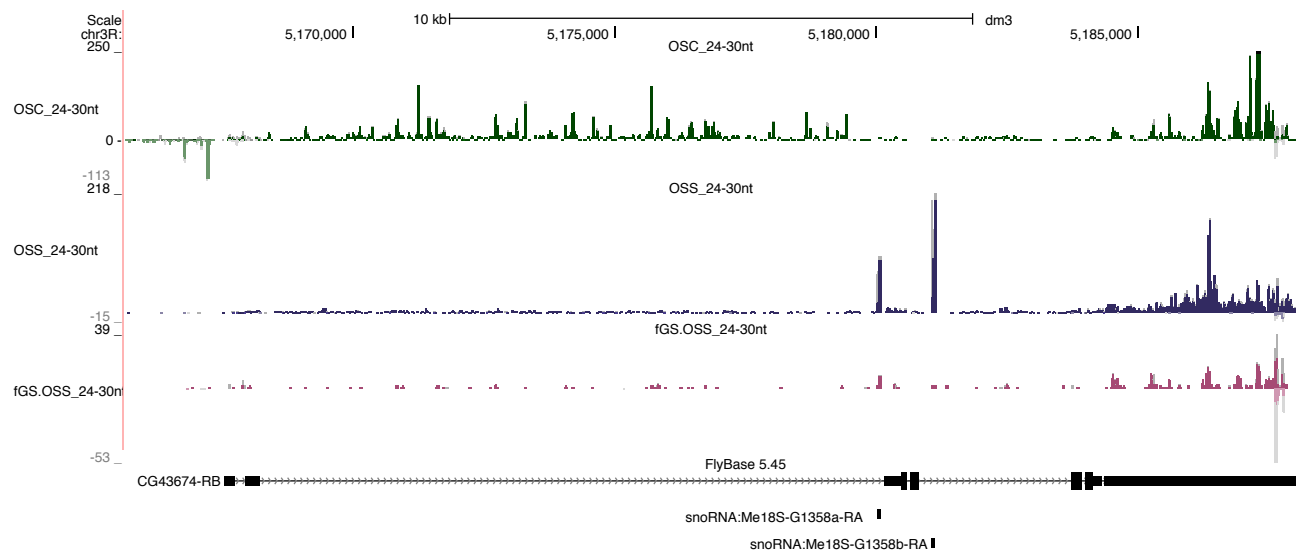
cta produces piRNAs spanning CDS/introns



mamo produces piRNAs spanning 3'UTR/introns

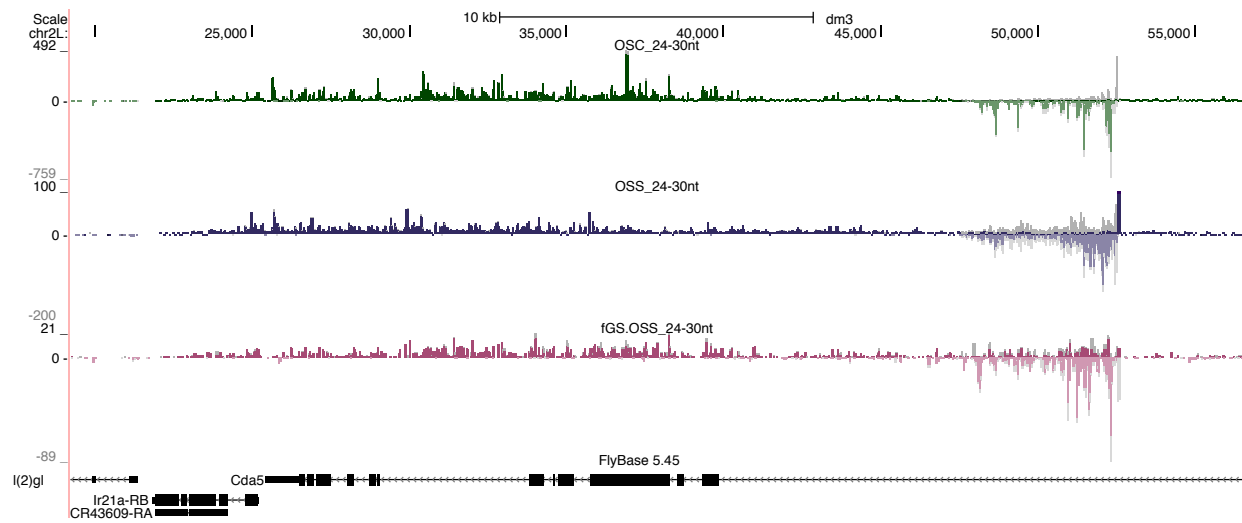


Msi produces piRNAs spanning CDS/introns

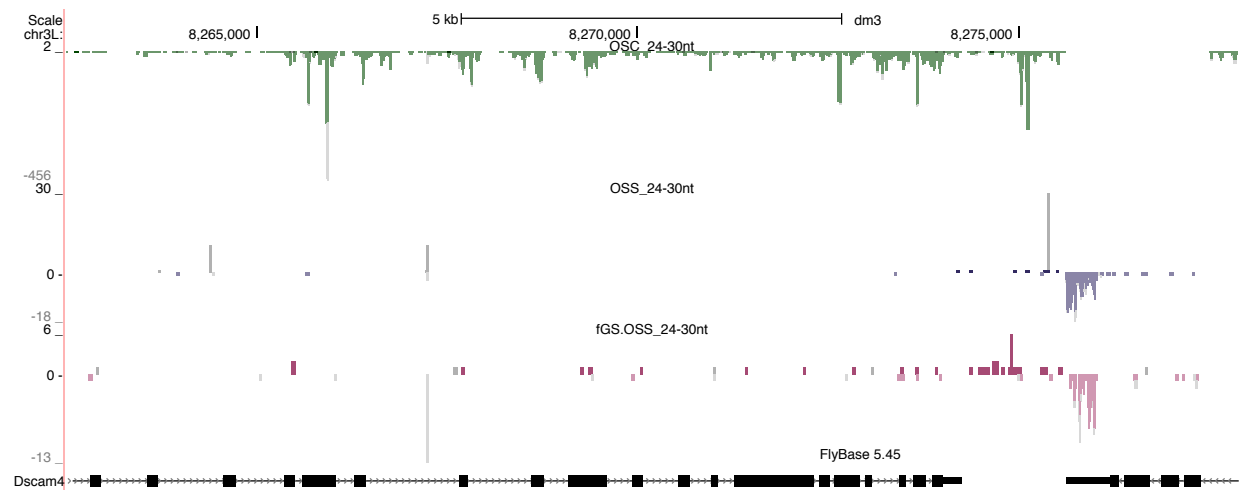


CG43674 has 3'UTR and intronic piRNAs

– Examples of antisense piRNAs

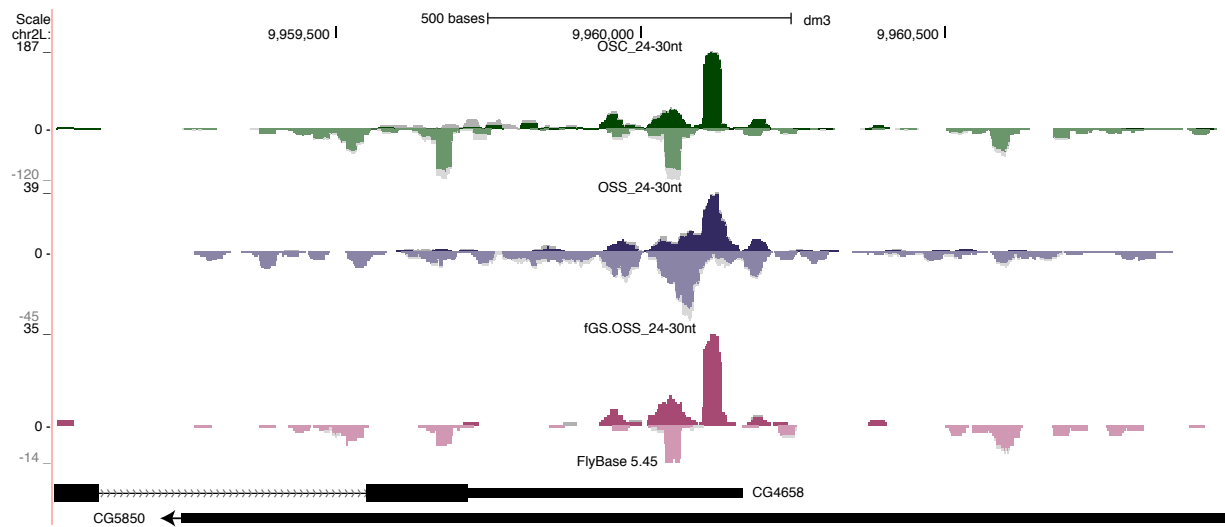


Cda5 produces both sense and antisense piRNAs

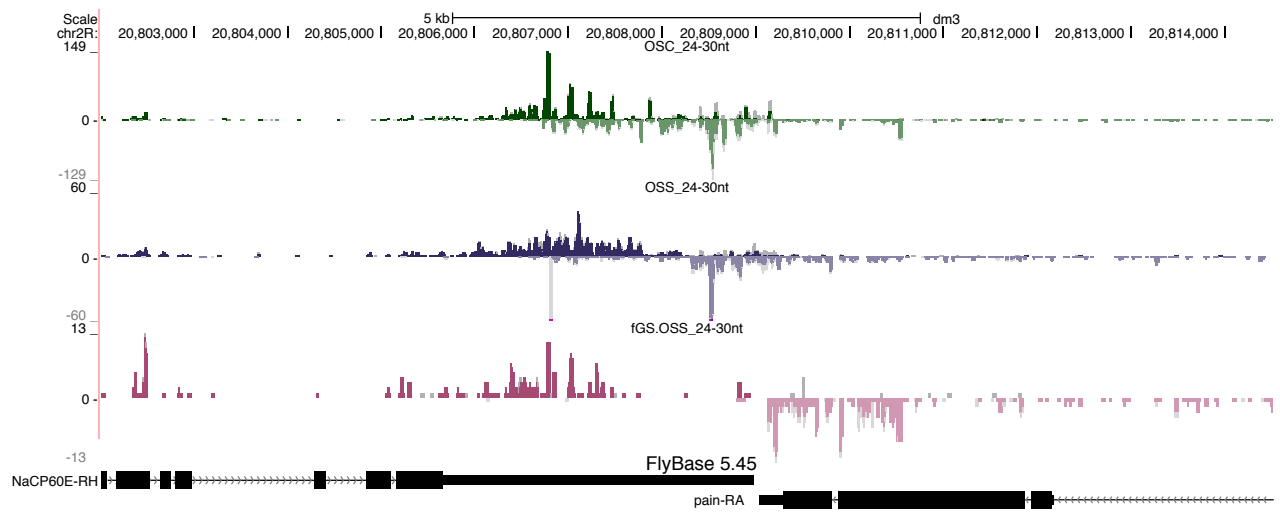


Dscam4 produces exonic antisense piRNAs

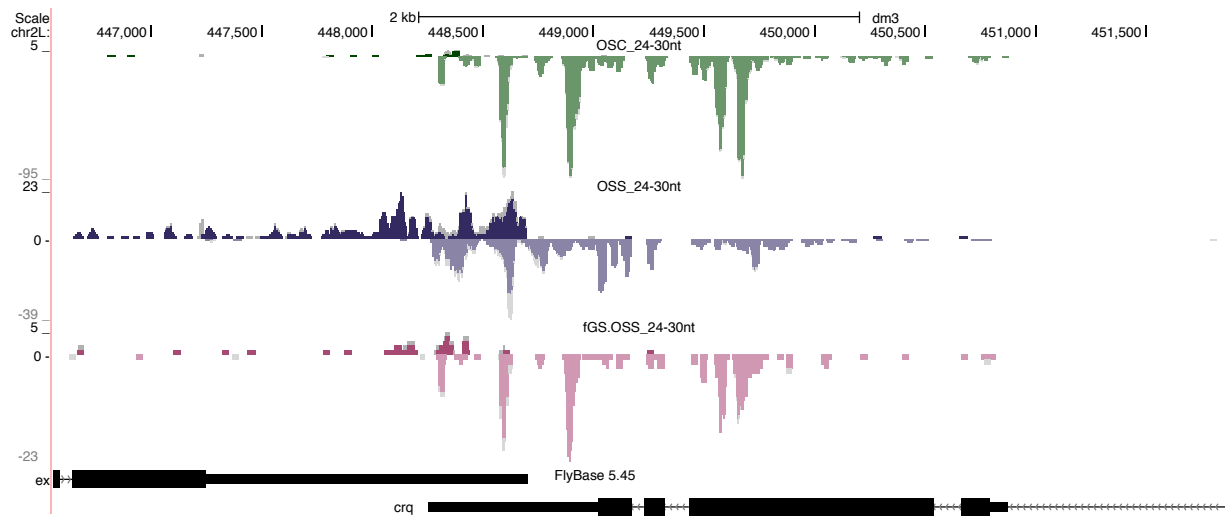
– Examples of bidirectional piRNA loci



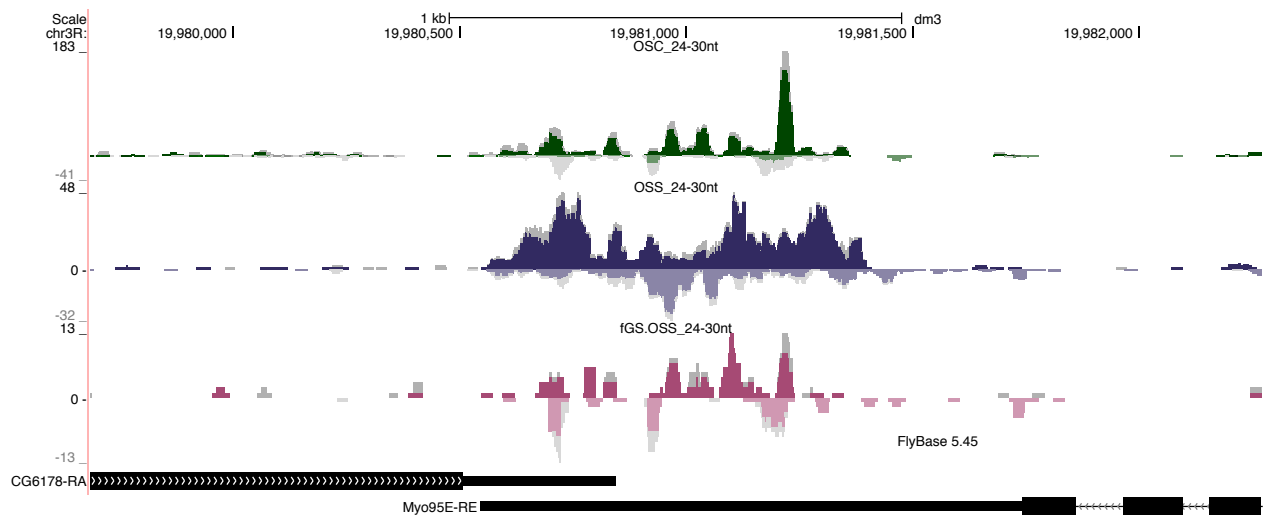
CG5850 and CG4658 3' UTR overlapping produce bidirectional piRNAs



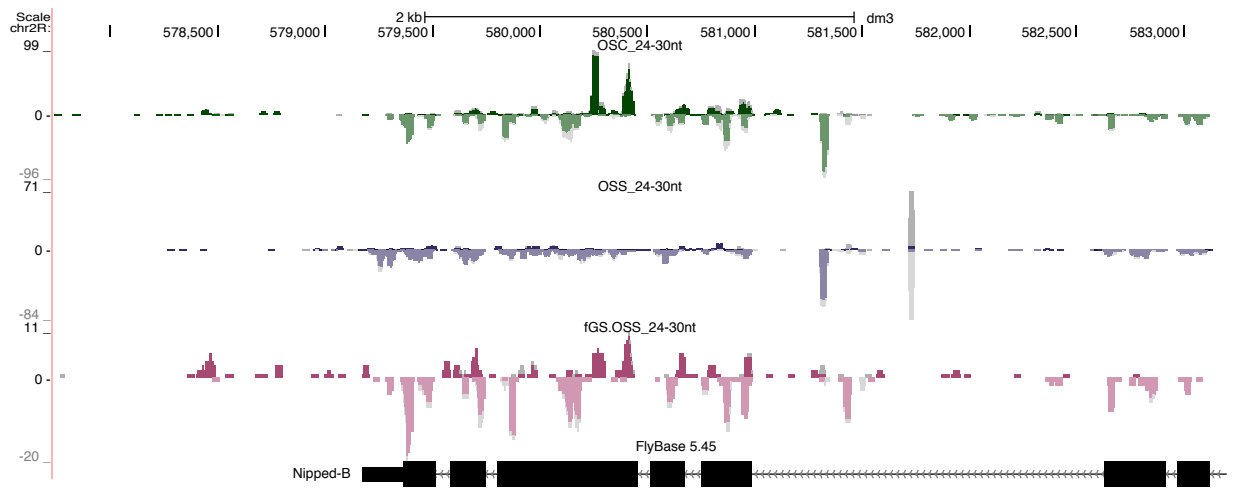
NaCP60E and pain 3' UTR overlapping produce bidirectional piRNAs



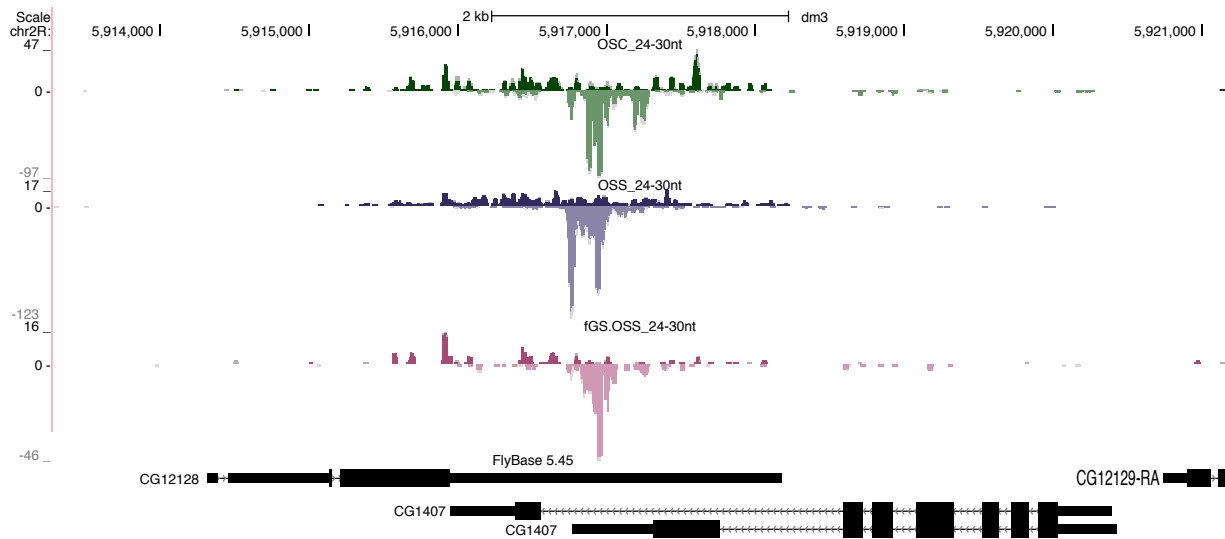
ex and crq 3' UTR overlapping produce bidirectional piRNAs



CG6178 and Myo95E 3' UTR overlapping produce bidirectional piRNAs



CG17883 and Nipped-B 3' UTR overlapping produce bidirectional piRNAs



CG1407 and CG12128 overlapping produce bidirectional piRNAs