SUPPLEMENTARY FIGURE LEGENDS for Okamura et al, "The hairpin RNA pathway generates endogenous siRNAs in Drosophila"

Figure S1. Small RNA size analysis effectively distinguishes between bona fide hairpin transcripts and fortuitous inverted repeats comprised of transcriptionally convergent or divergent gene duplicates. (A) EINVERTED hit containing divergently oriented trypsin genes generates many small RNAs; blue hits are on the (-) strand and red hits are on the (+) strand. (B) Size analysis shows a broad range of strand-selective, probable degradation fragments, suggesting that there is not a single transcript that traverses this region. (C) EINVERTED hit representing a novel tandem pair of short hairpins that we provisionally named *chou39-1* and *chou39-2*. (D) Size analysis shows that predominantly 22 nt RNAs are produced from these loci. Although these are miRNA-like, the short terminal loop suggests that they are not conventional miRNAs. Nevertheless, they appear to have been produced by some type of hairpin precursor. (E) EINVERTED hit contained within the CG18854 gene annotation; the vast majority of reads are from the (-) strand. (F) Size analysis shows a tight distribution of small RNAs centered around 21 nt.

Figure S2. Details of hp-CG18854. (A) Small RNAs were cloned predominantly from the 5' UTR and 3' UTR of the annotated CG18854, but its coding potential is limited. Genome view was provided by UCSC (http://genome.ucsc.edu/). (B) Clones mapping to the 3' UTR cluster exhibit clear phasing, although it is not as precise as with hp-CG4068 (Fig. 1B, Fig. S3). (C) The 5' UTR and 3' UTR clusters are complementary and form an extended duplex. (D) Closeup of the dominant clones mapped to one segment of the duplex illustrates clearly phased duplexes with ~2nt-3' overhangs.

Figure S3. Details of hp-CG4068. (A) A series of 20 tandem repeats are found between CG6903 and CG4068. Cloned small RNAs map to the (-) strand, and appear to initiate in the 3' UTR of CG4068. The extent of the CG4068 3' UTR has been confirmed by cDNA clones. Any individual repeat forms a hairpin (B), while any pair of repeats forms a longer hairpin (C); even longer hairpins involving additional repeats are possible. The sequences and structures of the first and the first+second hairpins in the tandem array are depicted. (D) Most of the hp-CG4068 clones map to many locations (up to 20), but the most distally-mapping clones correspond to unique locations in the CG4068 3' UTR.

Figure S4. Details of hpRNA1. A collection of seemingly heterogenous reads mapped to this hairpin (as summarized on the left schematic); the thickness of the bar is proportional to the number of reads. However, segregation of the green-highlighted clones revealed a preferred cleavage register yielding siRNA duplexes (middle schematic). Apparent phasing of the yellow-highlighted clones potentially reflects an alternative cleavage register (right schematic).

Figure S5. Details of hp-pncr009. One major siRNA duplex with canonical 2nt-3' overhangs could be mapped to its hairpin; some of these reads also map to hp-CG32207 and hp-CR32205. Another region with uniquely mapping reads exhibited phasing (i.e., adjacent-mapping reads). These and other small RNA reads that map uniquely to this hairpin provided evidence for its independent expression and processing. Additional rare clones are not depicted.

Figure S6. Details of hp-CG32207. Two major siRNA duplexes could be mapped to its hairpin; some of these reads also map to either hp-pncr009 or hp-CR32205. However, one of these siRNA reads mapped uniquely to this locus. These and other small RNA

reads that map uniquely to this hairpin provided evidence for its independent expression and processing. Additional rare clones are not depicted.

Figure S7. Details of hp-CR32205. Two major siRNA duplexes could be mapped to its hairpin; these reads also map to either hp-CG32207 and pncr009. These and other small RNA reads that map uniquely to this hairpin provided evidence for its independent expression and processing. Additional rare clones are not depicted.

Figure S8. The pncr009, hp-CG32207, hp-CR32205 loci are clustered within 70kb of each other in the chr3L:19,417,000-19,482,000 region. The CR32205 and pncr009 genes were previously annotated as putative non-coding RNAs. As seen in Figs. S5-S7, some reads map to two or three of these hpRNA loci. The schematic was generated using gbrowse at Flybase (http://flybase.bio.indiana.edu/cgi-bin/gbrowse/dmel).

Figure S9. Details of an hpRNA candidate downstream of *koi*. It contains an extended duplex segment that is in the size range of hpRNAs (and larger than that of typical pre-miRNAs). Its loop region includes a repeating element that matches to over 50 locations. However, its stem region is composed of unique sequence, and generates a small number of cloned small RNAs. The base of the stem generates eight 21nt RNAs while the top of the stem generates a 25 nt and a 26 nt RNA. While it matches minimally to our hairpin criteria, the putative siRNAs were not arranged as a typical siRNA duplex. Therefore, we tentatively designate this a candidate hpRNA locus. Perhaps its cleavage pattern might become clearer with additional clones, as with hpRNA1, for which the available clones potentially suggest dual cleavage patterns.

Figure S10. Biogenesis and structural analysis of hpRNAs. (A) Northern analysis of hp-CG4068G in S2 cells treated with various dsRNAs. Its biogenesis requirements are similar to other hpRNA-derived siRNAs, in that its accumulation is most dependent on Loqs, Dcr-2 and AGO2 (boxed regions). Note also that laddering is induced in the AGO2-knockdown cells; a minor amount of laddering is also detected in Dcr-2-knockdown cells. (B) 5' and 3' end analysis of small RNAs. Bantam migrates more quickly following ß-elimination of 3' ends, and its levels appear to be reduced as well. Conversely, bantam migrates more slowly following CIP treatment to remove 5' phosphate. hp-CG18854A and hp-CG4068G are resistant to ß-elimination, but sensitive to CIP.

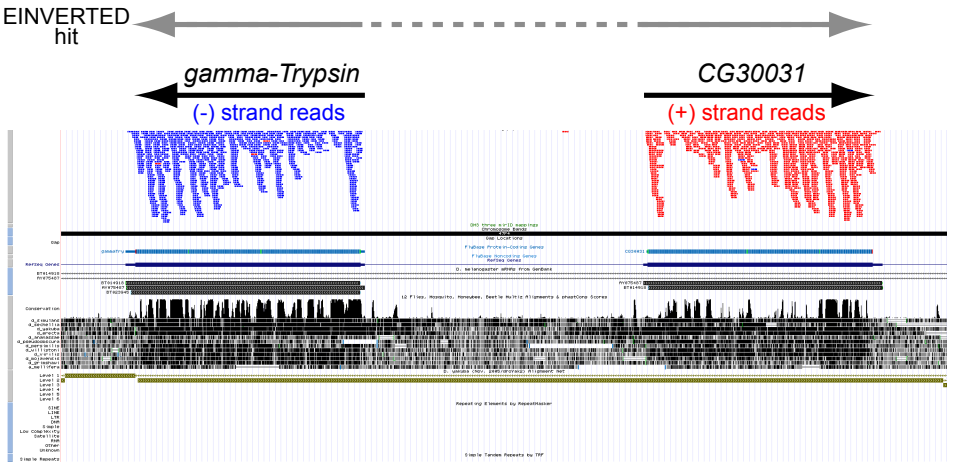
Figure S11. The 3' UTR of CG4068 does not appear to be a target of hp-CG4068. (A) A sensor containing the CG4068 3'UTR was not repressed in the presence of various hp-CG4068 expression constructs. (B) The CG4068 3' UTR sensor was not de-repressed in the presence of various ASOs.

Figure S12. Derepression of *mus308* in Dcr2- and Ago2-deficient cells. S2 cells were treated with dsRNA against GFP, Dcr-2 or Ago2. qPCR was used to measure *mus308* level in each condition, as normalized to *rp49* level. The graph depicts the mus308 level in *Dcr-2*- and *Ago2*-depleted cells, relative to *GFP* dsRNA-treated cells. These data were pooled from six experiments; standard deviations are shown.

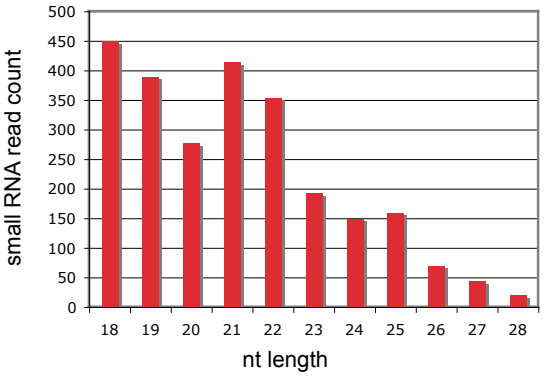
Figure S13. CG8289 may be a target of hp-CG18854. (A) hp-CG18854 is derived from CG8289. These genes are located on different chromosomes, but share significant similarity. (B) hp-CG18854 is not a simple inverted repeat of CG8289, however, as

portions of the CG8289 sequence are transposed with respect to each other in hp-CG18854. (C) Ectopic hp-CG18854 specifically repressed a CG8289:GFP fusion.

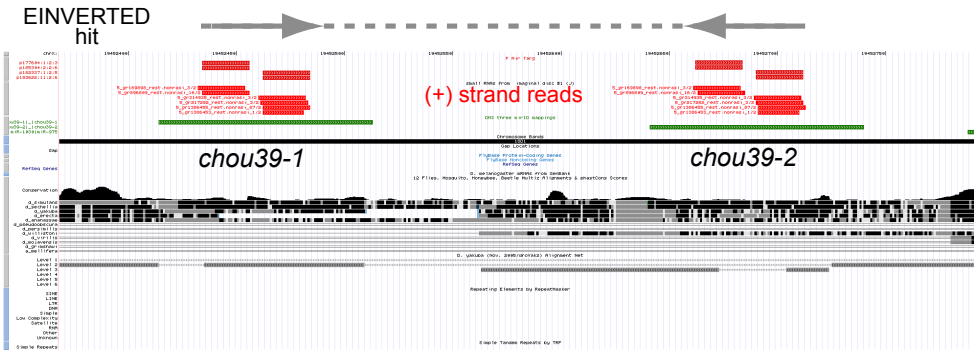
a degradation RNAs chr2R:7,235,869-7,240,000



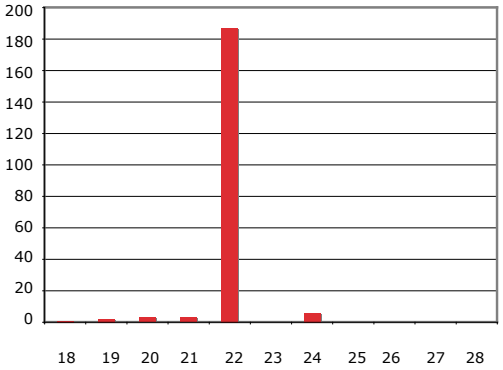
b



c novel short hairpins chrX:19,452,369-19,452,794

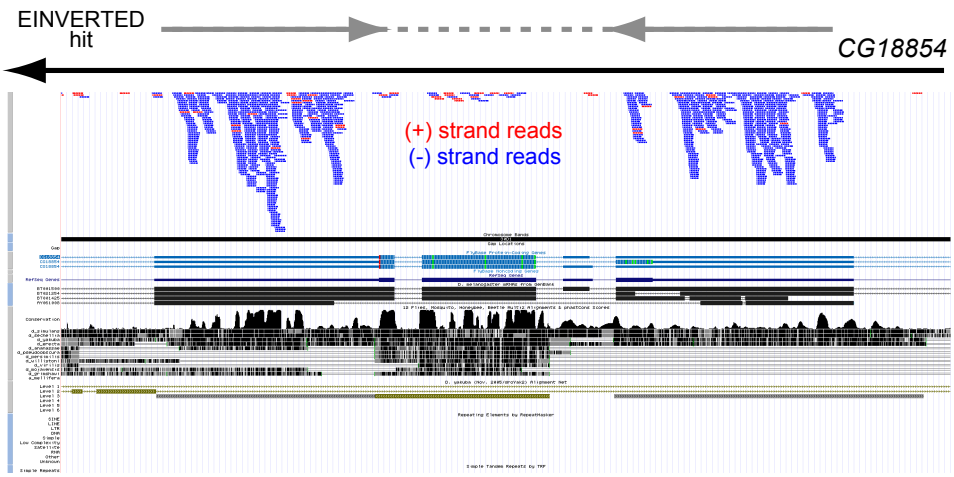


d

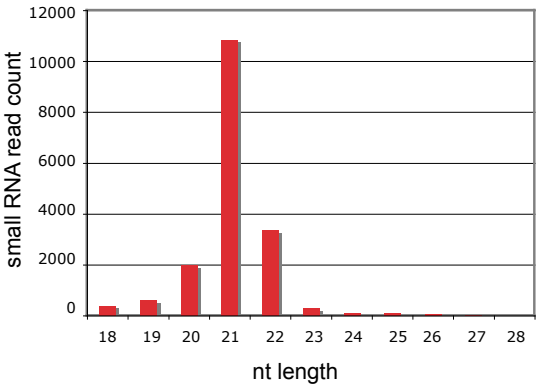


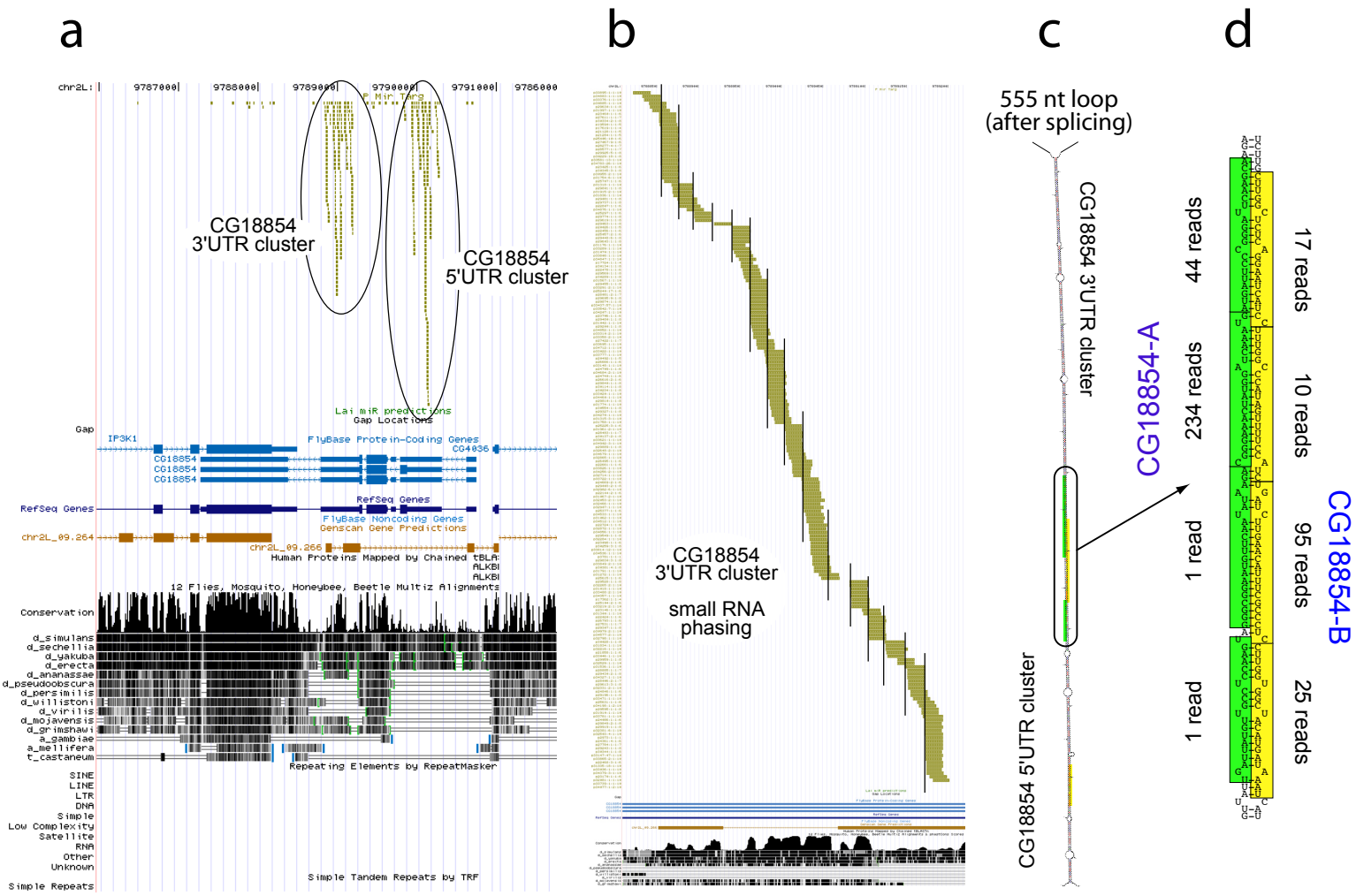
```
                                left-arm small RNA          right-arm small RNA
chou39-1 caaacuugucggcacuggagUACUAAAUAUAUUGCACAUGUUcaugcaUGUUCAACAUAUUUAGCACUCAaguaccgaaaaaugaacaacgaaacga
         .........((((.(((.((((.(((((((((.(((.(((((((....)))))))..))).)))))))).)))).))).)))).................. (-30.4)
chou39-2 uguacuugucggcacuggagUACUAAAUAUAUUGCACAUGUUcaugcaUGUUCAACAUAUUUAGCACUCAaguaccgaaaaaugaacaacuaaaagc
         (((.....((((.(((.((((.(((((((((.(((.(((((((....)))))))..))).)))))))).)))).))).)))).....))).......... (-30.5)
```

e hairpin RNAs chr2L:9,788,600-9,790,490



f

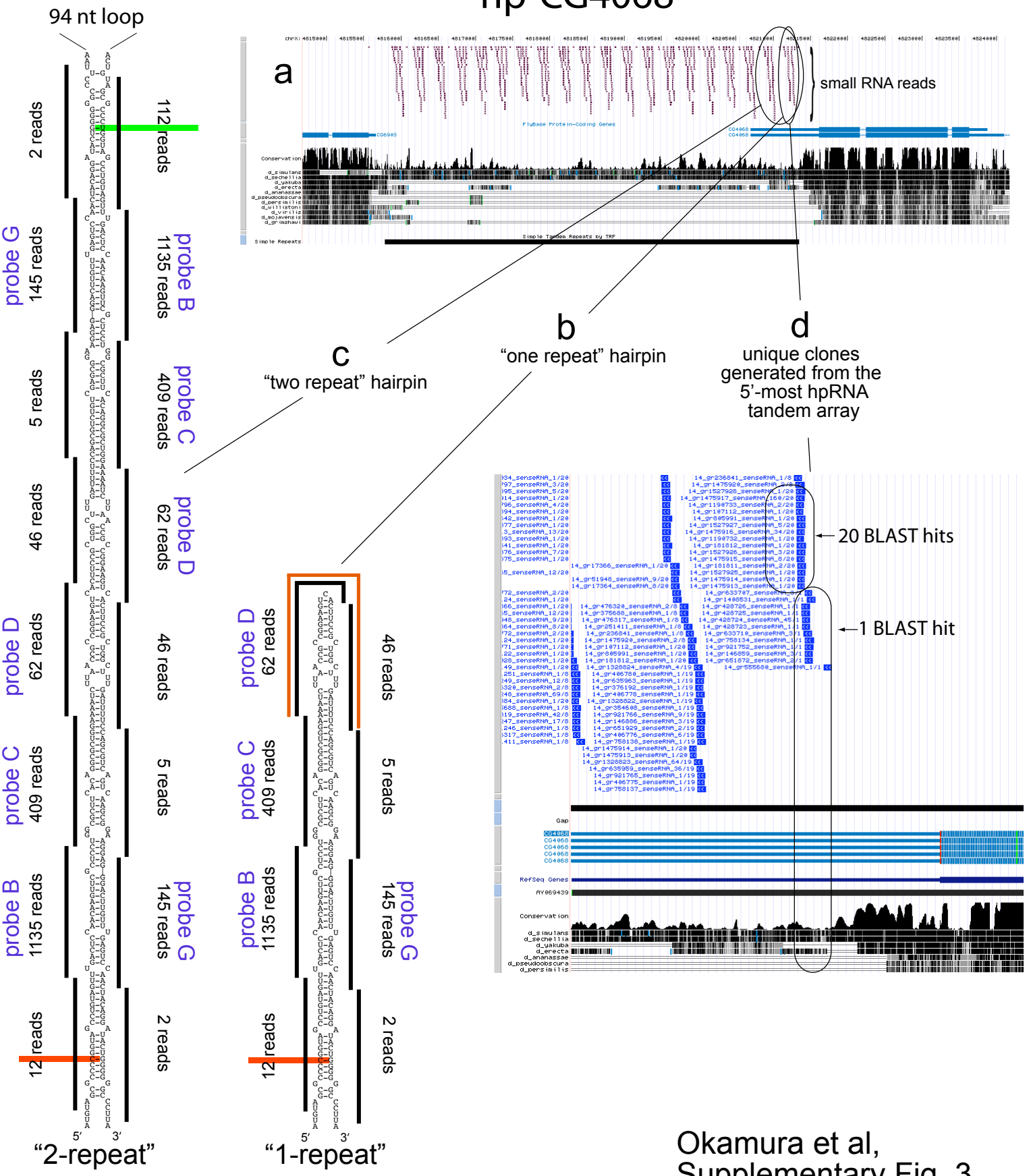




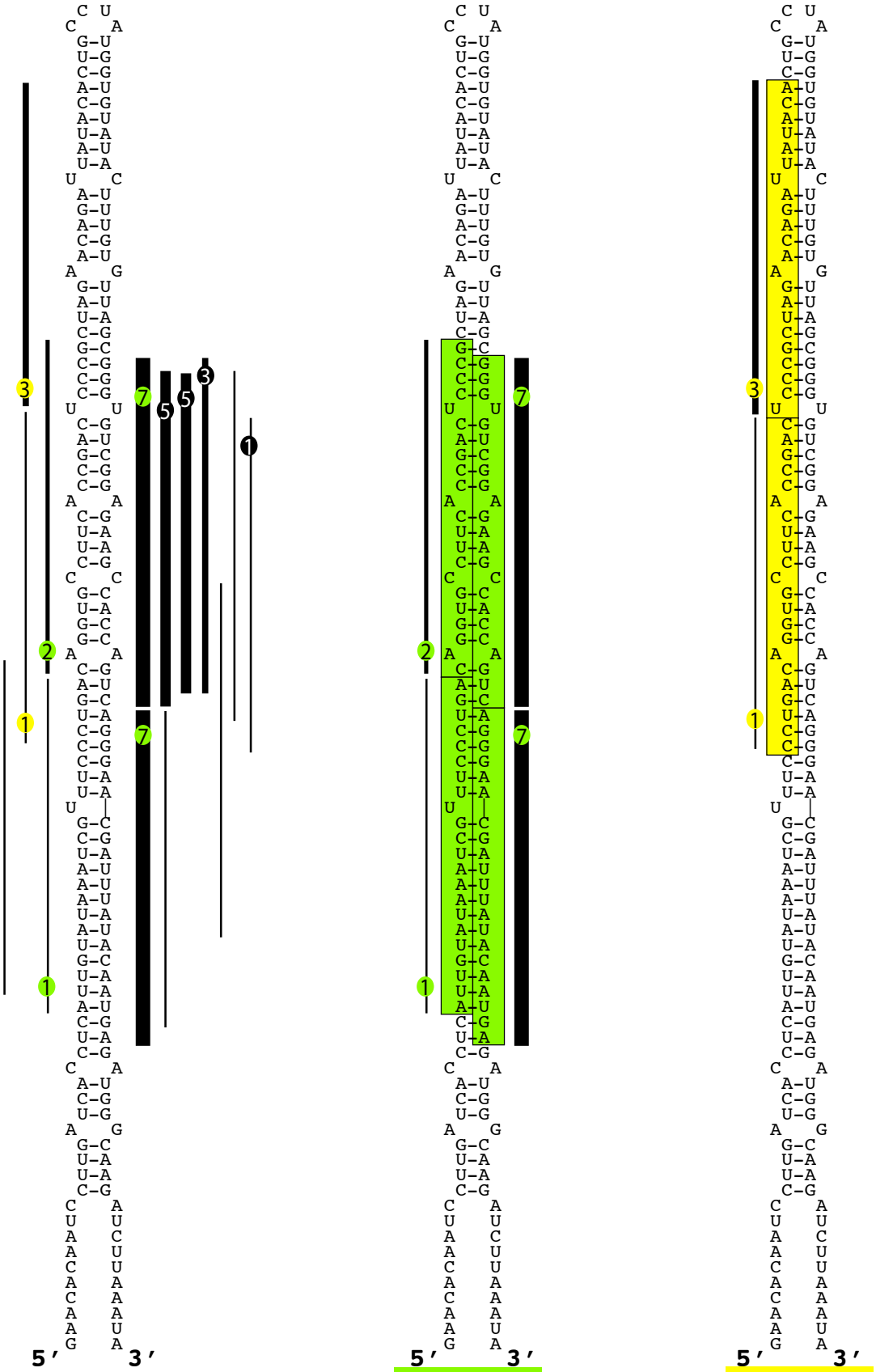
hp-CG18854

Okamura et al,
Supplementary Fig. 2

hp-CG4068



Okamura et al, Supplementary Fig. 3



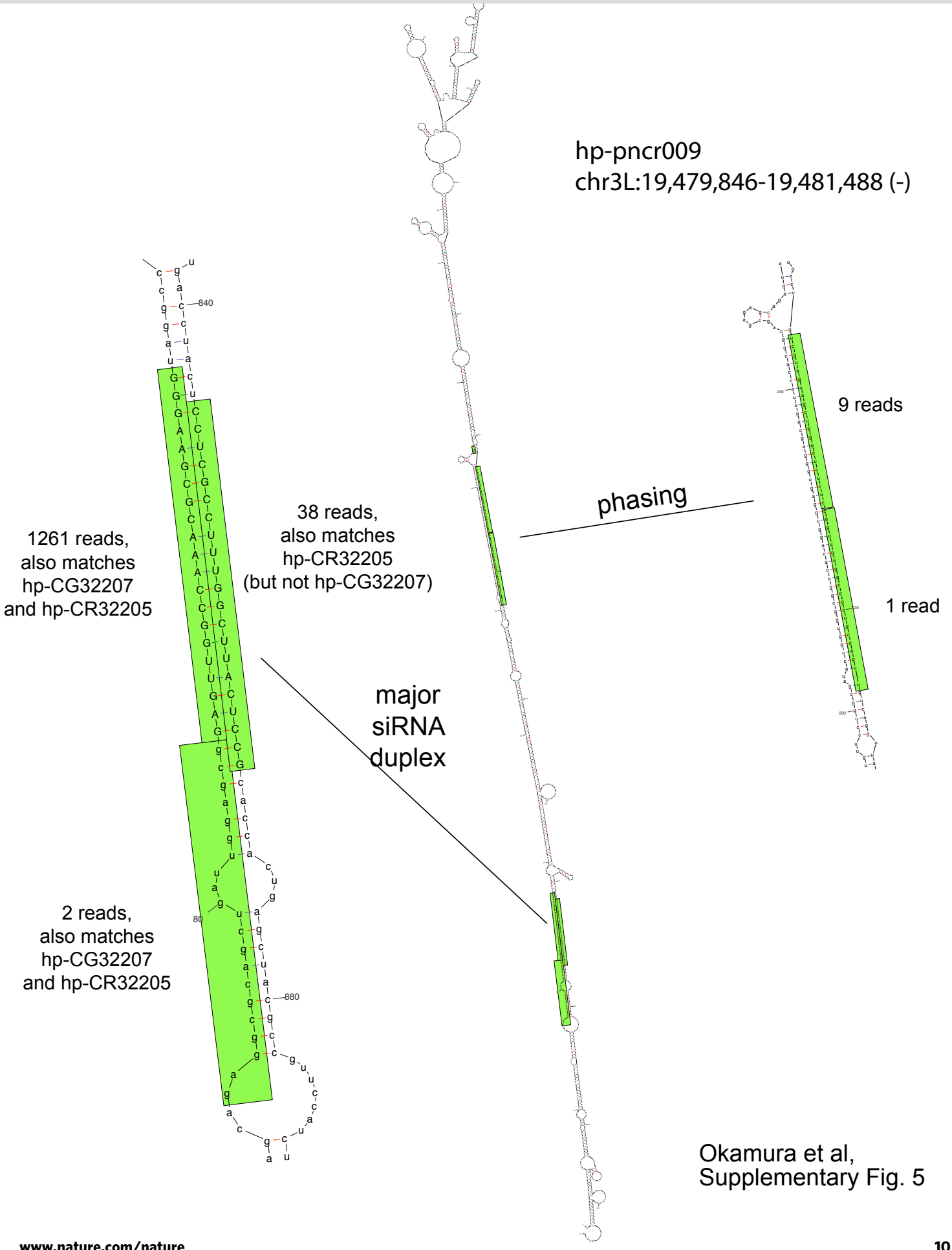
hpRNA1

chr3R:15,746,415-15,746,598 (-)
(downstream of CG4770)
unlabeled bars represent single clones

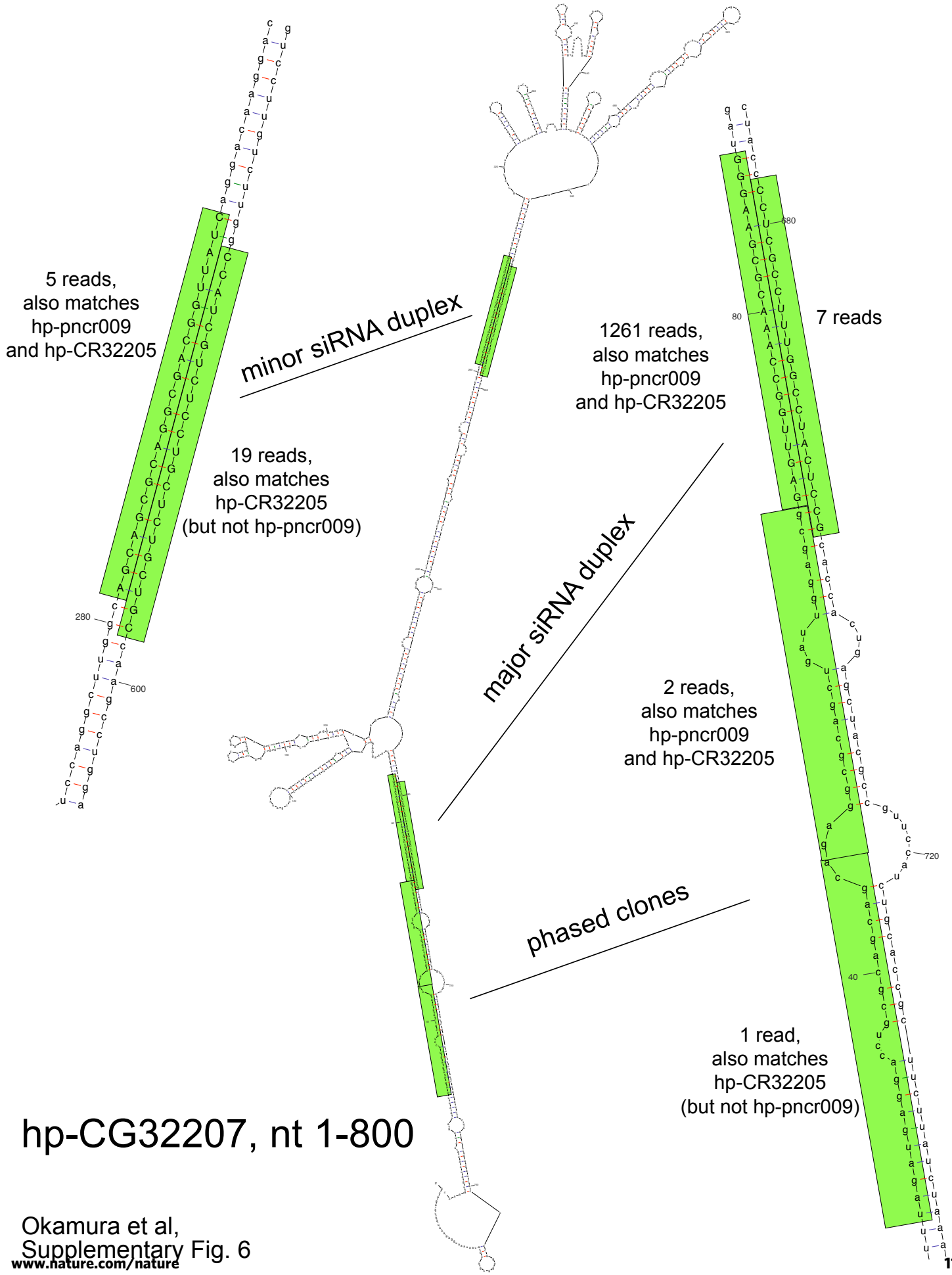
cleavage register 1

cleavage register 2?

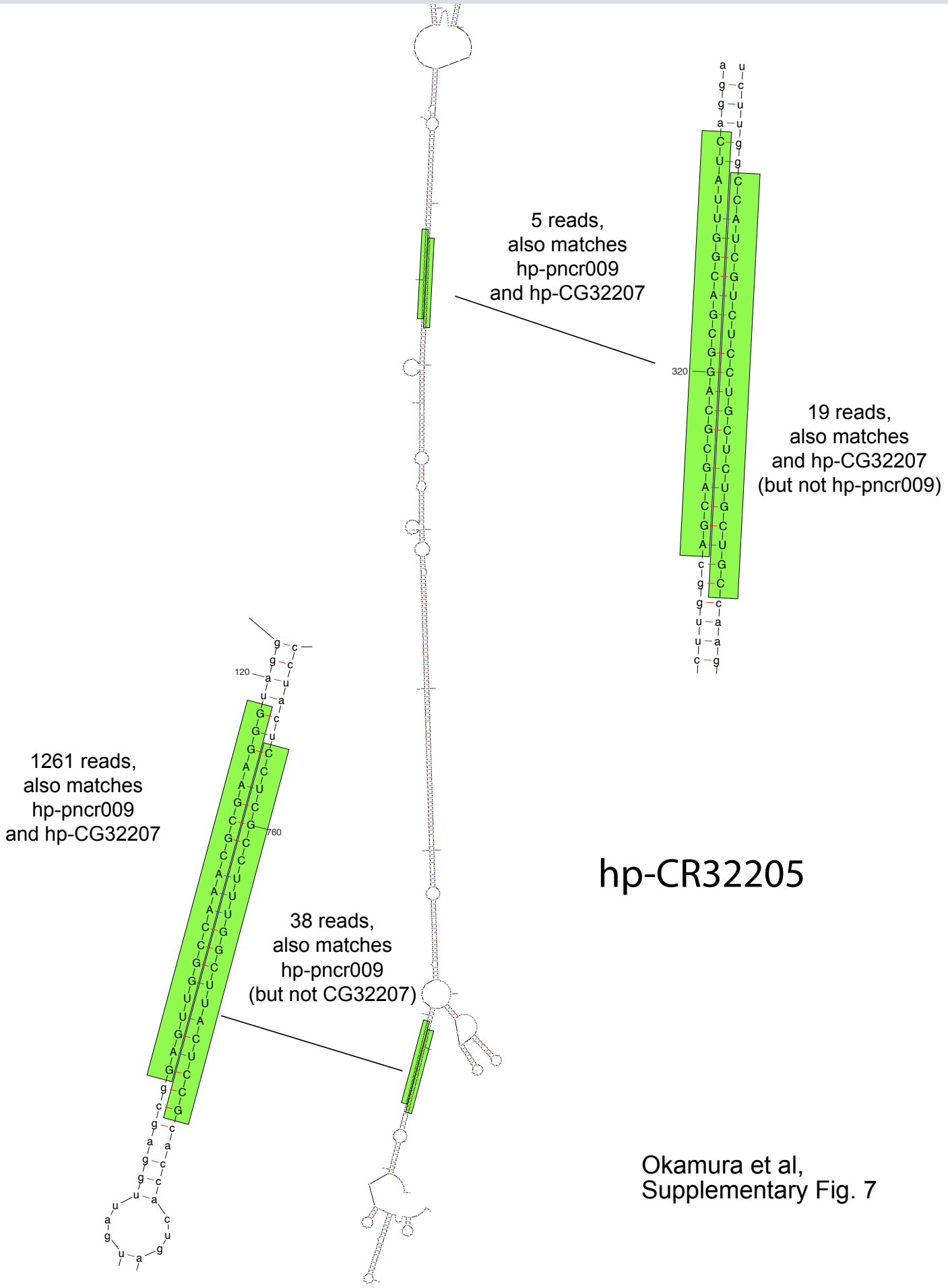
Okamura et al, Supplementary Fig. 4



Okamura et al,
Supplementary Fig. 5

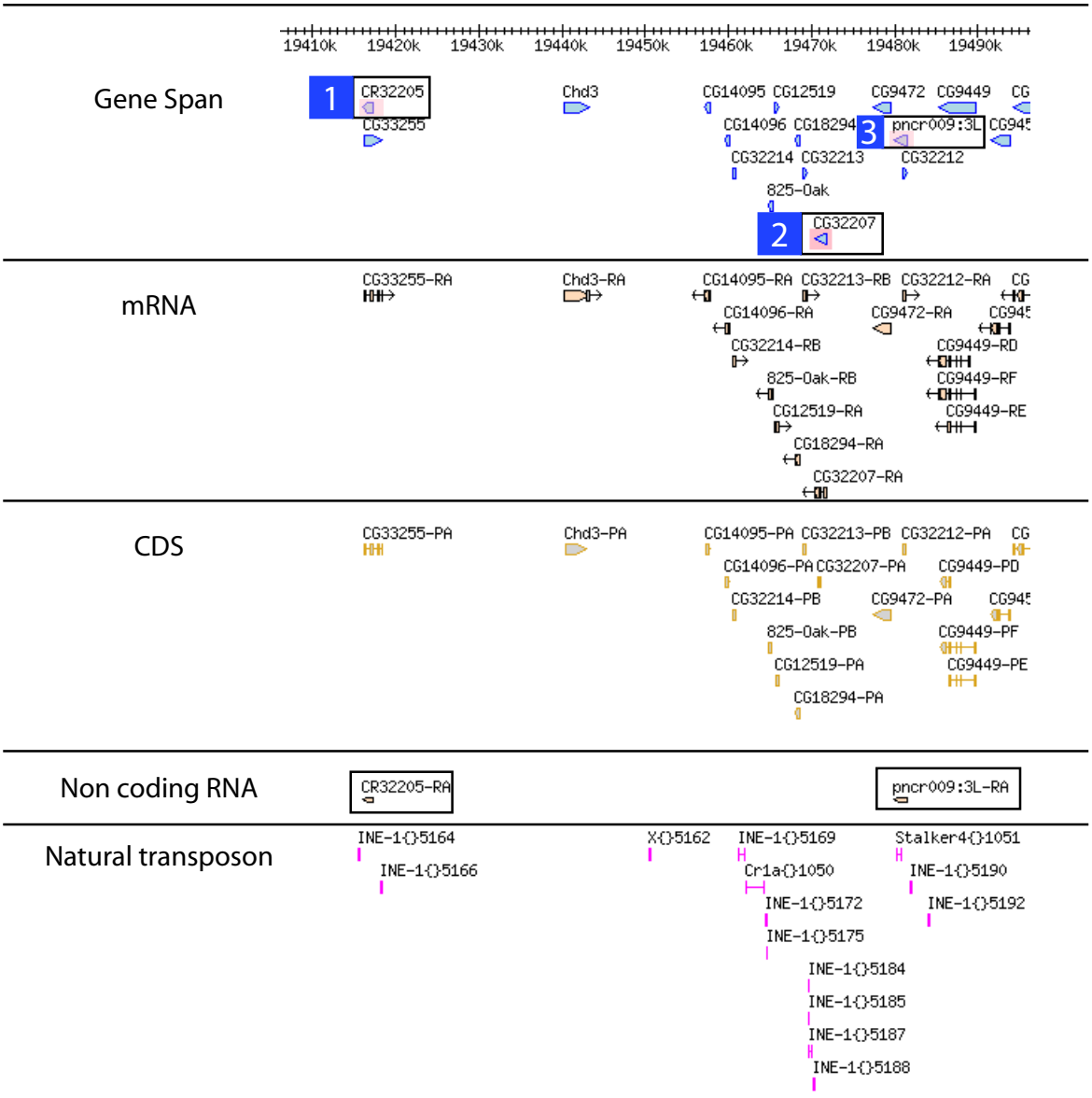


Okamura et al,
Supplementary Fig. 6

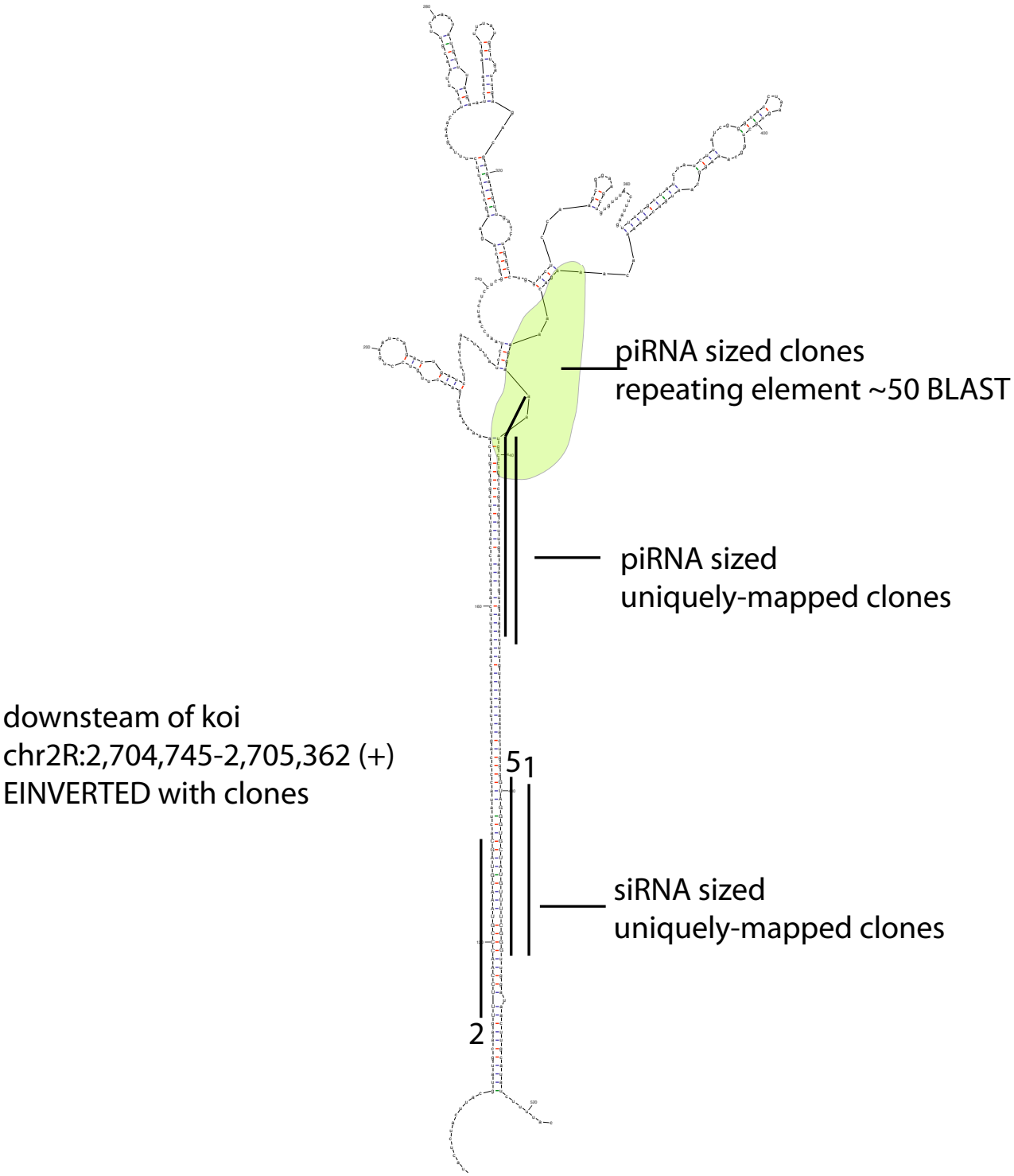


Okamura et al,
Supplementary Fig. 7

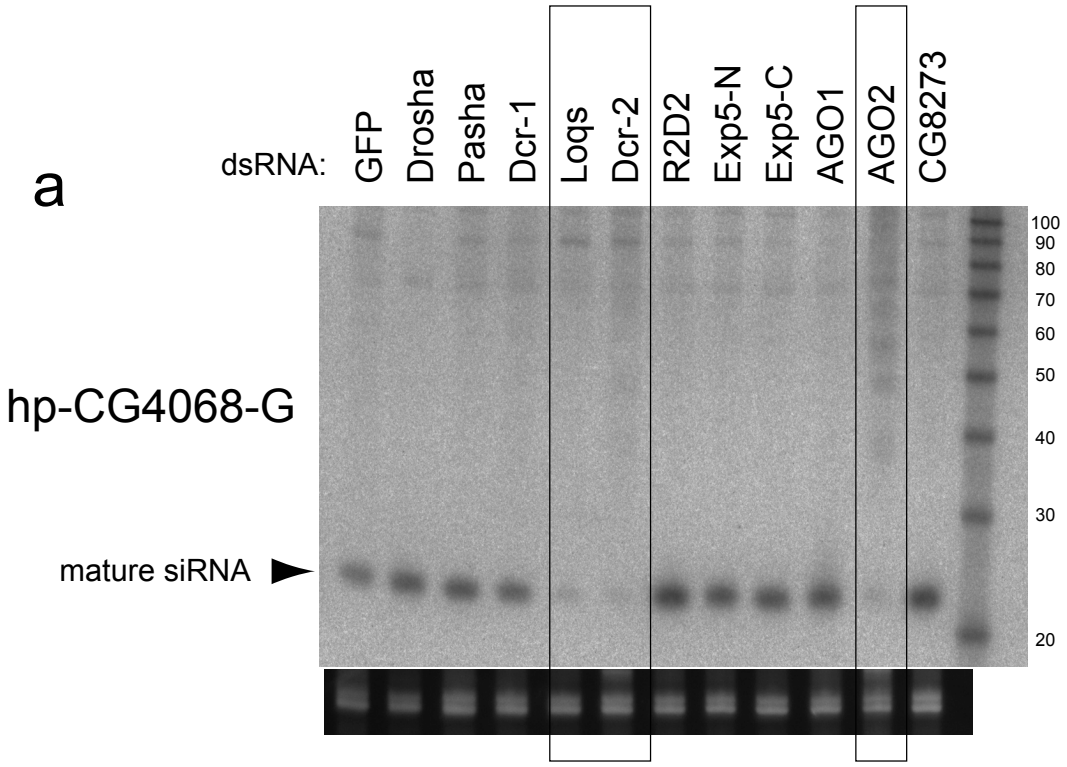
three pncr009-family hpRNA loci
in the chr3L:19,400,000-19,500,000 region

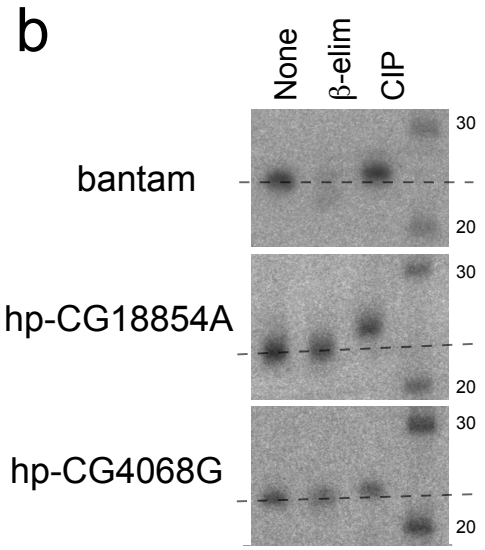


Okamura et al,
Supplementary Fig. 8

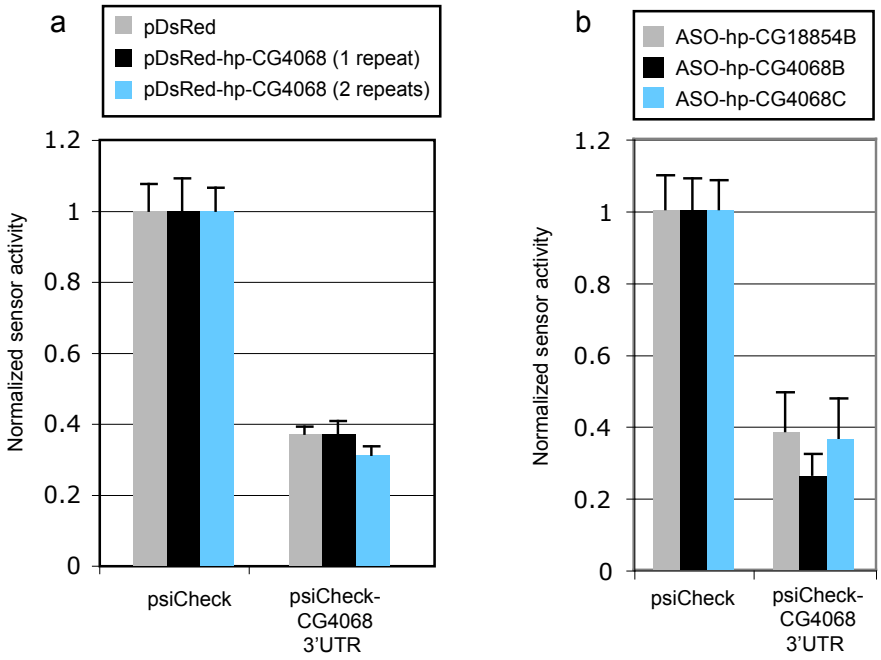


Okamura et al,
Supplementary Fig. 9

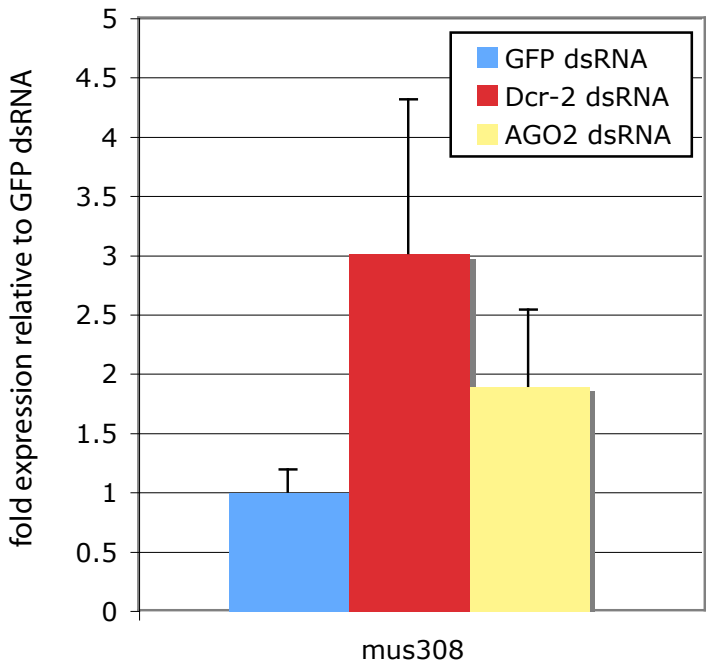




Okamura et al,
Supplementary Fig. 10

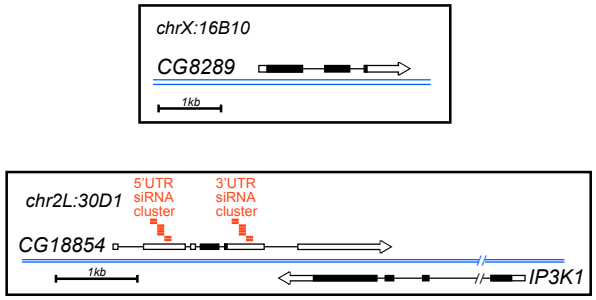


Okamura et al,
Supplementary Fig. 11

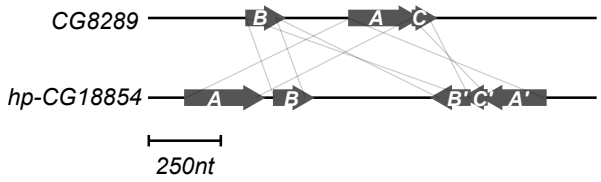


Okamura et al,
Supplementary Fig. 12

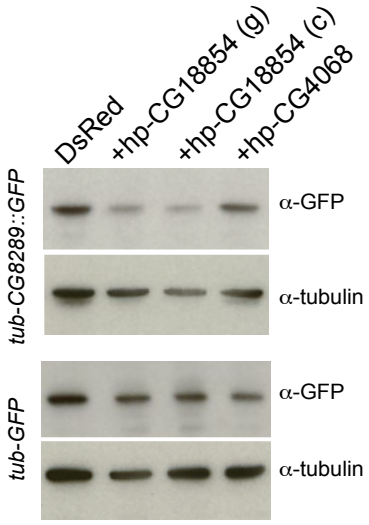
a



b



c



Okamura et al,
Supplementary Fig. 13