Evaluation of microRNA variant maturation prior to genome edition

Isabelle Busseau[1,2], Sophie Mockly[1,3], Élisabeth Houbron[1], Hedi Somaï[1] and Hervé Seitz[1,2]

1. Institut de Génétique Humaine, UMR 9002 CNRS and university of Montpellier, Montpellier, France
2. Corresponding authors; telephone: (+33)434359936; fax: (+33)434359901;
email: isabelle.busseau@igh.cnrs.fr (I.B.) and herve.seitz@igh.cnrs.fr (H.S.)
3. Current address: Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC, Canada

For the purpose of Open Access, a CC-BY public copyright licence has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission.

Distributed under a Creative Commons Attribution | 4.0 International licence: CC-BY 4.0

Abstract

Assessment of the functionality of individual microRNA/target sites is a crucial issue. Genome editing techniques should theoretically permit a fine functional exploration of such interactions, allowing the mutation of microRNAs or individual binding sites in a complete *in vivo* setting, therefore abrogating or restoring individual interactions on demand. A major limitation to this experimental strategy is the influence of microRNA sequence on its accumulation level, which introduces a confounding effect when assessing phenotypic rescue by compensatorily mutated microRNA and target site. Here we describe a simple assay to identify microRNA variants most likely to accumulate at wild-type levels even though their sequence has been mutated. In this assay, quantification of a reporter construct in cultured cells predicts the efficiency of an early biogenesis step, the Drosha-dependent cleavage of microRNA precursors, which appears to be a major determinant of microRNA accumulation in our variant collection. This system allowed the generation of a mutant *Drosophila* strain expressing a *bantam* microRNA variant at wild-type levels.

Keywords: microRNA; *bantam*; Drosha; genome editing

1 Introduction

microRNAs (miRNAs) are small regulatory RNAs which guide a repressive ribonucleoproteic complex onto specific RNA targets. miRNA targets are recognized by sequence complementarity, with most known targets in animals exhibiting a perfect complementarity to the miRNA "seed" sequence (nt 2–7), and imperfect complementarity to the rest of the miRNA sequence [Bartel, 2018]. While computational tools, and high-throughput molecular biology methods, typically predict hundreds of RNA targets for each miRNA, *in vivo* genetics usually identifies just a few key targets responsible for miRNA-controlled phenotypes [Seitz, 2009 ; Ecsedi et al., 2015 ; Title et al., 2018 ; Aeschimann et al., 2019 ; Gutiérrez-Pérez et al., 2021].

Such discrepancy between molecular biology and bioinformatics on one hand, and *in vivo* genetics on the other hand, is a central question in miRNA biology. The development of CRISPR/Cas9-mediated genome editing now offers a great opportunity to assess it rigorously, with the possibility to mutate miRNA genes, individual miRNA binding sites on target RNAs, or both of them simultaneously (offering the possibility to restore a single interaction with compensatory mutations in the miRNA and its target) [Ecsedi et al., 2015]. It can be expected that such precise genetics experiments will disprove the biological functionality of many published miRNA/target interactions, as exemplified by the *bantam*/*enabled* interaction in *Drosophila* [Becam et al., 2011 ; Bassett et al., 2014].

miRNA genes are typically transcribed as long precursors called "pri-miRNAs". In animals (see Bartel, 2018 for a review), most pri-miRNAs are cleaved in the nucleus by the Drosha endonuclease, liberating a short (≈70 nt) stem-loop-folded "pre-miRNA". The pre-miRNA is exported to the cytoplasm, where most pre-miRNAs are cleaved by the Dicer endonuclease: the cleavage reaction liberates the loop (which is quickly degraded) from the stem. The stem is composed of two small RNAs which are essentially complementary to each other. That duplex is then loaded on a protein of the Ago subfamily of the Argonaute family, and one strand is ejected and quickly degraded (see Iwakawa and Tomari, 2022 for a review). In general for a given miRNA gene, the same RNA strand tends to remain associated to Ago (that strand is called the "miRNA"), while the other strand tends to be most frequently ejected and degraded (that other strand is called the "miRNA*").

Known determinants of miRNA maturation from their pri-miRNA and pre-miRNA precursors rely on sequence and secondary structure motifs, none of them being located in the miRNA/miRNA* duplex [Han et al., 2006 ; Gu et al., 2012 ; Fang and Bartel, 2015]. But in addition to these well-defined motifs, miRNA maturation relies on the stability of the stem-loop structure, and mutations affecting the double-stranded structure of the stem can also perturb miRNA biogenesis. This dependency creates a major obstacle to the generation of useful genome-edited miRNA genes, complicating the interpretation of biological phenotypes when a mutated miRNA is not only modified in its sequence, but also in its expression level [Ecsedi et al., 2015].

Here we describe the generation of genome-edited alleles of the *bantam* miRNA in *Drosophila*, showing that miRNA seed mutation can profoundly perturb miRNA biogenesis. In that example, a simple cell culture assay can identify miRNA mutants which are inefficiently processed, therefore limiting the risk of preparing mutant animal strains with defective mutant miRNA biogenesis. We applied this method to generate a *Drosophila* strain where the *bantam* seed sequence has been mutated to another hexamer, and where mutant *bantam* accumulates at very similar levels than wild-type *bantam*.

2 Results

In an effort to generate a series of mutated alleles of the *Drosophila bantam* miRNA gene, we used CRISPR/Cas9-mediated genome editing to prepare three *Drosophila* lines where the *bantam* seed is mutated, and a fourth line where the *bantam* pre-miRNA hairpin is truncated (see Figure 1**A**). In each seed mutant, miRNA* nucleotides facing the miRNA seed nucleotides were mutated too in order to preserve pre-miRNA secondary structure.

miRNA expression was then assessed by Small RNA-Seq in L1 larvae, where *bantam* is expressed at high levels [Brennecke et al., 2003]. We analyzed larvae heterozygous for the wild-type allele to

provide an internal control. Sequencing biases between wild-type and mutant *bantam* were measured with the help of spiked-in synthetic RNA oligos exhibiting the same sequence as the wild-type and intended mutant miRNA except for a central trinucleotide tag distinguishing them from miRNAs expressed from the larvae (see Figure 1**B**).

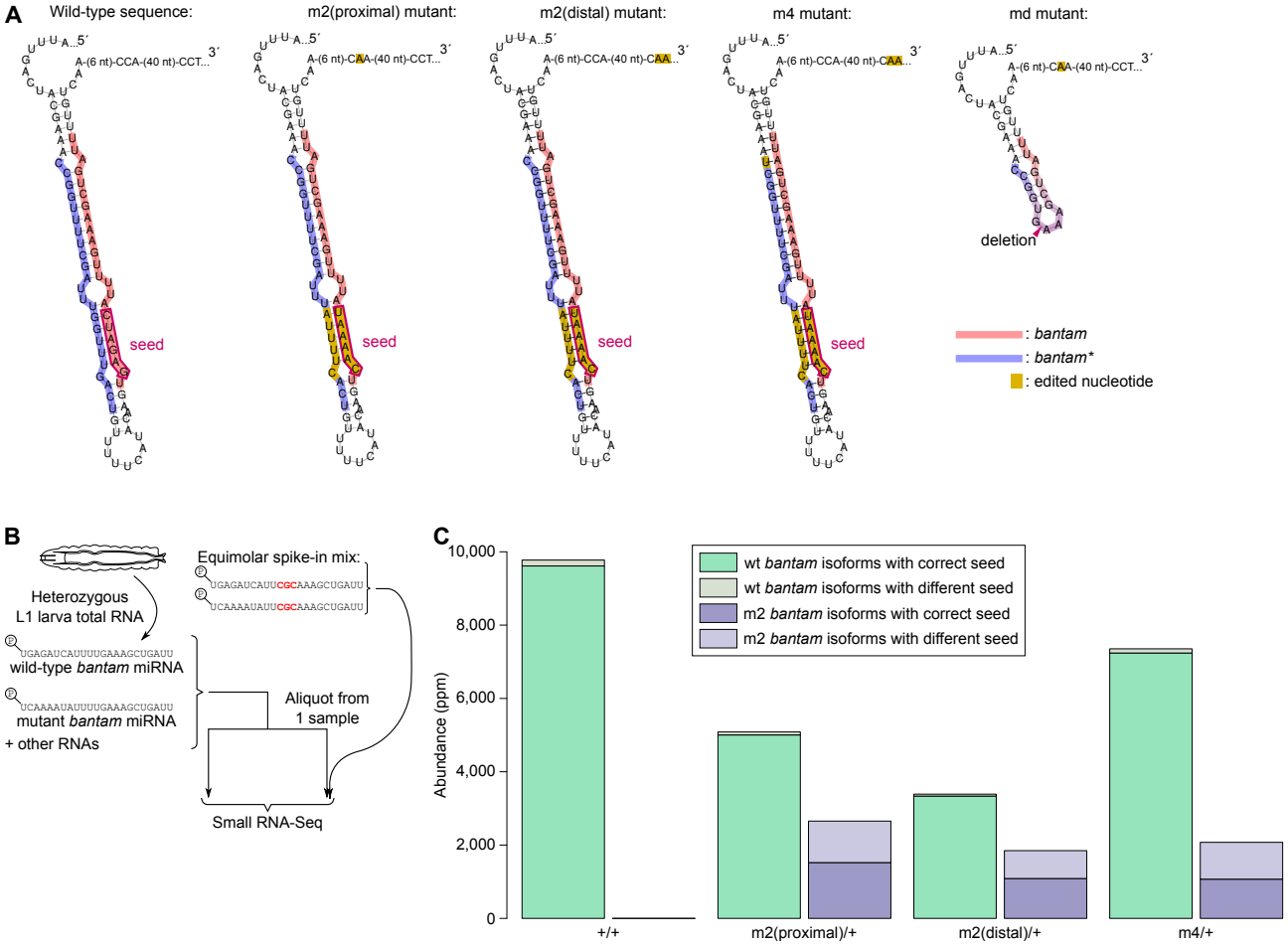


Figure 1: **Perturbed miRNA *in vivo* accumulation after seed mutation. A.** Predicted secondary structures of wild-type and mutated *bantam* loci. Mutated nucleotides are highlighted in yellow. In each construct, mutations were introduced in the miRNA and miRNA* sequences as well as in the PAM used for genome editing, to prevent cleavage of already-edited chromosomes (two distinct PAMs were used, proximal: 9 nt and distal: 52 nt downstream of the *bantam* hairpin). The seed of wild-type *bantam*, and the intended seed of the edited alleles, are framed in dark purple. **B.** Procedure for miRNA quantification. Whole-body L1 larva RNA was analyzed by Small RNA-Seq. An aliquot of one sample (the "m2(proximal)/+" sample) was diverted for sequencing bias measurement: it was supplemented with an equimolar mix of synthetic RNA oligos emulating wild-type and mutant *bantam* miRNAs (except for a trinucleotide tag, here highlighted in red) then analyzed by Small RNA-Seq. Sequencing biases between wild-type and mutant *bantam* were inferred from spike-in read counts in that library, and the observed wild-type/mutant ratio was used to correct miRNA read counts. **C.** Measured abundances of wild-type and m2 *bantam* isoforms in the 4 TM6-balanced L1 larva libraries (the "+/+" control is also heterozygous for the TM6 balancer chromosome). Wild-type isoforms "with correct seeds" are the isoforms having a GAGAUC hexamer at positions 2–7; m2 isoforms "with correct seeds" are the isoforms having a CAAAAU hexamer at positions 2–7. Read counts for m2 *bantam* with correct seed were corrected for sequencing biases using the observed spike-in read counts.

The replacement of the *bantam* seed sequence (as well as facing *bantam** sequences, and the PAM located 9 nt downstream of the stem-loop) greatly reduced miRNA accumulation (compare wt to m2 *bantam* abundances for m2(proximal)/+ heterozygotes on Figure 1**C**). It also perturbs the

balance between miRNA isoforms: while wild-type *bantam* expresses mostly isoforms sharing the same 5´ extremity, hence the same seed sequence, m2 *bantam* is expressed as a balanced mixture of isoforms with the intended CAAAAU seed and isoforms with alternative seed sequences (compare dark and pale purple on Figure 1**C**). Because it is conceivable that the PAM mutation disrupts some unknown important sequence or structure motif, we prepared another version of the m2 edited hairpin, keeping that PAM unchanged and mutating another, distant PAM. That modification did not improve mutant miRNA accumulation (see m2(distal)/+ on Figure 1**C**). In an attempt to reinforce miRNA/miRNA* asymmetry, we introduced an additional mutation, strengthening duplex base-pairing at the miRNA* 5´ end Aza-Blanc et al., 2003 ; Schwarz et al., 2003 ; Khvorova et al., 2003. While that mutation was predicted to increase miRNA accumulation, here it proved ineffective since m4 heterozygotes do not accumulate more mutant miRNA than m2 (see m4/+ on Figure 1**C**).

We reasoned that the deleterious effect of the seed mutation may derive from a destabilization of the pre-miRNA stem double-helix: while the wild-type *bantam* seed contains 3 (C+G) and 3 (A+U) nucleotides, the mutant seed contains 1 (C+G) and 5 (A+U) nucleotides. In mammals, Drosha-mediated cleavage requires the double helix to be long enough, at least because its partner DGCR8 needs to bind to a stem of sufficient length Han et al., 2006. It is therefore possible that a (A+U)-rich *bantam* seed decreases the pairing stability of the stem to the point that it is less efficiently recognized by the Drosha machinery.

We therefore aimed at designing a test to monitor the ability of candidate hairpins to be cleaved by Drosha in cultured cells, similar to the assay described by Cui et al., 2021. The insertion of the wild-type *bantam* hairpin in the 3´ UTR of a firefly luciferase reporter inhibits reporter expression in S2 cells, relatively to the insertion of the truncated "md" hairpin, which is unsuitable for Drosha processing (see Figures 2**A** and **B**). These results are in agreement with the notion that Drosha-mediated cleavage promotes degradation of the host transcript (reviewed in Lee and Shin, 2018), and provides the basis for a simple *ex vivo* assay to estimate the ability of candidate pri-miRNAs to serve as Drosha substrates.

We therefore compared 9 distinct mutant pri-miRNA sequences for their ability to repress a host luciferase reporter gene. Sequence and predicted secondary structure for these 9 variants are shown in Supplementary Figure 1. We detected a significant reduction of luciferase expression for each of these candidates compared to the md hairpin, indicating that they all undergo some level of cleavage by Drosha (see Figure 2**B**). When compared to the wild-type hairpin, every candidate was significantly less efficient at repressing their host reporter gene, except for the m6 construct (see Figure 2**B**), suggesting that the m6 allele may be matured with an efficiency similar to that of wild-type *bantam*. We also inserted *bantam* hairpin variants in a GFP host gene to assess their repressive effect *in vivo* in *Drosophila* ovaries (see Figure 2**C**). The reporter with the uncleavable md variant produced high GFP both in the germarium and developing egg chambers, while the reporter with the m2 variant produced lower level of GFP in egg chambers (see Figure 2**C**). In contrast, GFP from the reporters with wt *bantam* or the m6 variant was barely detectable both in the germarium and in developing egg chambers. These *in vivo* results mirror the *ex vivo* data, confirming that the presence of a m6 or wt hairpin in the 3´ UTR of a reporter mRNA decreases its expression to a similar extent, while the m2 variant exerts a weaker repressive effect.

These results prompted us to prepare a *Drosophila* strain where the *bantam* gene has been mutated into the m6 variant (see Figure 3**A**). Whole-body RNA from heterozygous L1 larvae was analyzed by Small RNA-Seq, using spiked-in RNA oligos for bias correction in each library. In terms of abundance of miRNA isoforms with the correct seed sequence, expression of the m6 allele is very similar to that of the wild-type allele, and we were unable to detect a significant difference between the abundances of these two miRNA sequences (see Figure 3**B**). The m2 allele, which was re-analyzed in parallel, shows a greater difference with the wild-type allele, although it also failed to reach statistical significance – likely because of the limited number of replicates in that genotype.

The availability of this allelic series allowed us to evaluate the phenotypes of homozygous edited *bantam* mutants. Both the m2 and m6 seed mutants, as well as the md deletion mutant, recapitulated the reported *bantam* null (called "Δ1") phenotypes [Hipfner et al., 2002 ; Brennecke et al., 2003]: larval growth defects (see Figure 3**C**), and larval and pupal lethality (not shown).

3 Discussion

Genome editing holds great promise for the functional assessment of individual miRNA/target interactions. Allowing the mutation of miRNA genes, or of their binding sites in target RNAs, or both of them simultaneously in a compensatory manner, this technique should allow to disrupt and restore miRNA/target interactions on demand in a complete *in vivo* setting. The analysis of physiological phenotypes controlled by individual interactions is much needed, as exemplified by the frequent misannotation of the biological roles of miRNA/target interactions when it is based on extrapolated *ex vivo* or *in silico* analyses [Mockly and Seitz, 2019 ; Mockly and Seitz, 2023].

Yet the interpretation of physiological phenotypes in such miRNA edition mutants is complicated by the influence of miRNA sequence on its accumulation level [Ecsedi et al., 2015]. It is therefore essential to generate miRNA mutant strains where not only the miRNA sequence has been edited, but its accumulation level is as close as possible to the wild-type miRNA abundance. The cell culture screening method that we describe here allowed us to identify a mutant construct which emulates wild-type miRNA expression levels for a seed-edited miRNA. Our method is easily generalizable to other miRNAs and other species.

This assay approximates Drosha cleavage efficiency by the reduction of expression of a host reporter gene. Additional mechanisms may alter reporter expression under the control of the inserted pre-miRNA hairpin sequences, therefore potentially blurring the prediction of Drosha cleavage efficiency. Also, miRNA expression levels depend on additional factors (transcription, Dicer cleavage, loading on the effector protein, post-loading stability, ...), which can also be influenced by miRNA sequence, and which are not captured by our assay.

We also observed that the mutation of the miRNA seed sequence (together with that of facing miRNA* nucleotides) affects the distribution of miRNA isoforms. While wild-type *bantam* harbors a very homogeneous 5´ end (yielding mature molecules which have mostly the same seed sequence), we note that m2 and m6 mutants tend to generate more heterogeneous 5´ ends (see Figures 1**C** and 3**B**). This result reveals a subtle dependence of Dicer cleavage site selection, of miRNA loading efficiency, or of miRNA stability, to miRNA sequence, even without any change in the predicted hairpin secondary structure. In the case of the *bantam* m6 mutant, a simultaneous increase in pan-isoform miRNA accumulation compensates for this effect, therefore achieving a highly accurate wild-type level of accumulation for the m6 isoforms with correct seed sequence. Further analyses will be required to characterize more finely such subtle consequences of sequence editing on various aspects of miRNA biogenesis and accumulation.

4 Materials and methods

Flies were raised under standard culture conditions at 25°C.

4.1 Molecular biology

Preparative PCR for plasmid construction were performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs cat. #M0491S) following the manufacturer's protocol. All constructs were verified by Sanger sequencing (Eurofins). Their annotated sequences are available at `https://github.com/HKeyHKey/Busseau_et_al_2023/tree/main/Sequences`.

4.2 Genome editing

Donor plasmids were prepared by PCR-amplifying chr3L bp 640650–643906 (dm6 assembly) from genomic DNA from *Drosophila* strain nos-Cas9 (Bloomington *Drosophila* Stock Center #54591) with primers d1319 and d1320 (see Supplementary Table 1 for oligo sequences), then cloning that product between the *Sac*I and *Xho*I sites of pBTKS-. Mutant constructs were generated by PCR-mediated mutagenesis on that plasmid, then replacing the *Age*I/*Hin*dIII fragment by the mutated fragment.

sgRNA-expressing plasmids were prepared by cloning oligo duplexes in the *Bbs*I site of plasmid pCFD3-dU6:3gRNA, which was a gift from Simon Bullock (Addgene plasmid #49410; `http:`

//n2t.net/addgene:49410; RRID:Addgene_49410; Port et al., 2014). sgRNA specificity was predicted using CRISPR targetFinder (http://targetfinder.flycrispr.neuro.brown.edu/). For constructs inserted using the PAM located 9 nt downstream of the hairpin, the cloned duplex was made by annealing oligos d1316 and d1343; for constructs inserted using the PAM located 52 nt downstream of the hairpin, the cloned duplex was made by annealing oligos d1938 and d1939.

Editions were performed following the protocols described by Port et al., 2014 ; Kane et al., 2017: donor plasmid (300 to 700 ng.µL^{-1}) and sgRNA producer plasmid (100 to 150 ng.µL^{-1}) were co-injected into nos-Cas9 embryos (Bloomington *Drosophila* Stock Center strain BL#54591) using TheBestGene service (https://www.thebestgene.com/) or in our own facility. G0 adults were then crossed to the balancer BL#23232 strain (TM6B-YFP/Dr), their progeny was screened by PCR using oligos specific for the intended mutation, and heterozygous stocks were established and maintained on the TM6B-YFP balancer chromosome from BL#54591. All edited *bantam* alleles were verified by Sanger sequencing (Eurofins).

4.3 Small RNA-Seq

Total RNA was TRIZOL-extracted (Life Technologies) from approximately 100 L1 heterozygous larvae of each genotype, hand-collected under a fluorescent binocular equiped with a GFP filter set. Samples were then depleted of S2 rRNA by Oligo Capture using biotinylated oligo d2227 and magnetic Dynabeads MyOne Streptavidin C1 (Invitrogen cat. #65001) following the protocol described by Kraus et al., 2019. Spike-in RNA oligos were then introduced in the RNA samples: for the experiment shown in Figure 1, an equimolar mix of oligos r25 and r26 was added to each RNA sample (1 µL of a mix containing 5×10^{-16} M of each oligo, per µg of total RNA). This spike-in mix was too dilute and spike-in counts were not usable for the four libraries. Another library was then prepared using an aliquot of the "m2(proximal)/+" RNA sample supplemented with a more concentrated oligo mix (1 µL of a mix containing 1.5×10^{-10} M of each oligo, per µg of total RNA), and its spike-in counts were used to correct for sequencing biases in the first four libraries.

For the experiment shown in Figure 3, an equimolar mix of RNA oligos r25, r26 and r64 was added to each RNA sample (1 µL of a mix containing 1.5×10^{-10} M of each oligo, per µg of total RNA).

Small RNA-Seq libraries were prepared by the MGX sequencing facility using the NEXTflex Small RNA-Seq Kit v3 (Bioo scientific), with pre-adenylated 3´ adapter sequence being NNNNTGGAATTCTCGGGTGCAAGG. Libraries were quantified on a Fragment Analyzer (Agilent) using kit "High sensitivity NGS" and by qPCR on a Roche LightCycler 480.

3´ adapter sequences were trimmed, and reads 18–30 nt long were selected, using cutadapt. Adapter-trimmed reads were mapped by Hisat2 on an index composed of the *Drosophila* dm6 genome, supplemented with sequences of the spiked-in RNA oligos as well as the sequences of the edited *bantam* hairpins. Full-length spike-in sequence occurences were counted in the cutadapt-trimmed fastq files, while wild-type and mutant *bantam* read counts were extracted from the hisat2 mapping output. Read counts were normalized to the total number of small RNA reads mapping on the hisat2 index.

4.4 Luciferase reporter assays

Reporter plasmids containing wild-type, md, m1, m2, m3 or m4 *bantam* variants were constructed by cloning the hairpin of interest (PCR-amplified with oligos d2289 and d2290 from the existing donor DNA, see above) into the *Sac*I and *Xba*I sites of the pmirGLO vector (Promega cat. #E1330).

Reporter plasmids containing m5, m6, m7, m8 or m9 variants were constructed using NEBuilder HiFi DNA Assembly (NEB cat. #E2621) by assembling a PCR fragment amplified using oligos d2436 and d2437 from the reporter plasmid pmirGLO containing wild-type *bantam*, together with oligo duplexes made by annealing oligos d2438 and d2439 (m5), d2440 and d2441 (m6), d2442 and d2443 (m7), d2444 and d2445 (m8), or d2446 and d2447 (m9).

S2R+ cells from DGRC were cultivated under standard conditions at 25°C in Schneider medium with 10% Foetal Bovine Serum and 1% penicillin-streptomycin. Cells were seeded at 3×10^5 cells.well^{-1} in 6-well plates and transfected 24h later with one pmiRGLO derivative construct per well (0.9 µg plasmid) using Effectene (Qiagen cat. #301425). After 3 days the cells were washed with PBS

and lysed with Cell Culture Lysis Buffer (Promega cat. #E1531) prior to Firefly and Renilla luciferase activity measurements which were performed in triplicate using the Dual-GLO kit (Promega cat. #E2920). The whole procedure was repeated three times independently for each construct.

4.5 *In vivo* GFP reporter assays

The GFP reporter plasmid containing a wild-type *bantam* hairpin in its 3´ UTR was constructed as follows: a GFP-*bantam* fragment was amplified using oligos d1360 and d1361 from the transgenic fly strain containing UAS-banD [Brennecke et al., 2003], digested with *Bam*H1 and *Xba*1, and ligated together with a *Xba*I-*Nde*I fragment containing the SV40 3´ UTR from vector pUASt-attB [Bischof et al., 2007] into *Bam*H1-*Nde*I-digested vector pUASpK10 [Koch et al., 2009]. GFP reporter plasmids containing md, m2 and m6 were generated by replacing the *Mlu*1-*Xba*1 fragment containing the *bantam* hairpin by a *Mlu*I-*Xba*I-digested PCR fragment generated using oligos d1379 and d1380 from the corresponding existing donor plasmids.

PhiC31 integrase-mediated transgenesis was performed by the BestGene facility. Transgenes were targetted into the attP2 docking site of stock BL#8622.

Tissue-specific expression of the GFP-reporter transgenes in the female germline was induced using the nos-Gal4-VP16 driver [Rorth, 1998]. Ovaries were hand-dissected from 3 day-old females maintained on medium supplemented with dry yeast, fixed in 4% PFA and stained with 1 µg.mL^{-1} DAPI following standard procedures, then mounted in Vectashield. Images were captured with a Zeiss Axioimager Apotome microscope.

4.6 Data and script availability

Small RNA-Seq data has been deposited at NCBI's SRA (`https://www.ncbi.nlm.nih.gov/sra/`) under accession numbers SUB8281052 (for Figure 1) and SUB11510745 (for Figure 3). Scripts and intermediary files are accessible at `https://github.com/HKeyHKey/Busseau_et_al_2023`.

5 Acknowledgments

We are grateful to Simon Nicot for guidance in cell culture, Lara Demont for help in the screening of edited flies, Laly Robledo for her implication in the setting of Luciferase assays, and Guillaume Canal for his help in optimizing spike-in oligo sequences. We acknowledge the MRI imaging facility, member of the France-BioImaging infrastructure, and the MGX sequencing facility, member of the France Génomique National Infrastructure, supported respectively by contracts ANR-10-INBS-04 and ANR-10-INBS-09 «Investments for the future" of French National Research Agency, as well as the Biocampus *Drosophila* facility. Stocks obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) were used in this study. This work was funded by the CNRS.

REFERENCES

Aeschimann, F., Neagu, A., Rausch, M., and Großhans, H. (2019). *let-7* coordinates the transition to adulthood through a single primary and four secondary targets. *Life Sci Alliance*, 2(2):e201900335. DOI:10.26508/lsa.201900335.

Aza-Blanc, P., Cooper, C. L., Wagner, K., Batalov, S., Deveraux, Q. L., and Cooke, M. P. (2003). Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol Cell*, 12(3):627–637. DOI:10.1016/s1097-2765(03)00348-4.

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell*, 173(1):20–51. DOI:10.1016/j.cell.2018.03.006.

Bassett, A. R., Azzam, G., Wheatley, L., Tibbit, C., Rajakumar, T., McGowan, S., Stanger, N., Ewels, P. A., Taylor, S., Ponting, C. P., Liu, J. L., Sauka-Spengler, T., and Fulga, T. A. (2014). Understanding functional miRNA-target interactions in vivo by site-specific genome engineering. *Nat Commun*, 5:4640. DOI:10.1038/ncomms5640.

Becam, I., Rafel, N., Hong, X., Cohen, S. M., and Milán, M. (2011). Notch-mediated repression of bantam miRNA contributes to boundary formation in the *Drosophila* wing. *Development*, 138(17):3781–3789. DOI:10.1242/dev.064774.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci USA*, 104(9):3312–3317. DOI:10.1073/pnas.0611511104.

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., and Cohen, S. M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell*, 113(1):25–36. DOI:10.1016/s0092-8674(03)00231-9.

Cui, Y., Lyu, X., Ding, L., Ke, L., Yang, D., Pirouz, M., Qi, Y., Ong, J., Gao, G., Du, P., and Gregory, R. I. (2021). Global miRNA dosage control of embryonic germ layer specification. *Nature*, 593(7860):602–606. DOI:10.1038/s41586-021-03524-0.

Ecsedi, M., Rausch, M., and Großhans, H. (2015). The *let-7* microRNA directs vulval development through a single target. *Dev Cell*, 32(3):335–344. DOI:10.1016/j.devcel.2014.12.018.

Fang, W. and Bartel, D. P. (2015). The menu of features that define primary microRNAs and enable de novo design of microRNA genes. *Mol Cell*, 60(1):131–145. DOI:10.1016/j.molcel.2015.08.015.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P. N., and Kay, M. A. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of Dicer processing in vivo. *Cell*, 151(4):900–911. DOI:10.1016/j.cell.2012.09.042.

Gutiérrez-Pérez, P., Santillán, E. M., Lendl, T., Wang, J., Schrempf, A., Steinacker, T. L., Asparuhova, M., Brandstetter, M., Haselbach, D., and Cochella, L. (2021). miR-1 sustains muscle physiology by controlling V-ATPase complex assembly. *Sci Adv*, 7(42):eabh1434. DOI:10.1126/sciadv.abh1434.

Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T., and Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5):887–901. DOI:10.1016/j.cell.2006.03.043.

Hipfner, D. R., Weigmann, K., and Cohen, S. M. (2002). The *bantam* gene regulates *Drosophila* growth. *Genetics*, 161(4):1527–1537. DOI:10.1093/genetics/161.4.1527.

Iwakawa, H. O. and Tomari, Y. (2022). Life of RISC: formation, action, and degradation of RNA-induced silencing complex. *Mol Cell*, 82(1):30–43. DOI:10.1016/j.molcel.2021.11.026.

Kane, N. S., Vora, M., Varre, K. J., and Padgett, R. W. (2017). Efficient screening of CRISPR/Cas9-induced events in *Drosophila* using a co-CRISPR strategy. *G3 (Bethesda)*, 7(1):87–93. DOI:10.1534/g3.116.036723.

Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115(2):209–216. DOI:10.1016/s0092-8674(03)00801-8.

Koch, R., Ledermann, R., Urwyler, O., Heller, M., and Suter, B. (2009). Systematic functional analysis of Bicaudal-D serine phosphorylation and intragenic suppression of a female sterile allele of BicD. *PLoS One*, 4(2):e4552. DOI:10.1371/journal.pone.0004552.

Kraus, A. J., Brink, B. G., and Siegel, T. N. (2019). Efficient and specific oligo-based depletion of rRNA. *Sci Rep*, 9(1):12281. DOI:10.1038/s41598-019-48692-2.

Lee, D. and Shin, C. (2018). Emerging roles of DROSHA beyond primary microRNA processing. *RNA Biol*, 15(2):186–193. DOI:10.1080/15476286.2017.1405210.

Mockly, S. and Seitz, H. (2019). Inconsistencies and limitations of current microRNA target identification methods. *Methods Mol Biol*, 1970:291–314. DOI:10.1007/978-1-4939-9207-2_16.

Mockly, S. and Seitz, H. (2023). Synthetic mir-34a against solid tumors: a predictable failure. *British Journal of Cancer*. In press.

Port, F., Chen, H. M., Lee, T., and Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci USA*, 111(29):E2967–E2976. DOI:10.1073/pnas.1405500111.

Rorth, P. (1998). Gal4 in the *Drosophila* female germline. *Mech Dev*, 78(1-2):113–118. DOI:10.1016/s0925-4773(98)00157-9.

Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115(2):199–208. DOI:10.1016/s0092-8674(03)00759-1.

Seitz, H. (2009). Redefining microRNA targets. *Curr Biol*, 19(10):870–873. DOI:10.1016/j.cub.2009.03.059.

Title, A. C., Hong, S. J., Pires, N. D., Hasenöhrl, L., Godbersen, S., Stokar-Regenscheit, N., Bartel, D. P., and Stoffel, M. (2018). Genetic dissection of the miR-200-Zeb1 axis reveals its importance in tumor differentiation and invasion. *Nat Commun*, 9(1):4671. DOI:10.1038/s41467-018-07130-z.

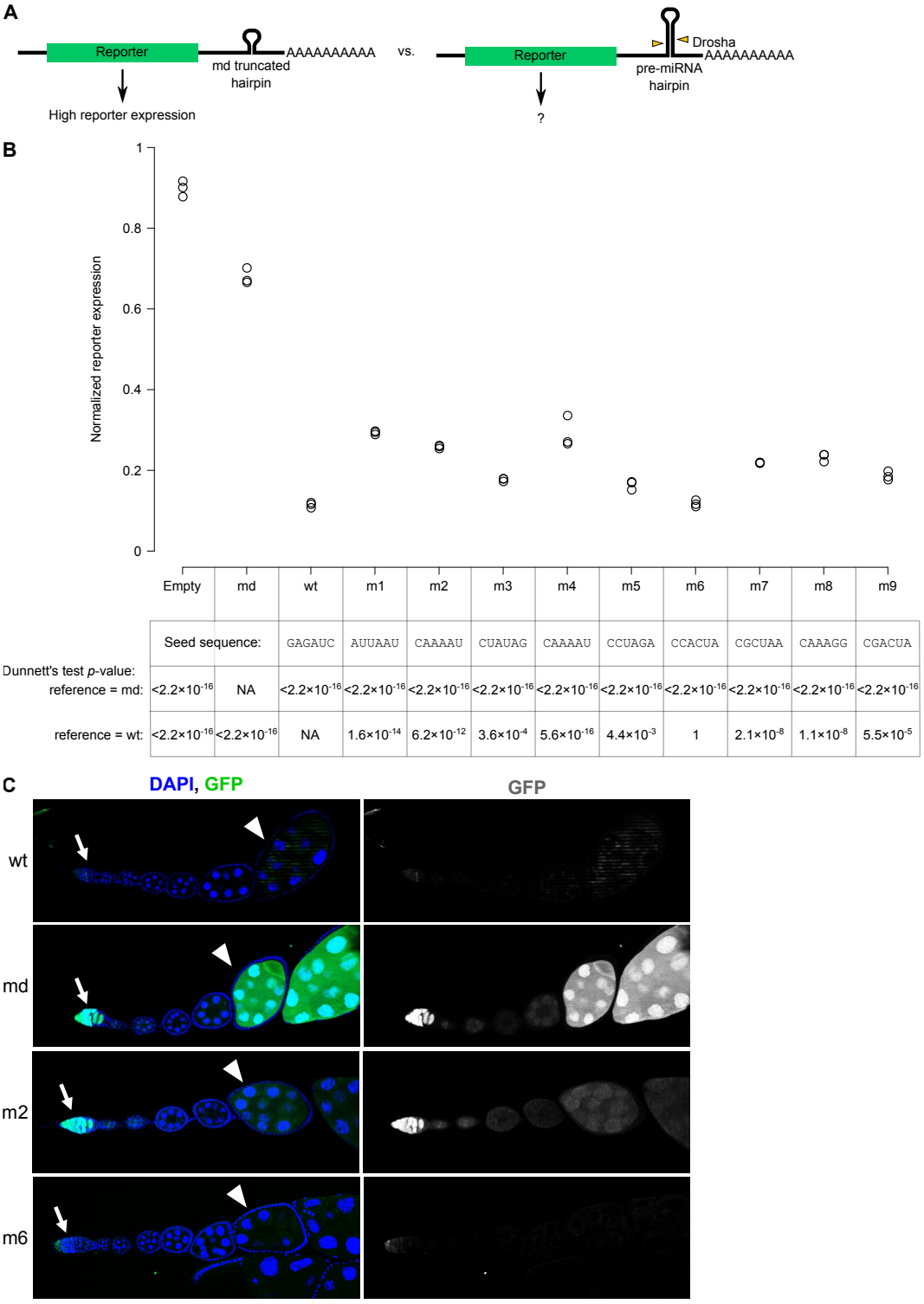


Figure 2: ***Ex vivo* screen for an estimation of Drosha substrate suitability. A.** Principle of the assay: insertion of a candidate pre-miRNA hairpin in the reporter's 3´ UTR attracts Drosha cleavage more or less efficiently, resulting in various levels of reporter expression. Efficiently processed hairpins induce a more degradation of the host transcript, hence a lower reporter expression. **B.** Candidate hairpins were inserted in the 3´ UTR of a firefly luciferase reporter, expressed from the same plasmid than a Renilla luciferase internal control. For each construct, the sequence of its mutated seed is given in the table under the graph. After subtraction of luciferase signals measured in untransfected cells, the firefly/Renilla luciferase activity ratio reveals the repressive action of inserted hairpins on firefly expression. The experiment was performed in three biological replicates, and firefly/Renilla luciferase ratios were compared to the ratio measured in two reference transfections (with an inserted "md" truncated hairpin, and with a wild-type *bantam* hairpin); Dunnett's test *p*-values are indicated in the table under the graph. **C.** Candidate hairpins corresponding to *bantam* wt, md, m2 and m6 were inserted in the 3´ UTR of a a UASp-GFP reporter. Expression was driven in the female germline using nos-GAL4-VP16, whose two peaks of activity reside in the germ cells of the germarium (arrows) and the developing stage 6-7 egg chambers (arrowheads).

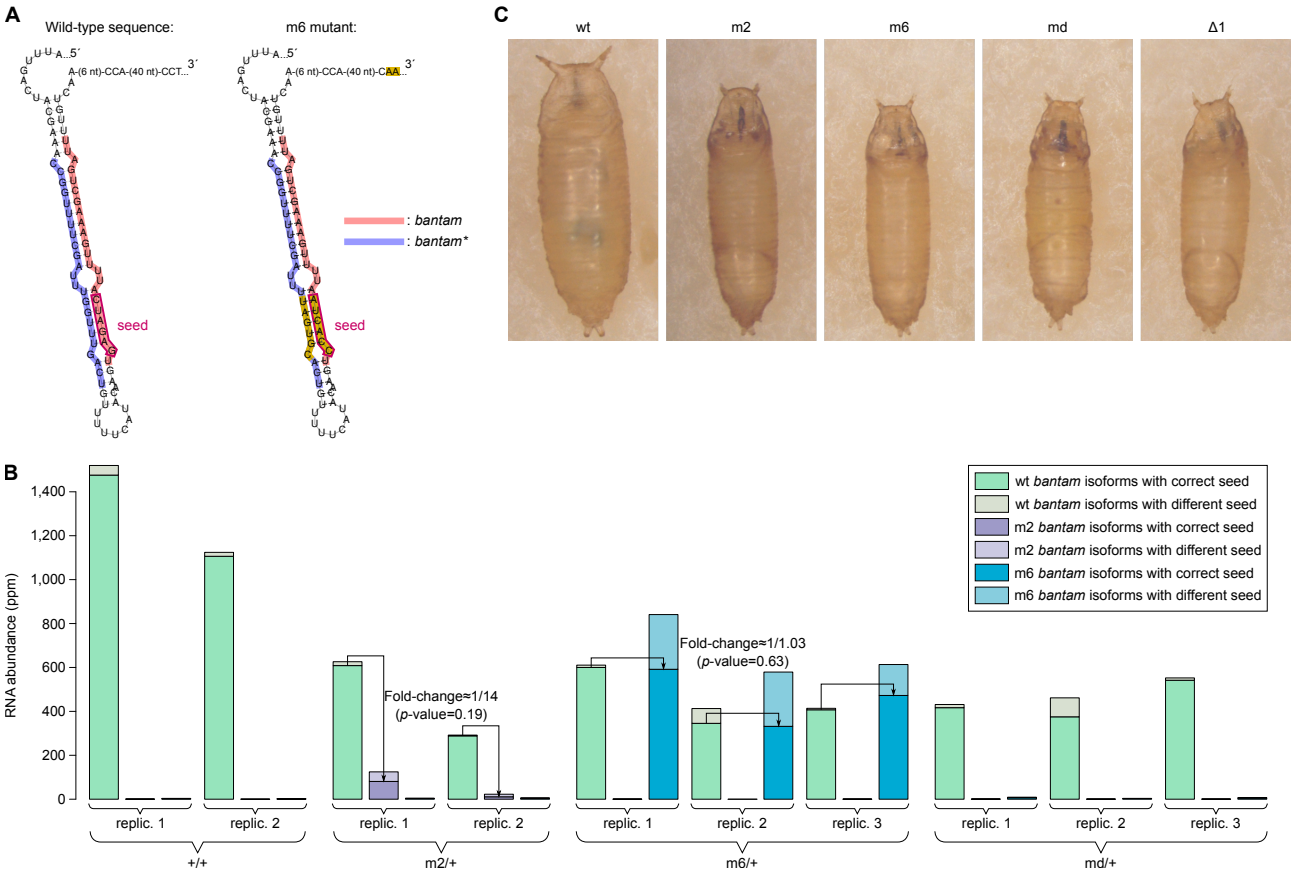
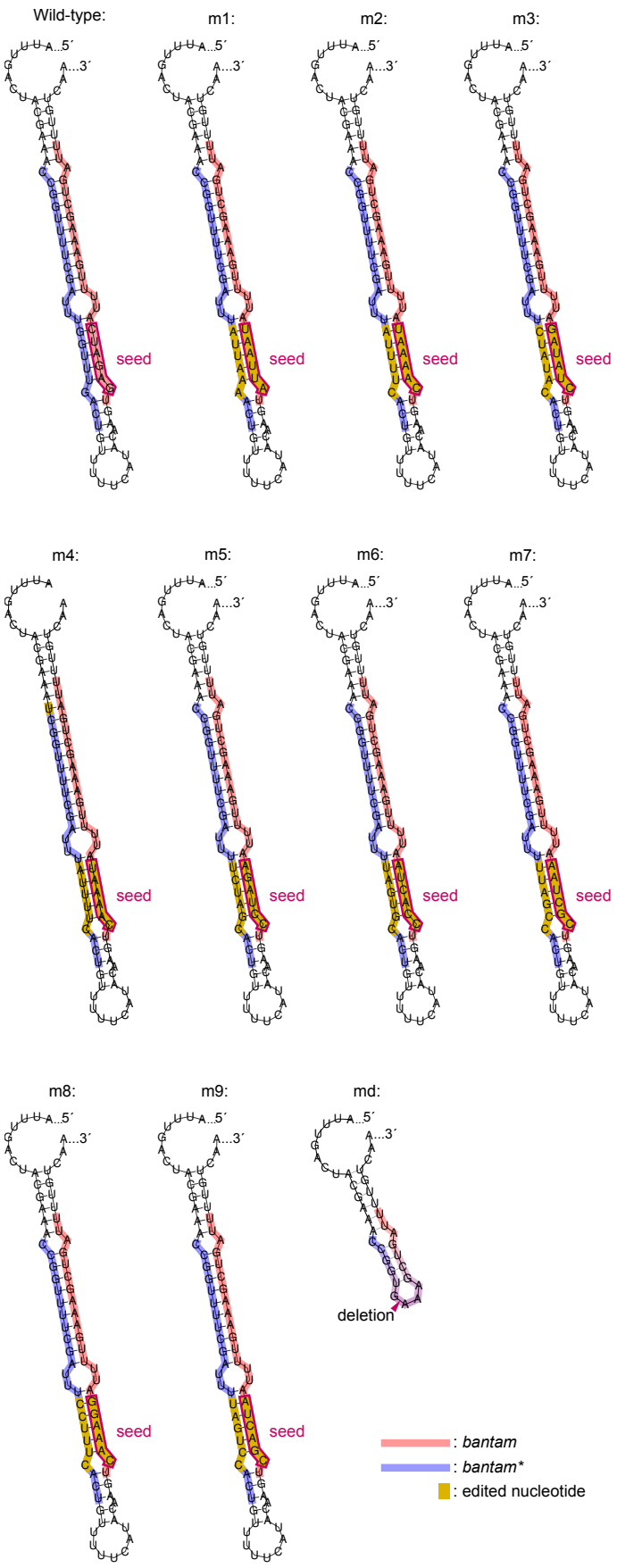


Figure 3: **Generation of an edited *bantam* allele with correct expression level. A.** Predicted secondary structure of wild-type and "m6" edited *bantam* loci. Same conventions as in Figure 1**A**. **B.** Measured abundances of wild-type, m2 and m6 *bantam* isoforms in TM6-balanced heterozygous L1 larvae (the "+/+" control is also heterozygous for the TM6 balancer chromosome). Wild-type isoforms "with correct seeds" are the isoforms having a GAGAUC hexamer at positions 2–7; m2 isoforms "with correct seeds" are the isoforms having a CAAAAU hexamer at positions 2–7; m6 isoforms "with correct seeds" are the isoforms having a CCACUA hexamer at positions 2–7. Read counts for m2 and m6 *bantam* with correct seed were corrected for sequencing biases using the observed spike-in read counts. Measured fold-changes between wild-type and m2 or m6 alleles are indicated (geometric mean of the fold-change across 2 or 3 biological replicates for m2/TM6 and m6/TM6 genotypes respectively; *p*-values were calculated using a paired t-test). **C.** Growth defects in wild-type, homozygous *bantam*-edited strains, and in Δ1 homozygous pupae [Hipfner et al., 2002], imaged at the same magnification.

6 Supplementary Data

DNA oligos for donor DNA construction:	
d1319	AAGAGCTCAAAGGGAGCGAAAAACACTGAG
d1320	AACTCGAGGATTTGATTTGAGCACCGAATG
DNA oligos for sgRNA-expressing cassette construction:	
d1343	GTCGGACCATCGGAATGTGGAATG
d1316	AAACCATTCCACATTCCGATGGTC
d1938	GTCGCGATATCGCTATATCTGTGT
d1939	AAACACACAGATATAGCGATATCG
DNA oligo for 2S rRNA capture:	
d2227	Bio-AGTCTTACAACCCTCAACCATATGTAGTCCAAGCAGCACT
DNA oligos for luciferase reporter construction:	
d2289	AAGAGCTCGACGCGTTTGTAACTCCAATG
d2290	GCATCTAGAAATCGACCATCGGAATGT
d2436	AAATCGAAAACCGGTTTCGTAGTC
d2437	ATTTTGAAAGCTGATTTTGTCAATGAATACC
d2438	ACCGGTTTTCGATTTTCTAGCACTGTTTTTCATACAAGTCCTAGAATTTTGAAAGCTGAT
d2439	ATCAGCTTTCAAAATTCTAGGACTTGTATGAAAAACAGTGCTAGAAAATCGAAAACCGGT
d2440	ACCGGTTTTCGATTTTAGTGCACTGTTTTTCATACAAGTCCACTAATTTTGAAAGCTGAT
d2441	ATCAGCTTTCAAAATTAGTGGACTTGTATGAAAAACAGTGCACTAAAATCGAAAACCGGT
d2442	ACCGGTTTTCGATTTTTAGCCACTGTTTTTCATACAAGTCGCTAAATTTTGAAAGCTGAT
d2443	ATCAGCTTTCAAAATTTAGCGACTTGTATGAAAAACAGTGGCTAAAAATCGAAAACCGGT
d2444	ACCGGTTTTCGATTTCCTTTCACTGTTTTTCATACAAGTCAAAGGATTTTGAAAGCTGAT
d2445	ATCAGCTTTCAAAATCCTTTGACTTGTATGAAAAACAGTGAAAGGAAATCGAAAACCGGT
d2446	ACCGGTTTTCGATTTTAGTCCACTGTTTTTCATACAAGTCGACTAATTTTGAAAGCTGAT
d2447	ATCAGCTTTCAAAATTAGTCGACTTGTATGAAAAACAGTGGACTAAAATCGAAAACCGGT
DNA oligos for GFP reporter construction:	
d1360	GCAACTACTGAAATCTGCCAAG
d1361	GTCCAATTATGTCACACCACAG
d1379	GCATCTAGAAATCGACCATCGGAATGTGG
d1380	CGACGCGTTTGTAACTCCAATGATTTG
Spiked-in RNA oligos for Small RNA-Seq bias measurement:	
r25	P-UGAGAUCAUUCGCAAAGCUGAUU
r26	P-UCAAAAUAUUCGCAAAGCUGAUU
r64	P-UCCACUAAUUCGCAAAGCUGAUU

Supplementary Table 1: **Oligonucleotides used in this study.** "Bio": biotin (with a 11-heteroatom linker to 5´ phosphate). "P": phosphate. Sequences are written 5´ to 3´.



Supplementary Figure 1: **Sequence and predicted secondary structure for the variants analyzed in Figure 2.** Same conventions as in Figure 1**A**.