

Figure S1. Characterization of TT-FHago2 and TT-Ago2 mESCs, related to Figure

1. (A) Western blot of TT-Ago2 (left) and TT-FHago2 (right) cells with titration of doxycycline (Dox). AB2.2 included for reference of wildtype expression. (B) Growth rate of TT-FHago2 mESCs supplemented with either no Dox, low (+) or high (++) Dox concentration (biological triplicates are plotted) after 96hrs doxycycline starvation. Error bars represent standard deviation. (C) Northern blot of miRNA expression levels in E7 cells or E7 cells overexpressing wildtype FLAG-HA tagged Ago2 or indicated mutants. Increased miRNA levels are observed with wildtype or catalytically inactive Ago2 expression, but not with Ago2 miRNA-binding mutant expression. (D) Scatter plot showing the log₂ expression levels of miRBase v19 annotated miRNA in TT-FHago2 mESCs after 96 hrs of doxycycline starvation (NoDox) or 96 hrs of doxycycline starvation followed by 48 hrs of high doxycycline treatment (++ Dox). Those miRNAs not destabilized by at least two-fold are colored in red and all other miRNA in black. Only miRNA that have > 2 reads in the datasets were considered.

Figure S2. TT-FHago2 and TT-Ago2 mESC FLAG-immunoprecipitation, related to

Figure 2. (A) Representative western blot of immunoprecipitation efficiency from TT-FHago2 mESCs. I=input, FT=flow-through. (B) Correlation of miRNAs cloned from TT-FHago2 and TT-Ago2 mESC total small RNA inputs. (C) Analysis pipeline of small RNAs sequenced in IP samples. (D) Determination of empirical false discovery rate (FDR) for enrichment cutoffs in reciprocal TT-FHago2 and TT-Ago2 small RNA enrichment calls. (E) MA plot highlighting enrichment in Ago2 for snoRNA and tRNA fragments as described as in Figure 3.

Figure S3. Additional analysis of protein-coding regions producing Ago-bound small RNA, related to Figure 3. (A) Barplot of the number of intronic and exonic regions of filtered Ago2-bound regions that overlap protein-coding genes. (B) DAVID gene annotation analysis output for significantly enriched categories. (C) Sequencing results from cDNA cloned from mESCs for three genes that share a promoter with TSS-miRNAs. First exon is black, second exon blue, and most abundant Ago enriched small RNA in red. (D) Scatterplot showing the lack of correlation between TSS-miRNA and mRNA transcript levels (mESC RNAseq, unpublished) from shared RNAPII promoters (top) and promoter-proximal RNAPII reads as measured by GRO-Seq (bottom).

Figure S4. TSS-miRNAs are Ago/Dicer-dependent, related to Figure 4. (A) IGV genome browser shots for two example TSS-miRNAs mapped to the mm9 genome and cloned from indicated samples. Collapsed reads are represented as blue or red denoting minus strand or plus strand, respectively. Gray barplots show coverage over the region. Arrows indicate the direction of transcription. (B) Left: MA plot of enrichment in TT-FHago2 FLAG IP compared to TT-Ago2 control. Annotated miRNA are colored blue and the example TSS-miRNAs highlighted in the text are colored red. Right: mESC copy number estimation of three example TSS-miRNAs. These correspond to estimates of 63,12 and 3 copies per cell for *Cpsf4l*, *Glul* and *Krcc1* TSS-miRNA, respectively. (C) Small RNA profiling from promoters that generate TSS-miRNAs and all other RNAPII promoters for wildtype and Drosha null mouse embryonic fibroblasts. Sequencing data from (Chong et al., 2010).

Figure S5. Comparison of RNAPII levels and localization from TSS-miRNA producing genes and *Cpsf4l* TSS-miRNA levels upon pausing factor knockdown, related to Figure 5. (A) Top Left: Meta analysis of Gro-Seq data at TSS-miRNA producing promoters aligned to the sense TSS. Bottom Left: Meta analysis of Gro-Seq data of matched control genes aligned to the sense TSS. Top Right: Boxplot of Gro-Seq reads from the promoter-proximal RNAPII peak for control matched promoters or TSS-miRNA producing genes. Control versus TSS-miRNA genes p-value < 0.07. Bottom Right: Meta analysis of RNAPII ChIP-Seq data aligned to the TSS for TSS-miRNA producing genes and the control matched set of promoters. (B) Left: Splint ligation-mediated detection of *Cpsf4l* TSS-miRNA and Snora15a upon stable knockdown of various pausing factors in mESCs. Right: Quantitative PCR measurement of *Cpsf4l* mRNA upon knockdown of indicated pausing factors. N=2, error bars represent standard deviation. Plko.1 samples are infected with the empty shRNA lentiviral delivery vector.

Figure S6. Single-cell based dual reporter TSS-miRNA activity assay and effect of TSS-miRNA overexpression on local gene expression, related to Figure 6. (A) Diagram of the bidirectional reporter constructs used in the assay. (B) Representative FACS data. (C) Scatter plot of single-cell bidirectional dual fluorescent reporter flow cytometry measured expression data. Points represent binned targeted mCherry log₁₀ by control eYFP log₁₀ expression in wildtype (blue) or *Dicer*^{-/-} mESCs (red) for the construct containing zero perfect matches to *Cpsf4l* TSS-miRNA in the mCherry 3' UTR. Points and error bars are mean and standard deviation of independent biological

replicates. (D) Normalized fold-repression is plotted for each control expression bin. (E) Ago2-dependent repression of a luciferase reporter construct containing a 3' UTR with a single perfect, bulged or seed mutant site complementary to the *Krcc1* TSS-miRNA. Values plotted are cells transfected with *Krcc1* TSS-miRNAs mimic at 100nM either with or without FHago2 expression (completed in biological triplicate). Inset: graphical representation of reporter constructs utilized in this experiment. (F) Left: *Krcc1* gene architecture diagram representation. Approximate location of *Krcc1* TSS-miRNA is denoted by purple line. Right: qPCR measurement of mRNA levels in wildtype or Dicer null mESCs transfected with either control siRNA or *Krcc1* TSS-miRNAs mimic. N=3, error bars represent standard deviation. (G) Left: *Cpsf4l* gene architecture diagram representation. Approximate location of *Cpsf4l* TSS-miRNA is denoted by purple line. Right: qPCR measurement of mRNA levels in wildtype or Dicer null mESCs transfected with either control siRNA or *Cpsf4l* TSS-miRNAs mimic. N=6, error bars represent standard deviation.

Figure S7. TSS-miRNAs are detected in various mouse and human tissues, related to Figure 7. (A) IGV genome browser shots for three selected TSS-miRNAs cloned from the indicated mouse tissue and mapped to the mm9 genome assembly. (B) IGV genome browser shots for two orthologous TSS-miRNA promoters with collapsed reads mapped to hg19 genome assembly from various human tissues. Red boxes highlight the orthologous TSS-miRNA region in mouse. Collapsed reads are represented as blue or red denoting minus strand or plus strand, respectively. Gray barplots show coverage over the region. Arrows indicate the direction of transcription. (C) Heatmap of TSS-miRNA

expression in various mouse tissues used in Figure 7A and mRNA expression from matched samples (Merkin et al., 2012) ordered based on the clustering of TSS-miRNA levels to show lack of correlation to mature mRNA production.

Table S1. Data summary for small RNA sequencing from TT-FHago2 cells to measure Ago-dependence, related to Figure 1.

Table S2. Data summary for identification of Ago2-bound small RNAs, related to Figure 2.

Table S3. Summary of filtered Ago2-bound regions within protein-coding genes, related to Figure 3.

Table S4. TargetScan predictions of *Cpsf4l* TSS-miRNA targets, related to Figure 5.

Extended Experimental Procedures

Cell Culture Conditions

mESCs were cultured under standard conditions (Tremml et al., 2008). Briefly, mESCs were grown on gelatinized tissue culture plates in Dulbecco's Modified Essential Media supplemented with HEPES pH 7.0, 15% fetal bovine serum (Hyclone/Thermo Scientific), 1000U/mL leukemia inhibitory factor (Chemicon/Millipore), 0.1 mM non-essential amino acids, 0.1 mM L-glutamine, 0.1 mM Pen/Strep and 0.11 mM β -mercaptoethanol. TT-FHago2 and TT-Ago2 clonal cell lines were propagated and maintained in 0.1 μ g/mL doxycycline (Sigma).

Generation of lentivirus and mESC infection

pSLIK lentivirus was generated as follows: human FHAgO2 was subcloned into pENTT entry vector and sequence verified. Doxycycline-inducible promoter and FHAgO2 sequence was recombined into pSLIK-Hygro using Gateway LR Clonase II (Invitrogen) and vectors were verified using restriction enzyme digestion and DNA sequencing. pSLIK lentivirus was generated by transfection into 293T cells alongside pCMV-dR8.91 packaging and pHDM.G envelope vectors. 293T cell media supernatant containing lentivirus was collected, filtered through a 0.45 μm syringe filter and supplemented with 1 mM HEPES pH 7. E7 cells were combined with equal parts fresh mESC media and lentivirus along with 4 $\mu\text{g}/\text{mL}$ polybrene (Millipore) in a six-well culture dish and centrifuged at room temperature for 1 hour (hr) at 2000 RPM in a Beckman-Coulter Allegra X-15R centrifuge. After overnight incubation at 37°C, cell media was replaced and supplemented with 150 $\mu\text{g}/\text{mL}$ Hygromycin. Cells were grown in the presence of Hygromycin until ready to be split out of the 6-well dish. All experiments were carried out with TT-Ago2 clone 1 or TT-FHAgO2 clone 3.

Western Blotting

mESCs were lysed in RIPA buffer (1% NP-40, 0.1% SDS, 0.5% Deoxycholate in PBS pH 7) containing 1x protease inhibitors (Roche) and protein was quantified using the Pierce BCA protein assay (Thermo Scientific). Protein was resolved on Novex gradient denaturing PAGE gels (Life Technologies) and transferred to PVDF membrane (Millipore). Antibodies used in this study: Ago2 (Cell Signaling), Tubulin (GenScript), HA (Roche), FLAG (Sigma). Secondary HRP-conjugated antibodies were from GE Healthcare.

Immunopurification

Doxycycline-induced cells were washed once with cold HEPES-buffered saline (HBS), trypsinized and collected. Cells were then centrifuged and washed once with cold phosphate buffered saline (PBS) and counted. An equal number of cells for each sample were centrifuged and then lysed in modRIPA buffer (10 mM Tris-Cl pH 7.4, 150 mM NaCl, 1% Triton-X 100, 0.1% SDS and 1 mM EDTA) containing 1x complete protease inhibitors (Roche). Cells were rotated at 4°C for 30 minutes to ensure cell lysis. Cell lysate was centrifuged to remove cellular debris and the supernatant was used for IP. FLAG M2 antibody (Sigma) was conjugated to protein G coated Dynabeads (Life Technologies) for at least 2 hrs in 0.1 M Na-Phosphate pH 8 at 4°C with rotation. Cleared extracts were incubated overnight at 4°C with rotation. Beads were collected and washed twice with PBS, 3x with modRIPA, 2x with modRIPA + 300mM NaCl and 1x with modRIPA for 5 minutes each with rotation at 4°C. RNA was eluted twice with 30 minute incubations with rotation at 4°C with 3x FLAG peptide (Sigma).

Luciferase and Fluorescent miRNA Reporter Assays

Cells were placed in 24-well plates 24 hrs prior to transfection at $2e^5$ cells/well. Cells were transfected utilizing Lipofectamine 2000 (Life Technologies) with pRL-CMV containing either targeted (perfect or bulged) or untargeted (seed mutant) sequences within the 3' UTR of the Renilla transcript, or pRL-CMV lacking a 3' UTR (empty) along with PGL3 untargeted control and pWhiteScript. Cells were transfected for 4 hrs in Optimem media (Life Technologies). After 24 hrs, cells were lysed in 1x Passive Lysis

Buffer (Promega). Renilla and Firefly luciferase were quantified using Promega Dual Luciferase Reporter Assay System. Values represent targeted Renilla luciferase relative to untargeted Firefly luciferase as a transfection control.

For the single-cell flow cytometry based TSS-miRNA activity assay, wildtype or Dicer knockout cells were plated at 2×10^5 cells/well in a 12-well dish 24 hrs prior to transfection. Cells were transfected with Lipofectamine 2000 (Life Technologies) with control reporter construct (0x) or a construct containing three perfectly complementary Cpsf4l TSS-miRNA target sites (3x) in the mCherry 3' UTR, along with rtTA expression vector and pWhiteScript. Cells were transfected for 4 hrs in Optimem media (Life Technologies) after which media was replaced with standard mESC media supplemented with $1 \mu\text{g/mL}$ doxycycline (Sigma). After 24 hrs, cells were harvested, analyzed and data processed with FlowJo and custom Matlab scripts as previously described (Mukherji et al., 2011).

RNA analysis

RNA was isolated using Trizol (Life Technologies) and treated with DNA-free DNase (Ambion). cDNA for quantitative PCR (qPCR) was synthesized using QuantiTect Reverse Transcription Kit (Qiagen). The cDNA was measured using PowerSybr (Applied Biosystems) on the Applied Biosystems 7500 Real Time PCR System. qPCR primer sequences are available upon request.

Small RNA sequencing and analysis

Briefly, a circularization-based protocol was adapted from (Churchman and Weissman, 2011) and included a Circligase-mediated step after reverse transcription from a 3' end ligated adaptor bypassing the 5' end adaptor ligation. This protocol allowed 5' end independent cloning of small RNA. The small RNA used in the library preparations were resolved on denaturing polyacrylamide gels to isolate RNA species below the abundant tRNA band (~18-75 nt based on low molecular weight ladder (New England Biolabs)). Sequencing was performed as described previously with barcoded samples for multiplexing (Gurtan et al., 2012). DNA sequences from the Illumina HiSeq 2000 Sequencing system were grouped by barcode, processed for adaptor removal and size exclusion of sequences < 15 nt with Cutadapt (Martin, 2011). Filtered reads were collapsed to unique sequences with the FASTX-Toolkit (from the laboratory of Gregory Hannon), trimmed to 21 nt and mapped to the full mouse University of California at Santa Cruz genome mm9 assembly with the Bowtie short read alignment tool (Langmead et al., 2009) allowing a single mismatch per alignment region with multiplicity of up to 500. Reads were quantified using a combination of BEDTools (Quinlan and Hall, 2010) commands and custom scripts. Reads were classified as pre-miRNA if they mapped to one annotated miRNA end and were longer than 30 nt. After quantifying and excluding rRNA and miRNA regions, per nucleotide genome coverage was determined for each dataset and regions of continuous coverage greater than expected from a poisson noise distribution were identified. These regions were merged for all datasets and intersected with gene annotation in the order illustrated in Figure S2C. These annotated regions named by the overlapping gene feature and 5' most nucleotide were then used for the quantitation of reads in each individual dataset. Reads of all lengths were considered for

normalization. Quantitation of reads that map to multiple positions in the genome were adjusted as described in (Ruby et al., 2006) and were included in quantification of repetitive regions, such as some annotated miRNA. For other non-repetitive regions, only uniquely mapped reads were used and only 20-30 nt length reads were used for quantitation to determine Ago enrichment. Normalized read values were used to perform analysis for differential expression or enrichment. For the classification of Ago-bound small RNA, a region must be enriched by more than 3-fold in both biological replicates or, if not detected in negative control, be represented by 2 reads in each biological replicate. Mouse gene annotations for analysis were from miRBase v19 and Ensembl. Protein-coding genes used in analysis exclude histone genes. Targetscan predictions were generated using version 6.0.

The gene ontology analysis was performed with DAVID and functional categories with Benjamini-Hochberg corrected p-values < 0.01 are shown. The secondary structure predictions were generated using RNAfold from the Vienna package with default settings. Two-tailed Student's t-test was applied to luciferase reporter assays where indicated and a Kolmogorov-Smirnov test was used to compare the distribution of cells in the single-cell FACs data.

mRNA cloning

For the cloning of full-length mRNA from TSS-miRNA promoters, cDNA was synthesized with SuperScriptIII First Strand Synthesis kit (Life Technologies) and inserted into a TOPO cloning vector (Life Technologies) for sequencing. For

overexpression, the *Cpsf4l* cDNA region was cloned into pcDNA 3.1+ overexpression vector and sequence verified.

Northern blotting and splint ligation detection of RNA

20µg of total RNA was separated on a 12% polyacrylamide/Urea/TBE gel and transferred to HyBond N+ membrane (GE Healthcare) using the TransBlot SD Semi-dry Transfer System (Biorad) and UV crosslinked. DNA oligo probes were γ -³²P end labeled using T4 polynucleotide kinase (New England Biolabs) and purified using Illustra G-25 MicroSpin columns (GE Healthcare). Blots were blocked in ULTRAHyb Oligo (Ambion) at 42°C for at least 30 minutes and hybridized with a radiolabeled DNA probe overnight at 42°C. Blots were washed twice for 30 minutes with 2xSSC and 0.5% SDS pre-warmed to 42°C. RNA was visualized with phosphor screens from GE-HealthCare.

The splint ligation protocol was performed as described in (Maroney et al., 2008) using 3ug of total RNA. The bridge sequences used were:

Cpsf4l TSS-miRNA (GAATGTCATAAGCGTCAGAGGGTGTGTGGGACAG)

Snora15 (GAATGTCATAAGCGATTTGTACTCACTTCTAATAAT)

with the sequence complementary to radio labeled adaptor underlined.

Chromatin-associated small RNA sequencing

Chromatin was isolated using the procedure described in (Wuarin and Schibler, 1994) from 8 x 15-cm² plates. Briefly, cells were washed once in cold 1X PBS, resuspended in Buffer A (10mM HEPES, pH7.3, 10mM KCl, 1.5mM MgCl₂, 0.34M

Sucrose, 10% Glycerol, 1mM DTT, 1x Complete EDTA free protease inhibitor cocktail (Roche)) and lysed with 0.5% Triton-X 100. The pellet was then washed twice in Buffer A. Nuclei were lysed with two washes of Buffer B (20mM HEPES, pH 7.3, 1mM DTT, 7.5 mM MgCl₂, 0.2mM EDTA, 0.3M NaCl, 1M Urea, 1% NP-40). The chromatin pellet was resuspended in Buffer C (50mM Sodium Acetate pH 5.5, 50 mM NaCl, 0.5% SDS) and phenol chloroform extracted twice. Size fractionation for RNA less than 200 nt was performed with miRNeasy Mini columns (Qiagen) with a gDNA Eliminator Mini spin column step incorporated for genomic DNA removal. Three micrograms of isolated RNA was treated as described in (Nechaev et al., 2010) with 5'-polyphosphatase (Epicentre) and Terminator 5' phosphate dependent exonuclease (Epicentre) treatment each followed by phenol chloroform extraction. The RNA was then size fractionated on a 15% denaturing gel to isolate different size ranges including the 50-100 nt range used to identify overlaps with TSS-miRNA regions. The gel purified RNA was ligated to miRNA linker #1 (IDT) and small RNA cloning was performed with the cDNA circularization based procedure using primers for 3' end sequencing. A full description of this dataset will be presented elsewhere.

Small RNA mimics

Small RNA mimics were ordered from Dharmacon as a siRNA duplex containing 2-3' nt overhangs. All control siRNA transfections utilized the Dharmacon siGenome Control non-targeting control #2. TSS-miRNA mimic sequences are as follows: Cpsf4l, sense: 5'CUGUCCACACACCCUCUGAC3', antisense: 5' GUCAGAGGGUGUGUGGGACAG 3'. Krcc1, sense:

5'CUCUCCGCCACCUCCACCGCAG 3', antisense: 5'

GCGGUGGAGGUGGCGGAGAGAG 3'. All TSS-miRNA mimic transfections were performed using Lipofectamine 2000 (Life Technologies).

- Chong, M.M., Zhang, G., Cheloufi, S., Neubert, T.A., Hannon, G.J., and Littman, D.R. (2010). Canonical and alternate functions of the microRNA biogenesis machinery. *Genes Dev* 24, 1951-1960.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-373.
- Gurtan, A.M., Lu, V., Bhutkar, A., and Sharp, P.A. (2012). In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. *RNA* 18, 1116-1122.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10, R25.
- Maroney, P.A., Chamnongpol, S., Souret, F., and Nilsen, T.W. (2008). Direct detection of small RNAs using splinted ligation. *Nat Protoc* 3, 279-287.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17, 10-12.
- Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593-1599.
- Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* 327, 335-338.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127, 1193-1207.
- Tremml, G., Singer, M., and Malavarca, R. (2008). Culture of mouse embryonic stem cells. *Curr Protoc Stem Cell Biol Chapter 1*, Unit 1C 4.
- Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* 14, 7219-7225.

Figure S1

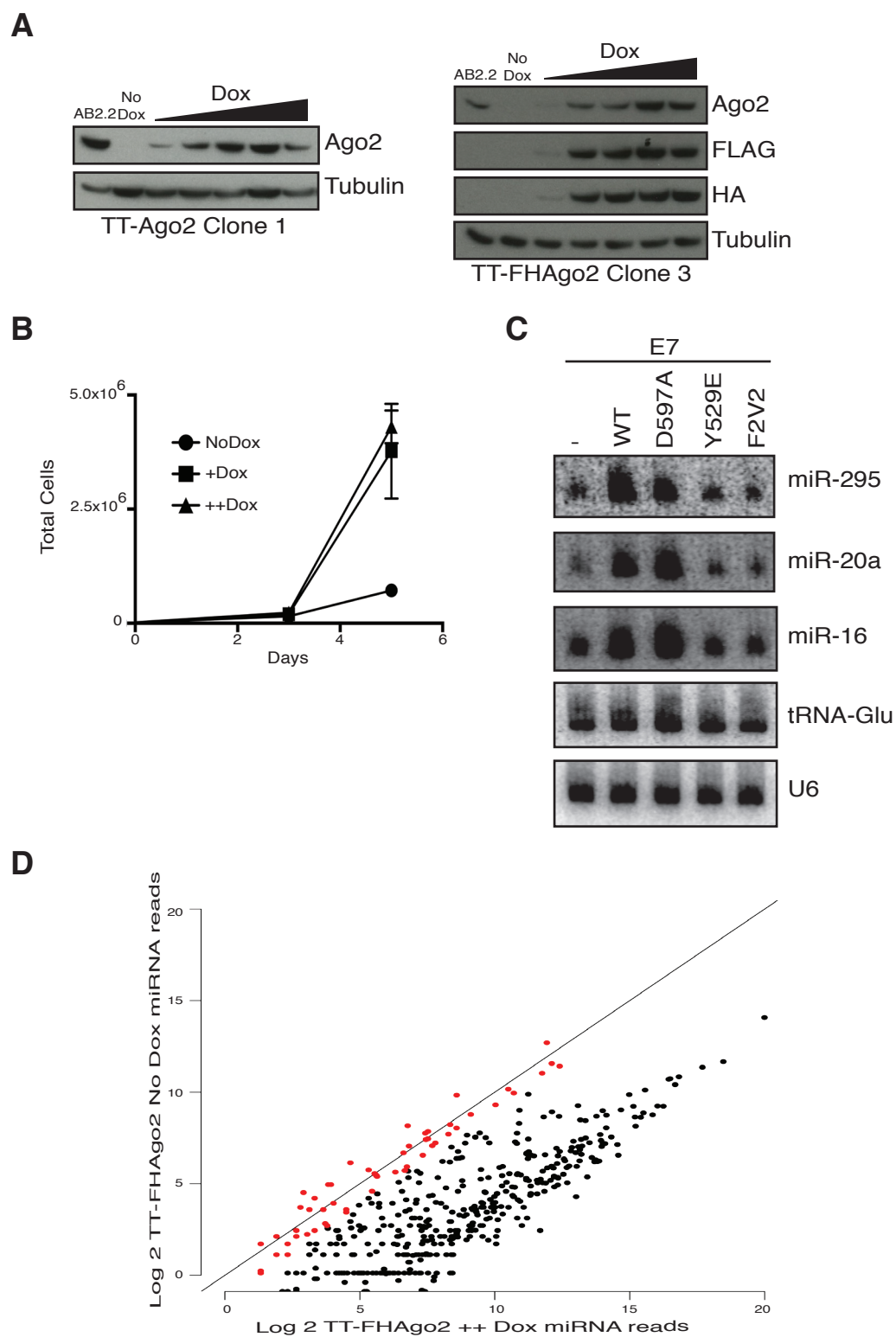


Figure S2

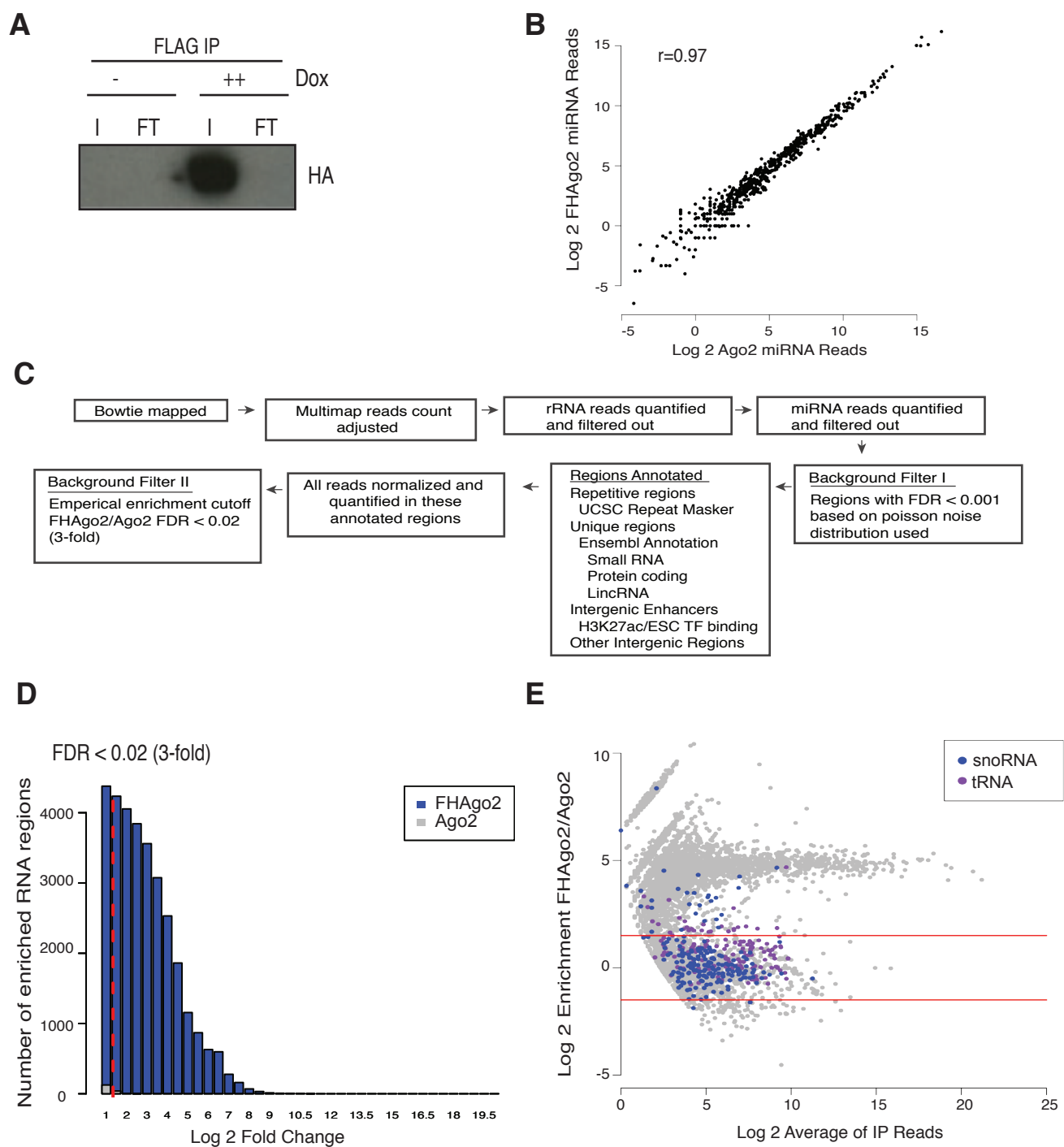
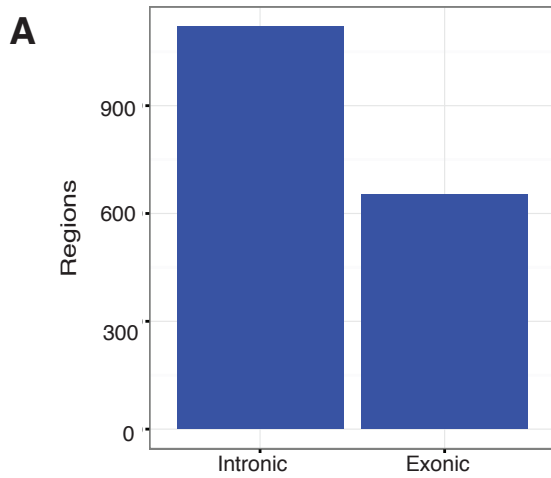


Figure S3

**B**

Category	Term	Count	p Value	Benjamini
SP_PIR_Keywords	Phosphoprotein	201	2.3×10^{-10}	6.1×10^{-8}
SP_PIR_Keywords	Nucleus	134	6.6×10^{-9}	8.9×10^{-7}
SP_PIR_Keywords	Transcription Regulation	61	1.3×10^{-5}	1.1×10^{-3}
SP_PIR_Keywords	Acetylation	80	5.4×10^{-5}	3.7×10^{-3}
SP_PIR_Keywords	Transcription	65	5.5×10^{-5}	3.0×10^{-3}

CCpsf4l

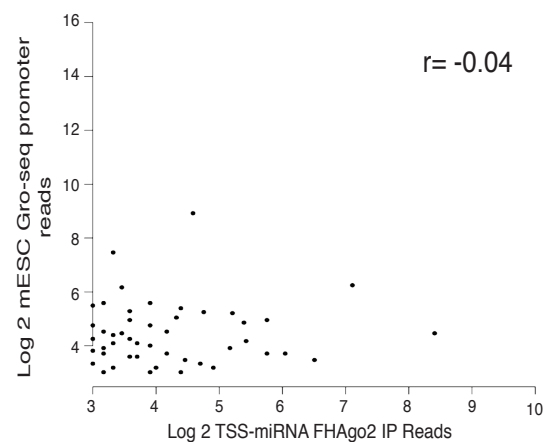
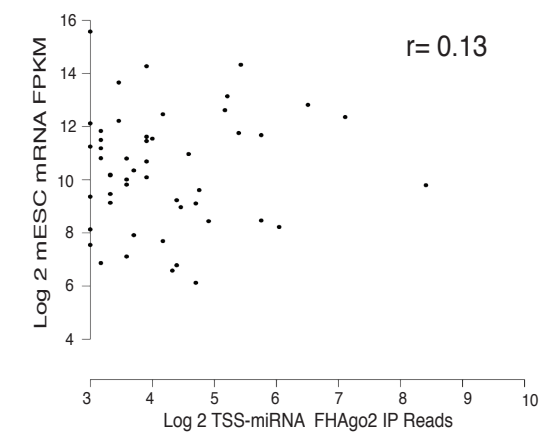
GGAGAGTGGCTGGGACCTAGTGGAGGAAC
 CTTTCCTCTGTCCACACACCCTCTGACTAC
 CCCCATCCCCTCCACCCAACACCTGGGCA
 TCGCTGTATGCAGTAGGTTTCGCATCAATGGG
 TAGTTATATCCCCACACATAACAATGGGACACT
 GGTTTGGCAACGGCTCTGTTCTGTGATCGG
 CAGAGATAAGCAGAGTCCAGTTCAGCTGTG

Glu1

AGAGCGGAGAATGGGAGTAGAGCAGAGTGT
 CTGAACAGCACGCTCACCCATCTCCTCTCCG
 CCTCGCTCTCCTGACCTGTTACCCATCCATC
 ATCCGGCCGGCCACCGCTCTGAACACCTTCC
 ACCATGGCCACCTCAGCAAGTTCCCACTTGA
 ACAAAGGCATCAAGCAAATGTACATGTCCCTG

Krcc1

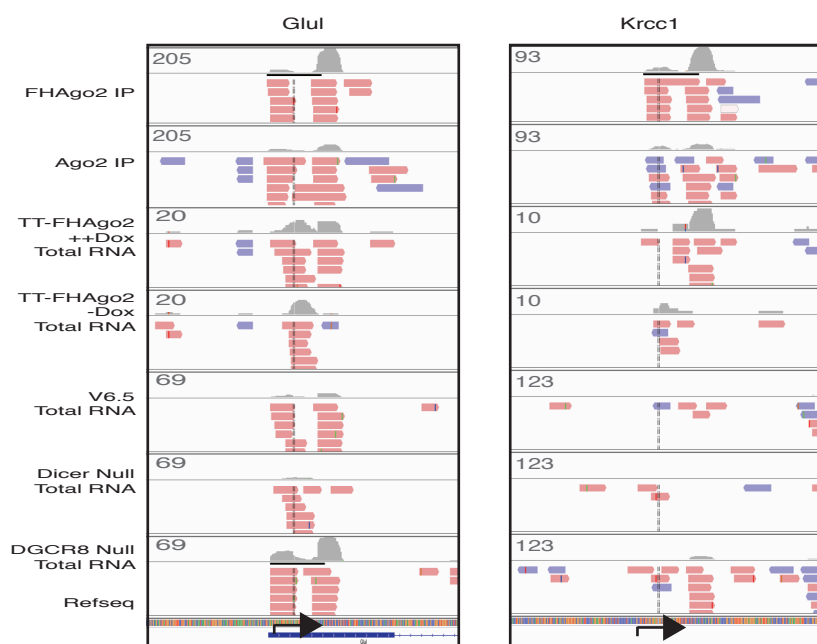
CTCCCACCCCGCCTGCTCTGCGGCGGAGGC
 GGCGGCGGCGGCAGCGGTGTCTCTGCCTCT
 CCGCCACCTCCACCGCAGCCTCTGACCCCGG
 GCCGGCGTTGTTGGCGGGGAGGGGGACAGT
 AGTTGTAGACGCCCGCCCTGCCTCAGAGAA
 GGTGAATGAGAGTTGAGTATTAGACAAAAATAG
 ATATAATCTCTACTCCATGGTTAGCTCCCTGAT

D

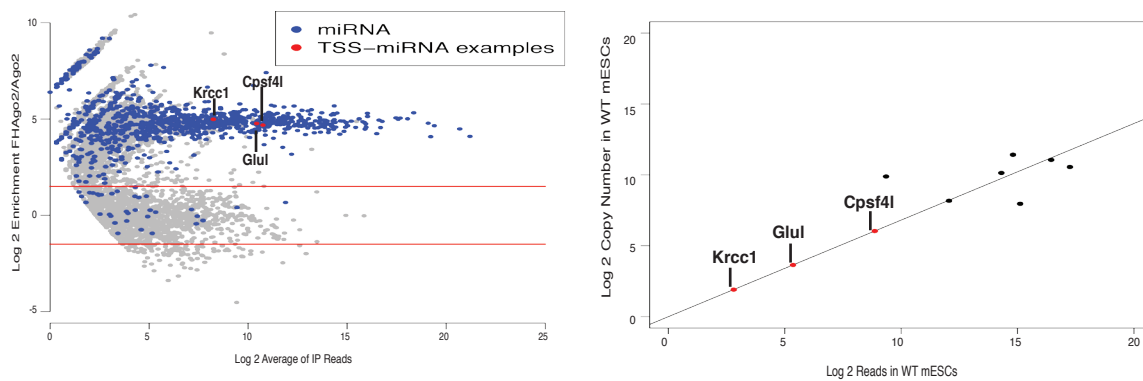
Zamudio et al.

Figure S4

A



B



C

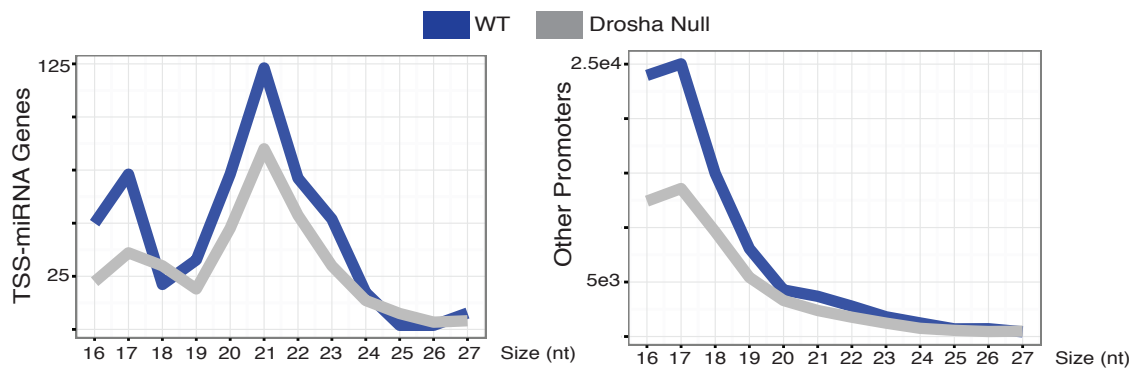
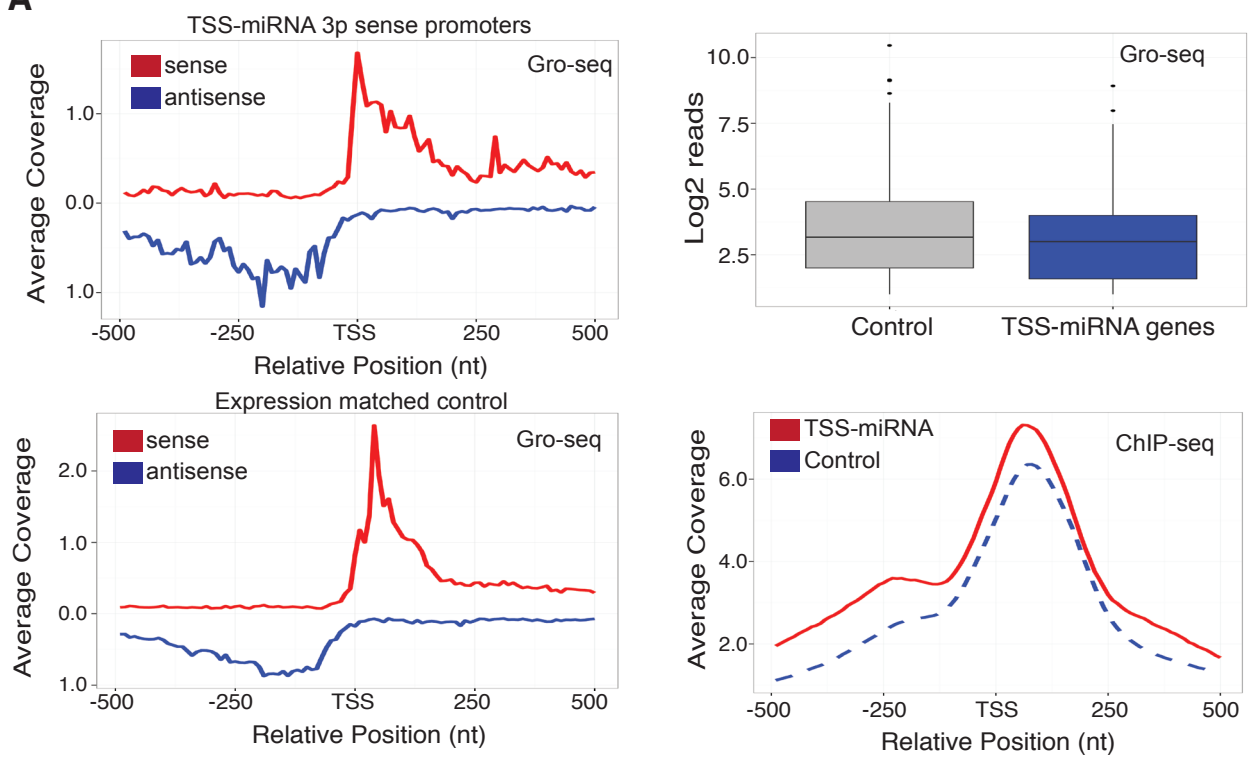


Figure S5

A



B

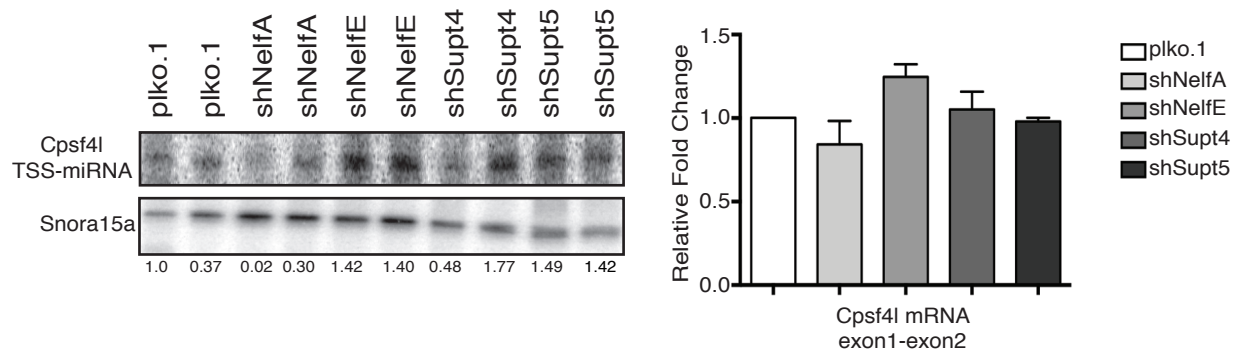


Figure S6

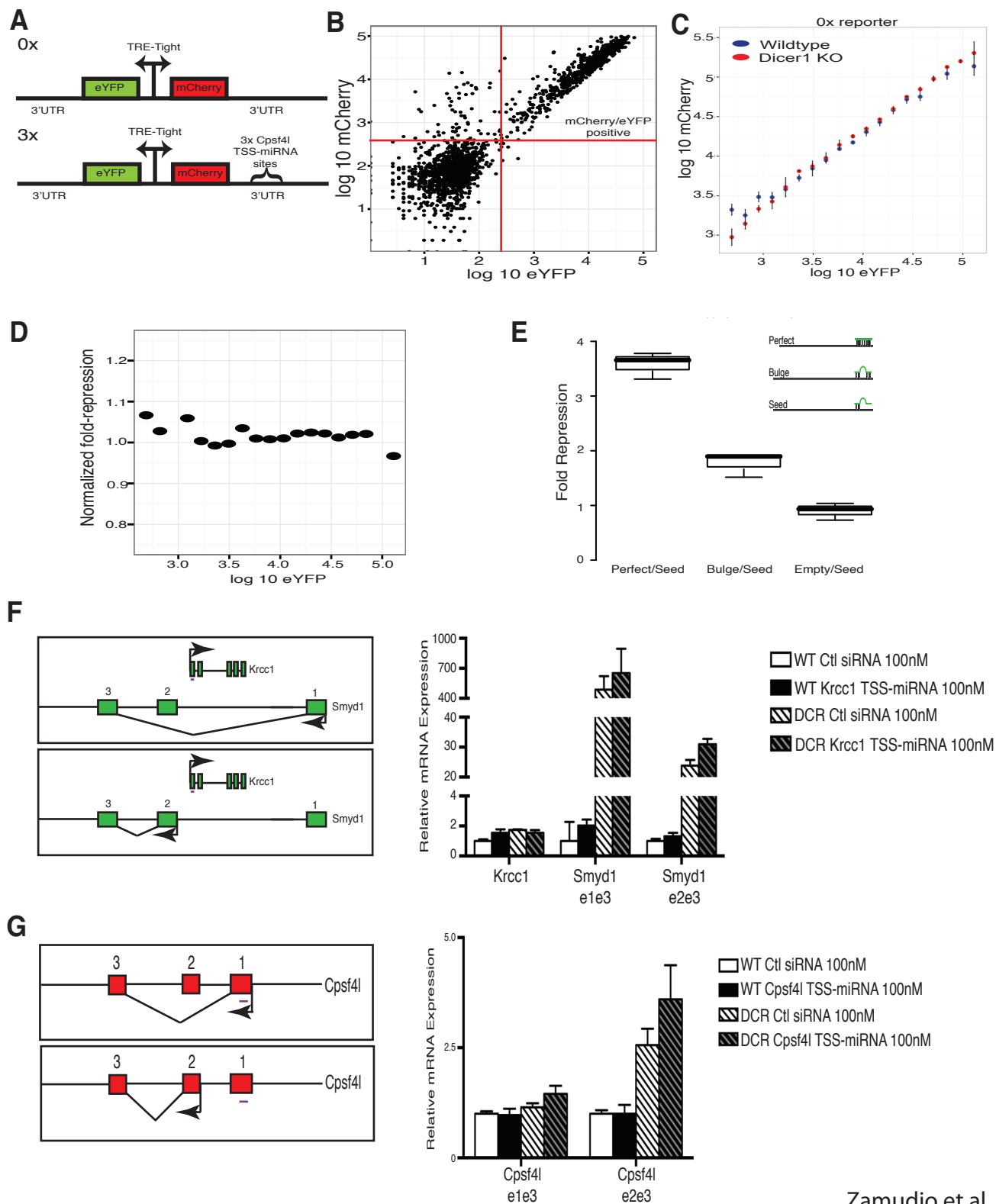


Figure S7

