

## SUPPLEMENTARY MATERIAL

### The human box C/D snoRNA U3 is a miRNA source and miR-U3 regulates expression of sortin nexin 27

Nicolas Lemus-Diaz, Rafael Rinaldi Ferreira, Katherine E. Bohnsack, Jens Gruber, Markus T.

Bohnsack

## SUPPLEMENTARY FIGURE LEGENDS

**Supplementary Figure S1. Quality control of small RNA sequencing.** Small RNA-seq was performed on total RNA extracted from HEK293 cells. **(A)** Correlation plot of sequenced libraries L001, L002 and L003 (L001-L002  $R_2=0.972$  L001-L003  $R_2=0.967$ ), colours of dots represent the read numbers of annotated miRNAs and snoRNAs, the side bar shows the values in logarithmic scale. **(B)** Read length distribution of individual small RNA libraries represented as histograms displaying the number of miRNAs (21-22 nt), C/D snoRNAs (70-100 nt) and H/ACA (>130 nt, not sequenced in this preparation). **(C)** Normalised numbers of reads of less than 40 nt in length and the mean expression of H/ACA 45 derived miRNA, mapped to different snoRNAs shown as boxplots. Error bars represent the variation between three independent library preparations.

**Supplementary Figure S2. Fraction of reads mapped to individual classes of miRNA and snoRNA (C/D box and H/ACA).** **(A-C)** Reads of small RNA and CLIP-derived libraries were aligned to the genome and then realigned to a subset of mi/snoRNAs. Bar plots represent the fraction of mapped reads to individual classes of subset RNAs displayed in different colours. Fractions derived from small RNA libraries of total RNA (A), representative data sets of miRNA pathway proteins (B), representative data sets of snoRNA interacting proteins (C).

**Supplementary Figure S3. Correlation of predicted seed pairing stability (SPS) and target abundance (TA) for U3 derived miRNA.** Thermodynamic parameters of RNA duplexes (7mer-

m8) of all possible heptamer sequences (16,384) and the number of non-overlapping 3' UTRs sites (8mer, 7mer-m8, 7mer-1a and 6mer) in reference mRNAs were retrieved from the TargetScan (context+++ code). Scatter plot represents the correlation between TA as logarithmic values versus SPS in kcal/mol ( $R=0.3482$ ,  $p < 2.2e-16$ , Spearman correlation) of all possible heptamer (7 nt sequences). The red dot shows the putative U3-miRNA seed sequence (AGUUUCU), and the reported low proficiency miRNA (miR-23a) is shown in blue.

**Supplementary Figure S4. DROSHA, XPO5 and DICER knock-outs cell lines characterisation. (A-B)** HTC116 parental and knockout cells were examined for protein expression by western blot (A) and miRNA expression by small-RNA seq (B). Barplots depict selected miRNAs as a percentage of all annotated miRNA reads and clustered according to their biogenesis pathway. Error bars represent the SEM of sequencing replicates: 3 parental cell line, 2 DROSHA KO, 2 XPO5 (KO#19 and KO#19-1) and 2 Dicer (KO#43 and KO#45).

**Supplementary Figure S5. Overexpression of Dicer reduces the amount of U3 present in the cytoplasm. (A)** Wild-type HCT116 cells (WT) and Dicer knockout (KO) cells that had been transfected with a plasmid for the expression of Flag tagged Dicer (+) or not (-) were examined for Dicer expression by western blotting. Tubulin was used as a loading control. **(B)** Total RNA extracted from cells as in (A) was separated by denaturing PAGE and analysed by northern blotting using probes hybridising to the indicated RNAs. The levels of cytoplasmic U3 in the cells knockout cells expressing or not expressing exogenous Dicer relative to the wild type, normalised to tRNA<sub>Met</sub>, were determined in three biological and two technical replicates, and data are shown as mean  $\pm$  standard error. Significance was calculated using the one sample t-test; ns – not significant, \*\* -  $p < 0.01$ .

**Supplementary Figure S6. Evolutionary conservation of the U3 miRNA target sequence in the 3' UTR of the SNX27 mRNA.** An alignment of sequences from the U3 snoRNA and the SNX27 mRNA 3' UTR from different species is shown. Multiple sequence alignment of vertebrate U3 snoRNA and SNX27 mRNA sequences was performed using Clustal W with default parameters. Retrieved from ENSEMBL, the sequences from the following species were analysed: *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Macaca mulatta* (rhesus macaque), *Dasyus novemcinctus* (armadillo), *Loxodonta africana* (elephant), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Oryctolagus cuniculus* (rabbit), *Canis lupus familiaris* (dog), *Bos taurus* (cattle), *Echinops telfairi* (hedgehog), *Gallus gallus* (chicken), *Xenopus tropicalis* (frog). The seed sequence of the U3-derived miRNA-like fragment is shown in white, italics and underlined, and the target site (mRNA) is shown in bold, italics and underlined. Nucleotides identified in as chimeric sequences by CLASH are marked in yellow (U3 snoRNA) and green (mRNA).

## SUPPLEMENTARY TABLES

**Supplementary Table S1. Oligonucleotides used in this study**

<b>Target</b>	<b>5'-3'sequence</b>	<b>Application</b>
U6 forward	CTCGCTTCGGCAGCACA	qPCR
U6 reverse	AACGCTTCACGAATTTGCGT	qPCR
U3 forward	CGAAAACCACGAGGAAGAGA	qPCR
U3 reverse	CACTCCCCAATACGGAGAGA	qPCR
GAPDH forward	GTCAGCCGCATCTTCTTTTG	qPCR
GAPDH reverse	GCGCCCAATACGACCAAATC	qPCR
U3 snoRNA	GCTCTACACGTTTCAGAGAACTTCTCTAGTAACACACTA TAGAAATGATCCC	Northern blot probe
miR-U3	CTCTACACGTTTCAGAGAACTT	Northern blot probe
U6	ATATGGAACGCTTCACGAATTTGCGTGTCATCCTTGCG C	Northern blot probe
ACA45	TCAAGGACTCTATTTTCTTAGTCTCCA	Northern blot probe
tRNA <sub>Met</sub>	ATGGTTTCGATCCATCGACCTCTGGGTTATGG	Northern blot probe

miR-U3 mimic	AAGUUUCUCGGAACGUGUAGdTdT	siRNA
Scramble	UCGUAAGUAAGCGCAACCCdTdT	siRNA
GL2	CGUACGCGGAAUACUUCGAdTdT	siRNA outcompeting
Anti-miR-U3	CTACACGTTTCAGAGAACT	miRNA LNA inhibitor
Anti-miR-Neg	TAACACGTCTATACGCCCA	miRNA LNA inhibitor
ACA-45 1 copy forward	GGCCAAAGGTAGATAGAACAGGTC	Reporter construct
ACA-45 1 copy reverse	TCGAGACCTGTTCTATCTACCTTT	Reporter construct
ACA-45 3 copies forward	GGCCAAAGGTAGATAGAACAGGTCCCGGAAAGGTAGATA TAGAACAGGTCCCGGAAAGGTAGATAGAACAGGTC	Reporter construct
ACA-45 3 copies reverse	TCGAGACCTGTTCTATCTACCTTTCCGGGACCTGTTCTA TCTACCTTTCCGGGACCTGTTCTATCTACCTTT	Reporter construct
U3 1 copy forward	GGCCAAGTTTCTCTGAACGTGTAG	Reporter construct
U3 1 copy reverse	TCGACTACACGTTTCAGAGAACTT	Reporter construct
U3 3 copies forward	GGCCAAGTTTCTCTGAACGTGTAGCCGGAAGTTTCTCT GAACGTGTAGCCGGAAGTTTCTCTGAACGTGTAG	Reporter construct
U3 3 copies reverse	TCGACTACACGTTTCAGAGAACTTCCGGCTACACGTTTC AGAGAACTTCCGGCTACACGTTTCAGAGAACTT	Reporter construct
miRNA-92a-3p forward	TCGAGACAGGCCGGGTTCTGTGCAATATATGCGGCCG CATAT	Reporter construct
miRNA-92a-3p reverse	GGCCGCATATATTGCACAGAACCCGGCCTGTCTCGAG AGTA	Reporter construct
Non-targeted forward	GGCCCACGTACGCGGAATACTTCGAAA	Reporter construct
Non-targeted reverse	TCGATTTCGAAGTATTCCGCGTACGTG	Reporter construct
U3-55K Forward	CCGGGCTTGACCTTCTCTGTGACATCTCGAGATGTCAC AGAGAAGGTCAAGCTTTTTG	shRNA cloning
U3-55K Reverse	AATTCAAAAAGCTTGACCTTCTCTGTGACATCTCGAGAT GTCACAGAGAAGGTCAAGC	shRNA cloning
Scramble Forward	CCGGTCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGG GCGACTTAACCTTAGGTTTTTG	shRNA cloning
Scramble Reverse	ATTCAAAAACCTAAGGTTAAGTCGCCCTCGCTCGAGCG AGGGCGACTTAACCTTAGGA	shRNA cloning

**Supplementary Table S2. Small RNA libraries analysed in this study**

Description	GEO accession number	Reference	Comment
HEK293	GSE136745	This study	Ion Torrent small RNA library
HeLa cellular fractionation	GSE50057	52	Ion Torrent small RNA library

snoRNPs HEK293	GSE43666	35	PAR-CLIP of endogenous NOP56, Fibrillarin, Dyskerin and FLAG-NOP58. Size selection of small RNAs
DGCR8 HEK293	GSE39086	31	HITS-CLIP of FLAG and endogenous DGCR8
DICER HEK293	GSE55333	32	PAR-CLIP of FLAG-Dicer1
Knockouts small RNA seq	GSE77989	26	DICER, XPO5, DROSHA KO small RNA libraries
DROSHA-fCLIP	GSE93651	30	DROSHA f-CLIP
Argonaute Hek293	GSE21918	33	PAR-CLIP FLAG-AGO 1 to 4
	GSE50452	34	CLASH libraries E1, E4 and E7
	GSE58127	35	RIP-seq using 2 antibodies against Argonaute

### Supplementary Table S3. Antibodies used in this study

Antibody	Company	Dilution
GAPDH (ab8245)	Abcam, Cambridge, MA, USA	1:5000
DROSHA (ab12286)	Abcam, Cambridge, MA, USA	1:1000
XPO5 (sc-271036)	Santa Cruz Biotechnology, Dallas, TX, USA	1:1000
DICER (ab14601)	Abcam, Cambridge, MA, USA	1:1000
U3-55K (sc-515661)	Santa Cruz Biotechnology, Dallas, TX, USA	1:500
SNX27 (ab77799)	Abcam, Cambridge, MA, USA	1:5000
Anti-rabbit 800 IRDye 800CW	LI-COR Biosciences, Lincoln, NE, USA	1:5000
Antiy-rabbit 680 IRDye 680LT (926-68023)	LI-COR Biosciences, Lincoln, NE, USA	1:5000

**Supplementary Table S4. Parameters resulting from fitting of the model of miRNA-mediated regulation to the analytical flow cytometry data generated from HEK293 cells**

$\lambda$  is inversely proportional to  $K_{on}$  and  $\theta$  is proportional to miRNA expression level. For details see materials and methods (RSS=Residual sum of squares).

Sample	$\lambda$	$\theta$	RSS
Control*	1.020	2.704	4.069
Empty*	0.9843	2.6989	4.196
miR-92a*	0.2773	5.6323	5.051
ACA45 1 copy	1.114	Constant ( $\theta= 2.7$ )	4.287
ACA45 3 copies	0.741		5.439
U3 1 copy	1.123		3.31
U3 3 copies	0.627		3.574
Mock	Constant mir92a ( $\lambda=0.27$ )	8.882	3.886
Outcompeted		6.579	3.876
Mock	Constant ACA45 ( $\lambda=0.27$ )	3.343	6.005
Outcompeted		3.071	3.702
Mock	Constant U3 ( $\lambda=0.27$ )	2.901	4.033
Outcompeted		2.584	9.716

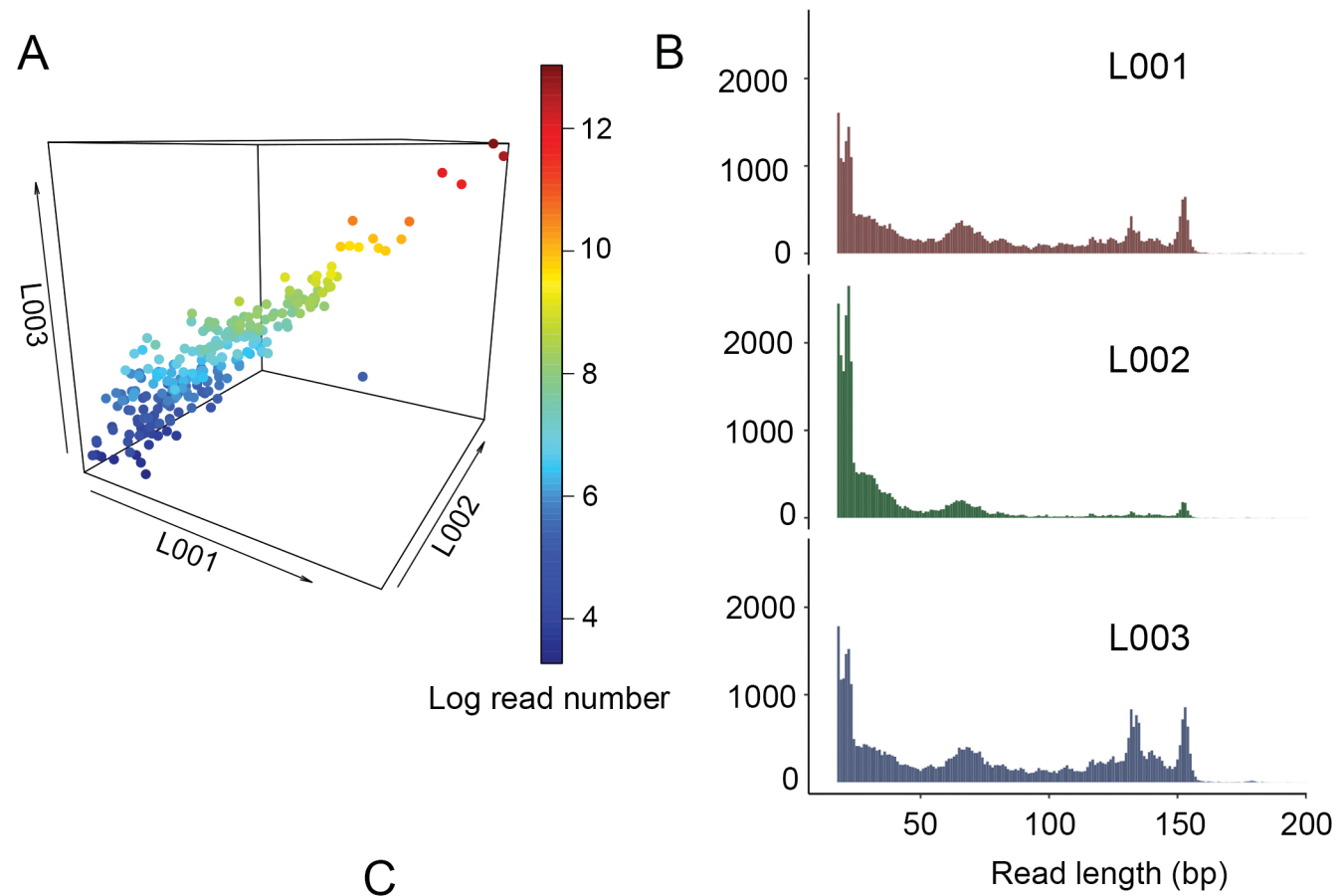
\*Empty, non-targeted and miR-92a used to calculate the initial values of parameters for constant  $\lambda$  and  $\theta$  for ACA45 and U3

**Supplementary Table S5. Parameters resulting from fitting of the model of miRNA-mediated regulation to the analytical flow cytometry data generated from HTC116 cells**

Lambda parameters were kept constant to determine relative expression using Theta.

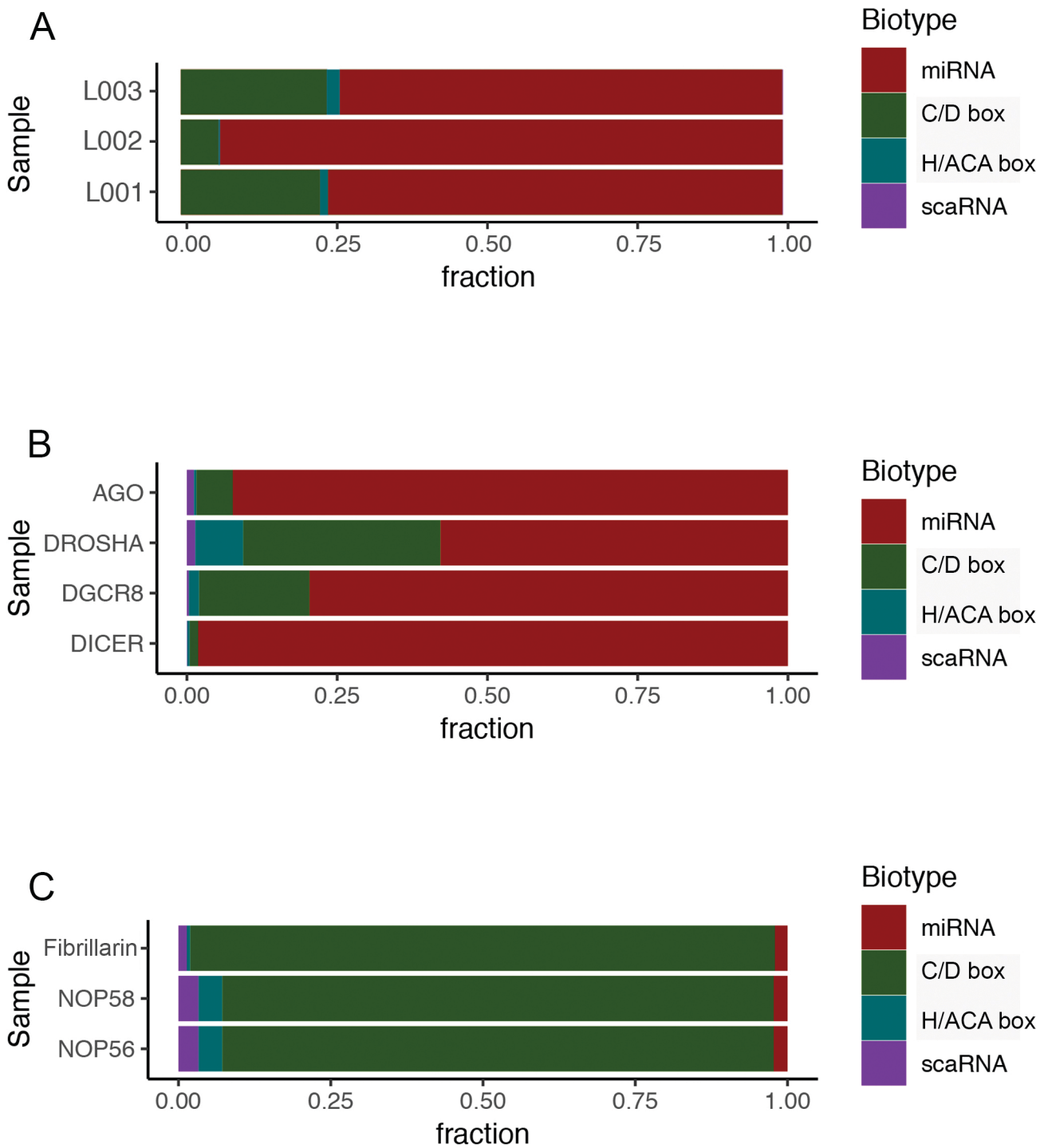
(RSS=Residual sum of squares)

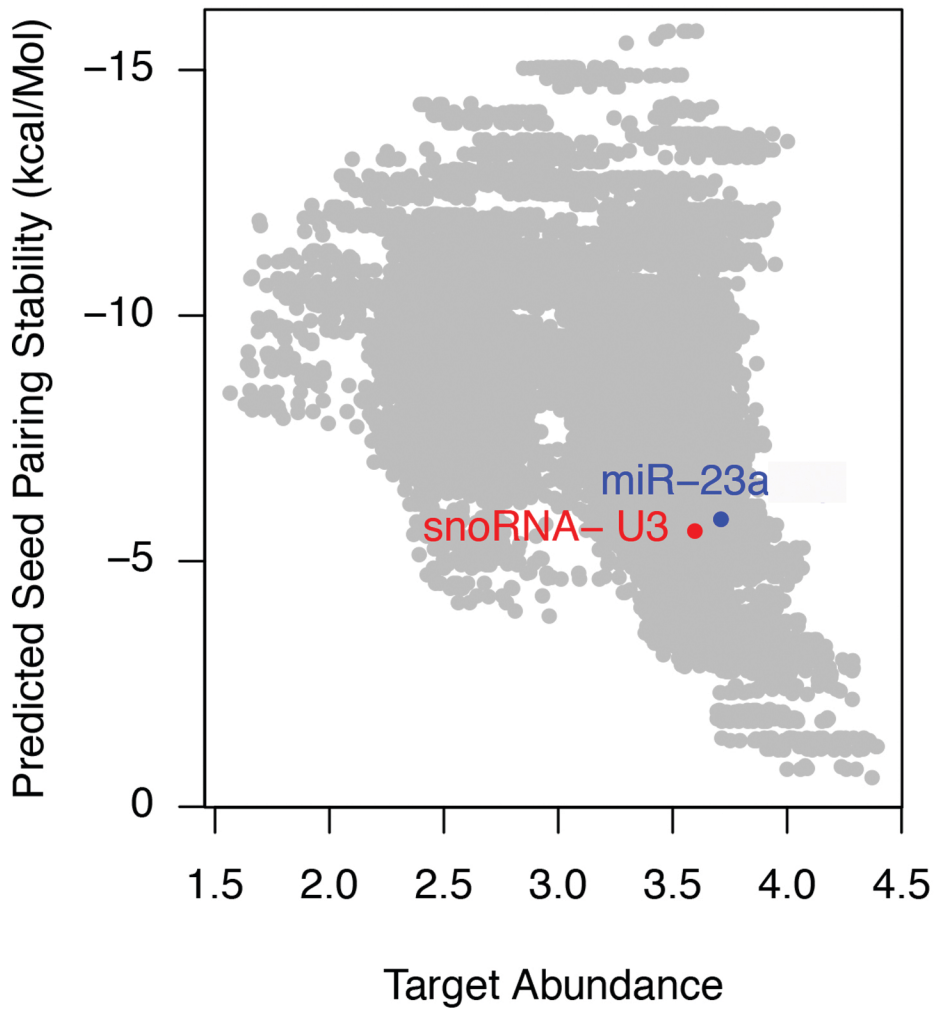
	WT		DROSHA KO		DICER KO	
	$\theta$	RSS	$\theta$	RSS	$\theta$	RSS
miR-21-5p	4.760	0.474	0.685	0.855	2.766	2.691
miR-92a	1.721	5.343	0.997	0.960	0.698	1.535
miR-320a	0.851	4.678	2.070	6.783	0.388	2.691
U3-1 copy	0.400	1.601	0.876	0.966	0.393	1.088
U3-3 copies	0.445	1.601	1.169	1.622	0.469	1.552



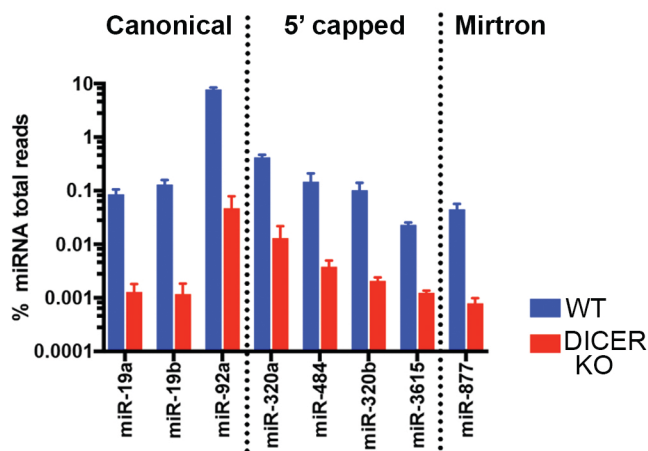
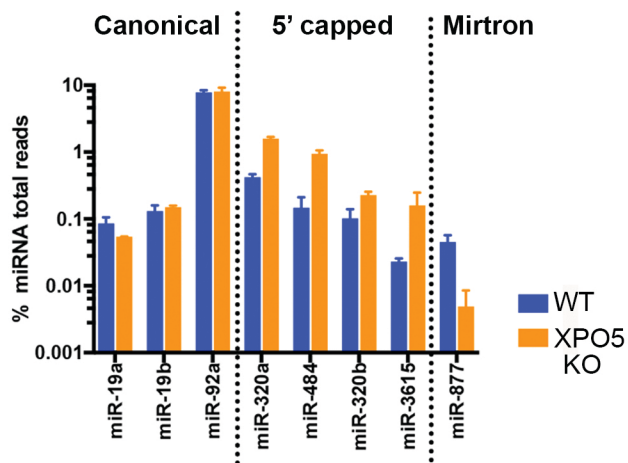
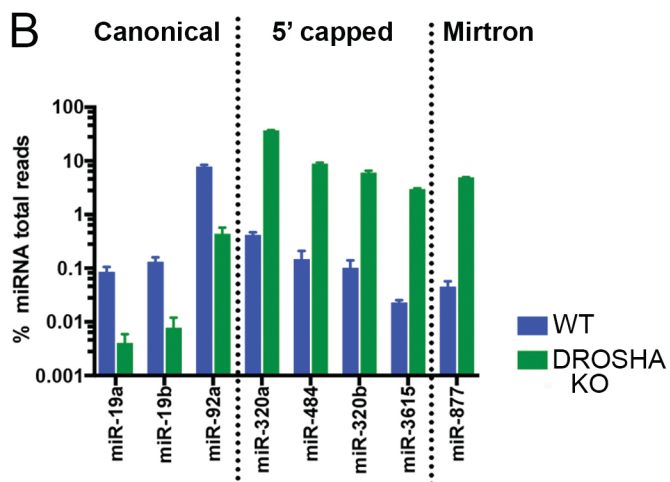
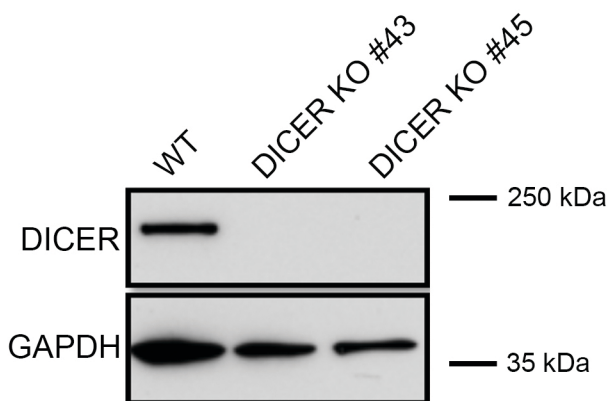
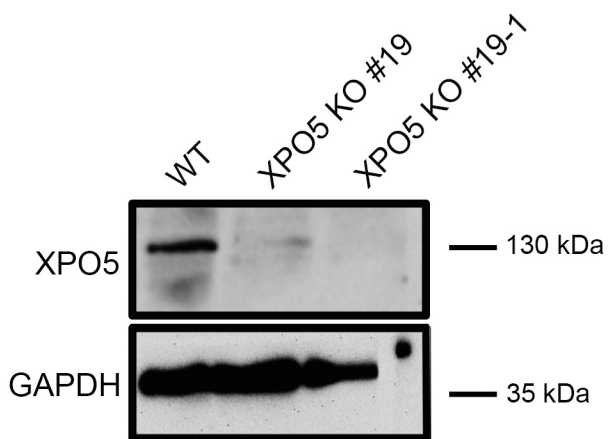
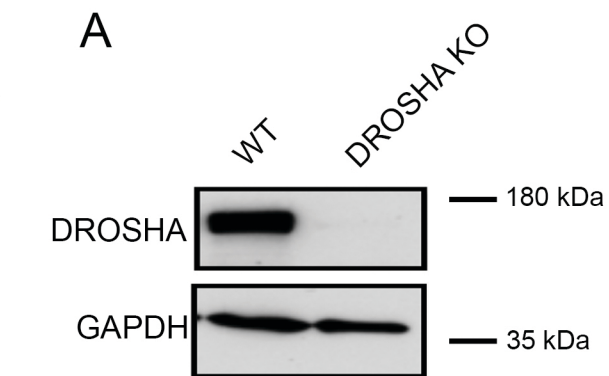
Lemus-Diaz *et al.* Supplementary Figure S1



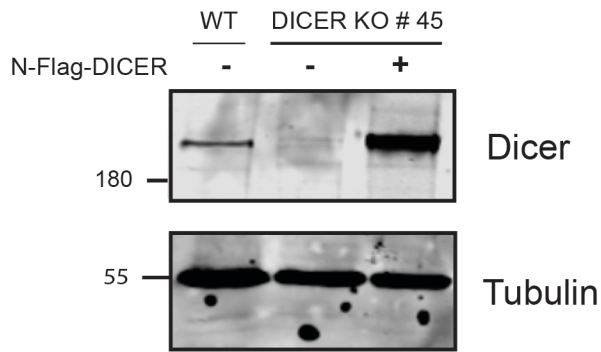




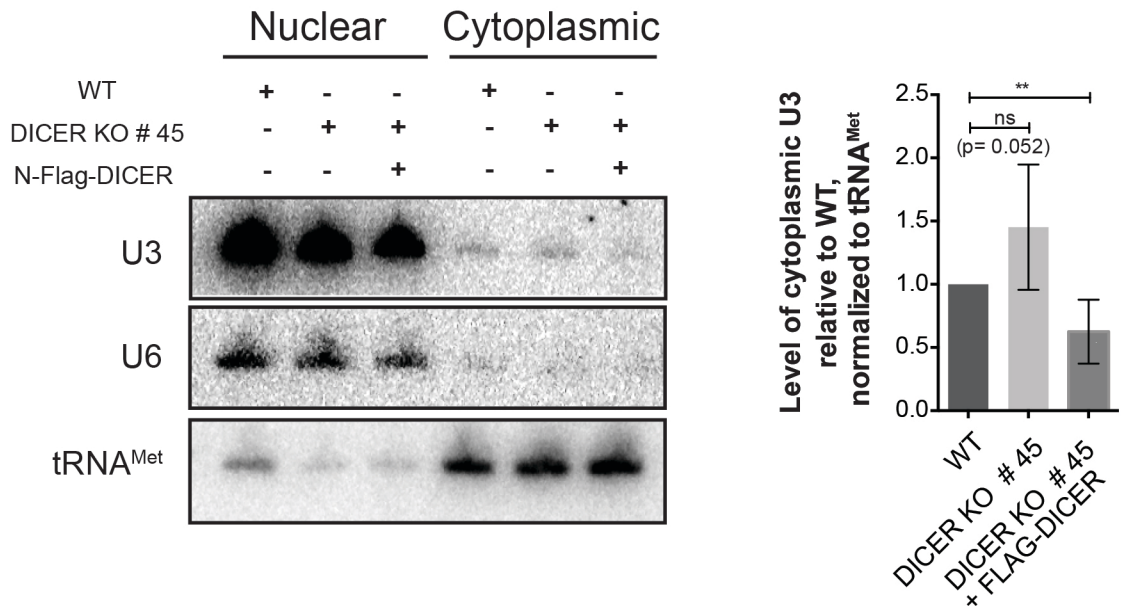
Lemus-Diaz *et al.* Supplementary Figure S3



**A**



**B**



Lemus-Diaz *et al.* Supplementary Figure S5

# SNX27 UTR

9201

9280

H. sapiens	UUU-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUCAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
P. troglodytes	UUU-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUCAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
M. mulatta	UUU-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUCAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
D. novemcinctus	CUC-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUCAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
L. africana	CUC-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUGAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
M. musculus	UUU-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUGAGAGAAAC	-UAAUUGU	-AUUUUCACUGCAG
R. norvegicus	UUU-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUGAGAGAAAC	-UAAUUGU	-AUUUUCACUGCAG
O. cuniculus	UUU-----	CCCAGGCCAUCCACAUUUAAUUU	UUUUGAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
C. lupus familiaris	UCC-----	CCCAGGCCAUGCACAUUUAAUUU	UUUUGAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
B. taurus	UUC-----	CCCAAACCAUGCACAUUUAAUUU	UUUUGAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
E. telfairi	CUC-----	CCCAGGCCGUGCACAUUUAAUUU	UUUUGAGAGAAACUC	-UAAUUGU	-AUUUUCACUGCAG
G. gallus	-----	-----	-----	-----	-----
X. tropicalis	-----	-----	-----	-----	-----

# snoRNA U3

2

112

H. sapiens	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
P. troglodytes	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
M. mulatta	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
D. novemcinctus	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
L. africana	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
M. musculus	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
R. norvegicus	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
O. cuniculus	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
C. lupus familiaris	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
B. taurus	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUUACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
E. telfairi	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUCGUGACUAGAGA	AGUUUCU	CUGAACGUGUAGAG	CACCGAAAACCACG
G. gallus	AGGACUAUACUUUCAGGGAUCAUUUCUGUAGUUGGUAACUAGAGG	AGUUGU	GAUCGAAAGUGU	CCGGCGUCACAAACC
X. tropicalis	AAGACUAUACUUUCAGGGAUCAUUUCUAUAGUUGUACCGUGUGA	AGUUGU	GCGUCUAAAGUGU	CUGAACUCACAAACC

Lemus-Diaz *et al.* Supplementary Figure S6