Supplementary Information for

A set of microRNAs coordinately controls tumorigenesis, invasion, and metastasis

Iacovos P. Michael[a], Sadegh Saghafinia[a,b], Douglas Hanahan[a,1]

[a] Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, 1015, Switzerland.

[b] Department of Computational Biology, University of Lausanne (UNIL), Lausanne, 1015 Switzerland

[1] To whom correspondence should be addressed. Email: douglas.hanahan@epfl.ch

This PDF file includes:

Supplementary Materials and Methods

Supplementary Figures S1-S9

Supplementary Datasets Legends

Supplementary Table S1

Supplementary References

Other supplementary material for this manuscript includes the following:

Supplementary Datasets S1-S5

Materials and Methods

The Bio-miRTa Algorithm

Input data: Initially, Bio-miRTa requires the miRNA ID (or IDs) and the species (human or mouse). Subsequently gene sets from functional perturbations, such as miRNA gain-of-function and loss-of-function studies, as well as from descriptive data, such as gene expression in different stages of cancer progression or other biological conditions being compared, is incorporated for the ranking of candidate gene targets in the second step.

Step 1: miRNA target extraction:

To identify potential targets of miRNAs, Bio-miRTa uses four different target prediction algorithms:

1. TargetScan (targetscan.org)
2. miRanda (microrna.org/microrna/getDownloads.do)
3. DIANA (diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index)
4. PITA (genie.weizmann.ac.il/pubs/mir07/index.html)

Computational target prediction algorithms predict numerous potential target genes for each miRNA, with a propensity of high number of false positives(1, 2). Therefore, to reduce the number of false positive targets, apart from including four different datasets, we also included two additional datasets of experimentally validated miRNA targets to the Bio-miRTa pipeline:

5. TarBase (diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index)
6. starBase (starbase.sysu.edu.cn/).

For each miRNA, Bio-miRTa extracts the predicted target genes in all six databases, and then, for each gene, a Prediction Score (PS) is defined as following:

$$PS_x = \sum_{i=1}^{6} PS_i \times Coeff.factor_{xi}$$

Where PS_i is the prediction score of dataset i (= 0.15 for prediction algorithms and 0.2 for experimentally validated datasets), and $Coeff.factor_{xi}$ is equal to 1 if and only if gene x is presented in dataset i, otherwise it is equal to 0. The maximum achievable PS_x is equal to 1.

As a result, a gene is considered as the potential target of the miRNA if $PS_x \geq 0.3$, *i.e.*, it is predicted to have an MRE by at least two target prediction algorithms, or one prediction algorithm and one experimentally verified database (**Fig. 4**).

Additionally, Bio-miRTa extracts the number of predicted MREs for the queried miRNA within 3'-UTR of the predicted target genes. This information is extracted for each target prediction algorithm separately, and reported in the final table.

In the cases where multiple related miRNAs with highly-similar or identical seed sequences are studied, such as miRNA clusters, Bio-miRTa runs the first step for each miRNA separately, and then takes the union of all target genes as the list of potential targets for the miRNA cluster. The final PS for a gene x is computed as follows:

$$PS_x = \frac{\sum_{i=1}^{n} PS_x^i}{\mathrm{n}}$$

where PS_x^i is PS_x for miRNA i, and n is the total number of miRNAs in the cluster (**Fig. S4**).

At the end of the first step, Bio-miRTa provides a list of potential target genes for the miRNA(s) of interest, which are ranked according to the PS.

Step 2: Experimetal dataset incorporation and final ranking of the candidate target genes.

In the second step Bio-miRTa has the ability to incorporate multiple datasets that are assigned with a Biological Score (BS) according to their inferred biological relevance. Eventually Bio-miRTa calculates the final BS for each gene as the following:

$$BS_x = \sum_{i=1}^{n} BS_i \times Coeff.factor_{xi}$$

Where BS_i is the biological score of dataset i (defined by the user), n is the number of datasets provided by the user, and $Coeff.factor_{xi}$ is equal to 1 if and only if gene x is presented in dataset i, otherwise it is equal to 0.

As a default, Bio-miRTa assigns a BS of 1 to the submitted datasets in the second step. Therefore, the BS of a given dataset would always be equal or greater than the PS. In additon, in the case of multiple datasets, Bio-miRTa allows the user to assign different BS to each dataset

according ot their perceived importance and biological relevance. This is illustrated in the section below describing the application of Bio-miRTa in this study and in the Results of the main text.

Finally, the list of potential target genes is further ranked by assigning a Final Score (FS) to each gene, which is the sum of two independent scores: the Prediction Score (PS) and the Biological Score (BS).

Additionally, Bio-miRTa allows the converse analysis, in which a potentially miR-regulated gene is known, and the question to be addressed is whether there are high scoring miRNAs that target MREs in that gene, identified in a reverse step 1, and also score in a reverse step 2, where biologically relevant miRNA expression datasets are incorporated.

Application of Bio-miRTA for the identification of miR-137 and miR-23b cluster biologically relevant gene targets, and *Inhßb* regulatory miRNAs.

We ran Bio-miRTa on miR-137 and the miR-23b cluster, which is composed of miR-23b, miR-27b, and miR-24-1. The analysis was performed considering the mouse as the selected species for target prediction. We also provided Bio-miRTa with the following three datasets of biologically relevant information:

1. The set of significantly down-regulated genes (more than 1.2-fold) upon over-expression of the miRNA in ßTC3 cell line (BS = 2).
2. The set of significantly down-regulated genes (more than 1.5-fold) in MLP versus IT tumor samples dissected from human patients (BS = 1).
3. The set of significantly down-regulated genes (more than 1.5-fold) in MLP versus IT tumor samples dissected from RipTag2 model (BS = 1).

The rational for the different weighting scores is explained in the Results section. The final lists for the miR-137 and the miR-23b cluster's biologically relevant target genes are shown in Datasets S2 and S3, respectively.

To assess the potential miRNAs targeting Activin-B (*Inhßb*), we ran Bio-miRTa in reverse for *Inhßb,* considering the mouse as the selected species. Thus, for step 1, we queried each of the six algorithms to identify miRNAs that define MREs within the 3' UTR for *Inhßb*, differentially assigning scores as described above to a maximum value of 1.0, if all six were concordant. To

calculate the BS, we used a single set of significantly up-regulated miRNAs (more than 1.5-fold) in PanNETs compared to normal pancreatic islet cells (3). The BS for this miRNA dataset was therefore defined as BS=1. The complete list of miRNAs implicated as modulators of *Inhßb* is shown in Dataset S4, while the top scoring miRNAs are shown in Figure S5.

Animal Studies

SCID/beige mice (from Taconic) were maintained in a pathogen-free barrier animal facility according to the Swiss regulations, and male, 8-10 weeks old, mice were used for all studies.

For orthotopic tumor models, fifty thousand ßTC3 cells in 50 µl of sterile PBS were injected in the parenchyma of the pancreas in a region proximal to the spleen.

As we have previously published(4), neuroendocrine cancer cells preferentially metastasize to the liver following tail vein injections; thus, to model the late stages of liver metastasis, we mainly used this route of inoculation instead of injections into the portal vein or spleen. One hundred thousand cells for both the ßTC3 and STC-1 in 200 µl of sterile PBS were injected into a lateral tail vein. To model the early stages of liver metastasis, we used intra spleen injections to deliver the cells via the portal vein circulation to the liver. One hundred thousand cells in 100 µl of sterile PBS were injected into the exteriorized spleen. The spleen was removed four minutes after injection to allow enough time for the cells to enter the circulation.

EdU Cell Proliferation Assays

The proliferation of cancer cells in-vivo was performed as we have previously described(4). Briefly, mice were injected intraperitoneally with 100 µg of EdU in PBS. One hour post-injection, they were anesthetized using pentobarbital sodium and transcardially perfused with 20 ml of PBS, followed with 20 ml of 1% PFA.

Mouse mRNA Sequencing

The RNA quality of the samples was assessed using the Fragment Analyzer (Advanced Analytical Technologies, Inc., Ankeny, IA, USA), and all samples had RQN > 9.0. We used 500 ng of total RNA to prepare the RNA-seq libraries using the Illumina TruSeq Stranded mRNA reagents

(Illumina; San Diego, California, USA). Subsequently, we used the Illumina TruSeq SR Cluster Kit v4 reagents for cluster generation and Illumina HiSeq 2500 for sequencing, generating stranded single-end reads of 100 nucleotides. The bcl2fastq Conversion Software was used to demultiplex the sequencing data (v. 2.20, Illumina; San Diego, California, USA) (5).

Cutadapt (v. 1.3) was used to remove adapter sequences and to trim the reads, while seq_crumbs (v. 0.1.8) was used to filter for low complexity. Subsequently, STAR (v. 2.4.0g) was used to align the reads against the Mus musculus.GRCm38.82 genome. The number of read counts per gene locus was summarized with htseq-count (v. 0.6.1) using the Mus musculus.GRCm38.82 gene annotation. RSeQC (v. 2.3.7) was used to assess the quality of the RNA-seq data alignment. Finally, STAR (v. 2.4.0g) was used to align the reads to the Mus musculus.GRCm38.82 transcriptome and RSEM (v. 1.2.19) to estimate the isoforms abundance(6).

DNA synthesis

The following fragments were synthesized using GeneArt Gene Synthesis service from Thermofisher:

A. *SORL1*-3'UTR Fragment: GRCh38/hg38 chr11:121,633,282-121,633,762; 481 bp. The MRE-mutant was also synthesized according to the sequence shown in Figure S4.
B. *ALK7*-3'UTR Fragments: The three fragments (Fr-A_miR-23b MRE: GRCm38/mm10 chr2: 58,273,978-58,274,903; 926 bp, Fr-B_miR-23b MRE: GRCm38/mm10 chr2: 58,269,120-58,270,042; 923 bp, Fr-C_miR-27b MRE: GRCm38/mm10 chr2: 58267587-58268507; 921 bp) containing the MRE for each miRNA were synthesized.
C. *Inhßb*-3'UTR Fragment: GRCh38/hg38 chr1:119,415,522-119,416,361; 840 bp. The MRE-mutant was also synthesized according to the sequence shown in Figure S5.

Each fragment contained AttB1 and AttP1 sites to facilitate subsequent cloning in the pDNR221 vector.

MicroRNA and cDNA cloning

A. miR-23b cluster: The genomic area (GRCm38/mm10 chr13:63,300,232-63,301,552; 1320 bp) encoding for the miR-23b cluster (including 251 bp downstream of miR-23b, and 277 bp upstream of miR-24-1) was PCR amplified from genomic DNA isolated from C57Bl6/N wild type mice. The Forward and Reverse primers contained AttB1 and AttP1 sites to facilitate subsequent cloning in the pDNR221 vector. The pENTR-miR-23b cluster was used to subclone the cluster to the PB-31 destination vector.
B. miR-137: The genomic area (GRCm38/mm10 chr3:118,433,746-118,434,058; 313 bp) encoding the miR-137 (including 110 bp downstream, and 129 bp upstream of miR-137) was PCR amplified and cloned as described above for the miR-23b cluster.
C. miR-130a/b and miR-301a/b: The following genomic areas were PCR amplified and cloned in the pDNR221 vector: miR-130a (GRCm38/mm10 chr2:84,740,815-84,741,478; 664bp), miR-130b (GRCm38/mm10 chr16:17,123,810-17,124,399; 590 bp), miR-301a (GRCm38/mm10 chr11:87,112,704-87,113,389; 686 bp), and miR-301b (GRCm38/mm10 chr16:17,124,143-17,124,796; 654 bp).
D. 3'UTR fragments: The 3'UTR fragments were cloned in pDNR221, and the resulting pENTR vectors were used for subcloning into the pcDHA-effLuc-RfA destination vector.
E. *ALK7*-3'-UTR MRE mutants: For the derivation of MRE mutants we used the Gibson assembly strategy. Briefly, the pENTR vector containing each fragment of the 3'UTR was digested with the appropriate enzymes (Fr-A; XhoI and EcoRI, Fr-B; StuI and NdeI, Fr-C; XbaI and BsaBI), and the PCR products of the primer pairs listed in the supplementary table S1 were used for assembly. The resulting pENTR vectors were used for subcloning into the pcDHA-effLuc-RfA destination vector.

Knockdown experiments

For the generation of gene knockdown, we generated vectors coding for tandem miR-E based shRNAs as we have previously described (4).

*piggy*Bac vectors

*Piggy*Bac transposon vectors used in this study have been previously described (4).

Transfections

One million cells were plated in 6-well plates 24 hrs before transfection. The following day the media was changed 2 hrs before transfections. The following liposomal complexes were set-up for transfections:

a. Single *piggy*Bac vector: 4.5 µg PB#1 + 1.5 µg PBase; 12 µl of Lipofectamine 2000
b. Two *piggy*Bac vectors: 2.25 µg PB#1 + 2.25 µg PB#2 + 1.5 µg PBase; 12 µl of Lipofectamine 2000

The complexes were incubated for 20-30 min at room temperature and then applied dropwise to the cells. Media was changed the next day. Stable cell lines were selected for resistance to the following antibiotics: Geneticin (G418; 1mg/ml), Hygromycin B (200 µg/ml), Puromycin (2 µg/ml), and Blasticidin (4 µg/ml).

Transient transfection with miRCURY LNA antagomirs

Transfections with antagomirs (miRCURY LNA; QIAGEN, MMU-MIR-27A-3P, MMU-MIR-27B-3P, MMU-MIR-23A-3P, MMU-MIR-23B-3P, Negative control A) were done using Lipofectamine RNAiMAX Reagent (Thermo Fisher Scientific).

Briefly, one million cells were plated in 6-well plates 24 hrs before transfection. The following day, the cells were transfected with 40nM of each antagomir MMU-MIR-*XYZ*-3P, or 160nM of negative control, along with 1µl of Lipofectamine RNAiMAX Reagent. The cells were collected in QIAzol lysis buffer 72hrs post transfection.

Cell culture treatment with recombinant proteins (cytokines) and antagomirs

Upon arrival, activin B (R&D Systems; Cat# 8260-AB-010) was resuspended according to the manufacturer's instructions, aliquoted and stored at -20^{o}C. Care was taken not to thaw the aliquots more than one time.

One million cells were plated in 6-well plates 24 hrs before addition of recombinant proteins. The following day, the media was changed, and the appropriate amount of recombinant protein was added.

Invasion assays

Invasion assays were performed as previously described (7). ßTC3 control or miR-23b cluster overexpressing cells were used in DOX containing media.

Reporter assays

The vector destination vector pcDHA-effLuc-RfA was made for assessing the activity of MREs in the 3'UTR of *ALK7*. Briefly, the enhanced Firefly Luciferase was cloned in the commercial pcDNA3.1/hygro vector downstream from the CMV promoter, followed by the Gateway cassette to facilitate the cloning of the fragments. The MRE containing fragments were synthesized and then cloned in the vector pcDHA-effLuc-RfA.

The 293T cells were first stably transfected with the vector PB-13/PB-RB, allowing doxycycline-inducible expression of the miR-23b cluster. For the reporter assays, 50 000 293T cells expressing the miR-23b cluster were plated in 24-well plates in basic DMEM media/10%FS. The following day the media was changed to media containing DOX (2 µg/ml) 1 hr before transfections. The following liposomal complexes were set-up for transfections:

a. 25 ng pcDHA-effLuc-UTR fragment + 100 ng LacZ + 125 ng PB-EGFP; 0.5 µl of Lipofectamine 2000
b. 5 ng pcDHA-effLuc-UTR fragment + 50 ng LacZ + 200 ng PB-EGFP; 0.5 µl of Lipofectamine 2000

The complexes were incubated for 20-30 min at room temperature and then applied dropwise to the cells. After 24 hrs of adding the complexes, the media was changed to fresh DOX media. Forty-eight hours after the luciferase and ß-galactosidase activity was measured according to the manufacturer's instructions. Briefly, the 24-well plates were washed with cold PBS and 120 µl of CCLR lysis buffer was added to each well. The plates were incubated on ice for 5 min, followed by 3 rounds of Freeze/thaw cycles.

For the luciferase assays, 20 µl from the cell lysate was transferred to each well of a 96-well white plate in duplicates. Forty microliters of the luciferase substrate were dispensed using the luminometer plate reader, and after 2 sec delay, three measurements were taken at 1, 5 and 10 sec.

For the ß-galactosidase assay, 20 µl from the cell lysate was transferred to each well of a 96-well clear plate in duplicates. Thirty microliters of 1xCCLR plus 50 µl of assay buffer were added. The plate was incubated at 37^{o}C for a maximum of 30 min, and the color development was monitored. At the appropriate time point, the reaction was stopped with 150 µl of Tris buffer per well. The absorbance in each well was measured at 420 nm.

Quantitative reverse transcription PCR

Total RNA was isolated using the QIAzol Lysis Reagent according to the instructions.

A. For microRNAs, one microgram of total RNA was used for reverse transcription using the miScript II RT Kit with the Hi-Flex buffer. Q-PCR was performed using miScript SYBR Green.
B. For genes, five hundred nanograms of total RNA were used for reverse transcription using the PrimeScript RT Master Mix. Q-PCR was performed using the Rotor-Gene SYBR Green.

Primer design for Q-PCR

PrimerBank was used as the source for the primers used for PCR quantitation of genes (8).

Western blots

Western blots were performed as has been previously described (REF). Briefly, 10-20µg of total protein extract were separated using sodium dodecyl sulfate polyacrylamide gel electrophoresis and subsequently transferred onto PVDF blotting-membrane. Subsequently, the membranes were blocked with 5% Bovine Serum Albumin Fraction V (BSA) in TBS-T (pH 7.6; 0.5% Tween 20), and incubated with the primary and secondary antibodies diluted in 5% BSA in TBS-T. Finally, the membranes were visualized using ECL and imaged with Fusion FX7. The following antibodies

were used: phospho-Ser465/467 Smad2 (Cell Signaling; Cat# 3108), Smad2 (Cell Signaling; Cat# 5339), SORL1 (Abcam; Cat#ab190684), and ALK7 (Sigma; Cat# HPA007982).

Immunostaining

ALK7 (Sigma; Cat# HPA007982) and T antigen (in-house antibody) chromogenic immunostaining was performed using the automated Ventana Discovery XT as has been previously described (4).

EdU/T antigen immunostaining: The immunofluorescence for T antigen and EdU was performed manually. After dewaxing and rehydration, sections were pretreated with heat in 10mM Tri Na citrate buffer pH 6 using the PT module (Thermo Fisher Scientific) for 20 minutes at 95°C. The anti-T antigen was diluted 1:1000 and incubated overnight at 4°C under agitation. After three washes, a donkey anti rabbit Alexa488 (Life Technology) was applied on the slide for 45 minutes at RT. Tissue was treated with 0.5% Triton in PBS1x for 20 minutes at RT before incubation with the Azide reaction cocktail (0.1M TBS pH7.4, 4mM CuSo4, 100mM Na-ascorbate and 10uM Azide Alexa594) to reveal the presence of EdU. DAPI was used to stain the nuclei.

CC3/T antigen immunostaining: After sample dewaxing and rehydration, heat-induced epitope retrieval was done with 1mM EDTA pH 8.0, and after washing, the samples were incubated overnight at 4°C with rabbit anti-CC3 (Cell Signaling; Cat# 9664) diluted 1:200 in 1%BSA. The next day, after washing, the samples were incubated for 40 min with the secondary donkey anti-rabbit HRP antibody diluted 1:1000 in 1% BSA, and Tyramide Signal Amplification (TSA) revelation was done with AlexaFluor568. After microwave treatment with 10mM Sodium Citrate ph6.0 and washing, the samples were incubated for 1 hr with rabbit anti-T antigen diluted 1:1000 in 1% BSA. After washing, the samples were incubated for 40 min with a donkey anti-rabbit AlexaFluor488 diluted 1:1000 in 1% BSA. Samples were washed, and nuclear staining was done with DAPI before mounting.

A

Met-like primary and metastasis miRNA signature				Metastasis-specific miRNA signature			
Up-regulated	fold-change	Down-regulated	fold-change	Up-regulated	fold-change	Down-regulated	fold-change
miR-449	8.4	miR-200a	-9.9	miR-431	3.8	miR-148a	-21.1
miR-181d	**4.7**	miR-141	-9.8	miR-189	2.9	miR-146a	-8.4
miR-137	**4.5**	miR-184	-7.3	**miR-23b**	**2.1**	miR-152	-5.3
miR-129	3.8	miR-182	-7.2	**miR-27b**	**2.1**	miR-10b	-4.0
miR-410	3.6	miR-429	-6.2	**miR-24**	**2.0**	miR-365	-3.5
miR-92	2.6	miR-200c	-5.2			miR-10a	-3.0
miR-132	**2.5**	miR-200b	-5.0			miR-148b	-2.9
miR-15b	2.3	miR-143	-3.8				
miR-124a	2.2	miR-145	-3.1				
miR-344	2.2	miR-126	-3.0				
miR-181a	**2.0**						

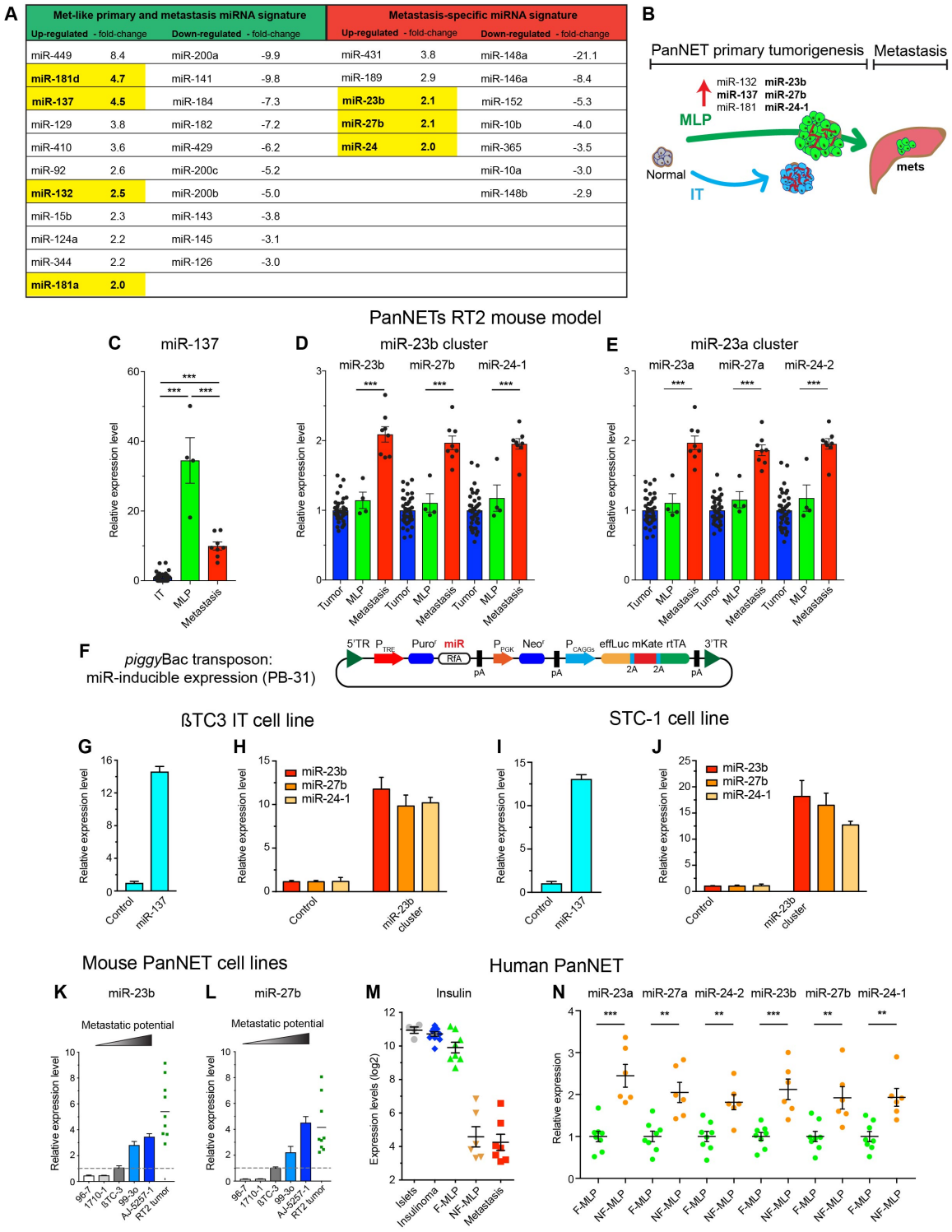


Fig. S1: Characterization of MLP-associated miRNAs during primary tumor growth.

(A) Table of upregulated and downregulated miRNAs in MLP and metastasis miRNA signature, and metastasis-specific miRNA signature in the RIP1-Tag2 mouse model of PanNETs. The miRNAs that are similarly altered in human PanNETs are highlighted yellow.

(B) Branched tumorigenesis pathway in the RIP1-Tag2 (RT2) mouse model of PanNET. MLP; metastasis-like primary, IT; islet tumor/insulinoma.

(C-E) Expression of miR-137 (C), the miR-23b cluster (D) and the = miR-23a cluster (E) in IT, MLP, and liver metastasis in the PanNETs RT2 mouse model. Note that miR-24-1 and miR-24-2 have identical sequence and therefore show the same expression levels. n = 8-37. Tukey's multiple comparison test. *** $p<0.001$.

(F) *PiggyBac* transposon system for inducible microRNA expression.

(G - J) Expression of miR-137 (G and I) and the miR-23b cluster (H and J) in ßTC3 (G and H) and STC-1 (I and J) cells upon doxycycline (DOX) induction (Control cells were also treated with DOX).

(K and L) Relative expression levels of miR-23b (K) and miR-27b (L) in RT2 PanNET-derived cell lines with varying metastatic potential, as well as primary RT2 tumors. Expression levels were normalized against the ßTC3 cell line. MiR-24-1 levels were not measured.

(M and N) Expression the miR-23a/b cluster of miRNAs in human PanNETs.

(M) Insulin expression in human PanNET subtypes; the MLPs were sub-divided into functional (F-MLP; insulin-high) and non-functional (NF-NLP; insulin low) sub-types.

(N) Expression of miRNAs in human F-MLP and NF-MLP PanNETs. n = 6-8. Students t-test. ** $p<0.01$, *** $p<0.001$.

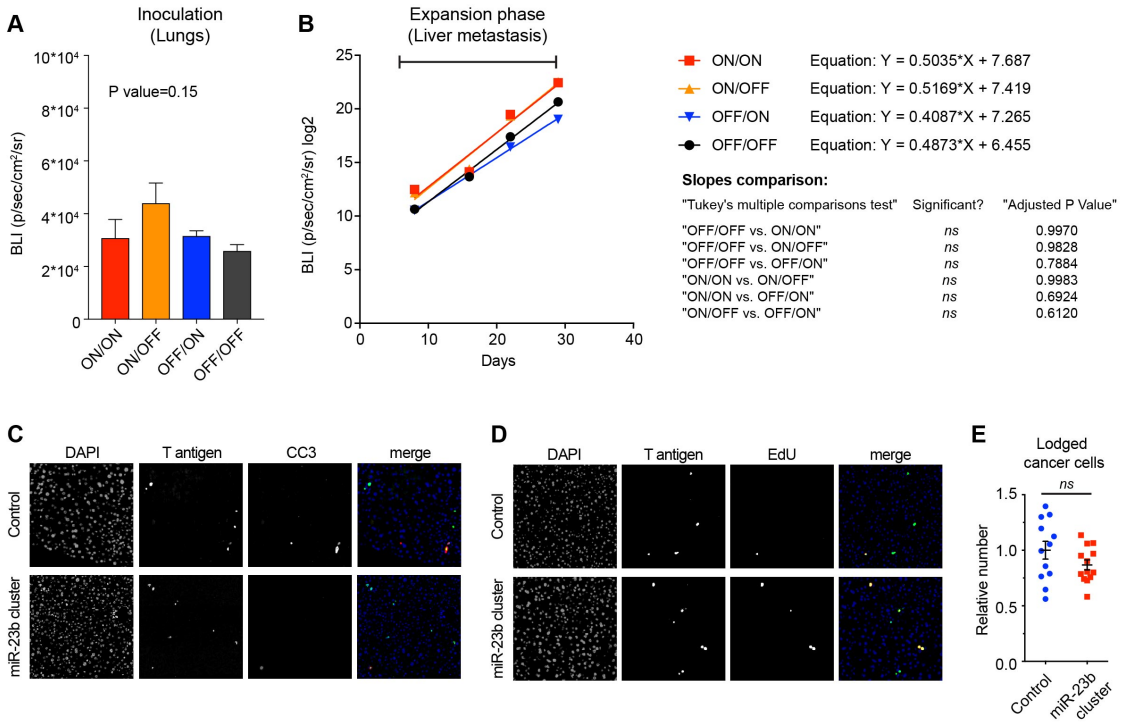


Fig. S2: Characterization of the miR-23b cluster role during metastasis.

(A) Quantification of bioluminescence signal in the lungs right after tail vein injections. n=5. Tukey's multiple comparison test.

(B) Linear regression and slope comparison of bioluminescence signal during the expansion phase of liver metastasis after tail vein injections (related to Figure 3). n=5. Tukey's multiple comparison test. *ns* = not significant.

(C, D) Representative images of immunostaining for the Tag oncoprotein plus EdU (C), and Tag plus Cleaved Caspase 3 (D), revealing lodged ßTC3 control and miR-23b overexpressing cells in the mouse liver, 24 hrs after intrahepatic injection. (Quantification is shown in Fig. 3 E and F).

(E) Quantification of cancer cells lodged in the liver. n = 10-12. Students t-test. *ns* = not significant.

Gene	Target Prediction Algorithms				Experimentally Validated Targets		PS	Differential Gene Expression	BS	FS
	TargetScan	miRanda	DIANA	PITA	TarBase	starBase		Mouse Double K/O		
ZEB1	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	**1.00**	1.32	**1**	**2.0**
MBNL1	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-141	miR-200c miR-141	miR-200c miR-141	**0.925**	1.42	**1**	**1.925**
DLC1	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-141	miR-200c miR-141	**0.90**	1.35	**1**	**1.90**
RECK	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-200c miR-141	miR-200c	miR-200c miR-141	**0.90**	1.72	**1**	**1.90**
OGT	miR-141	miR-200c miR-141	miR-200c miR-141	miR-141	miR-200c miR-141	miR-200c miR-141	**0.85**	1.42	**1**	**1.85**

miR-200c: refers to miR-200b/200c/429

miR-141: refers to miR-200a/141

Fig. S3: Validation of Bio-miRTa using a publicly available dataset for the miR-200 family.

Bio-miRTa was used to identify the candidate gene targets of the miR-200 family during tumorigenesis of the RIP1-Tag2 model. The transcriptome data with accession number E-MTAB-6717 were used for the analysis. The top five genes are shown; ZEB1, DCL1, RECK, and OGT were previously experimentally validated as miR-200 targets. Dataset S1 contains the complete list of ranked genes.

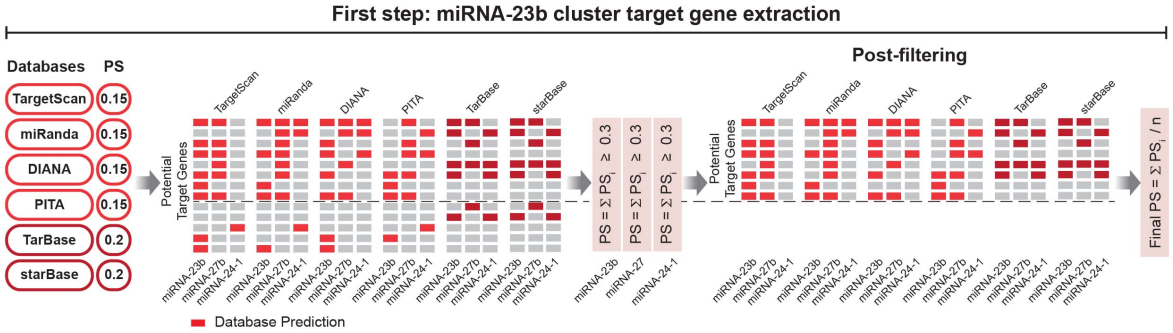


Fig. S4: Identification of potential gene targets for multiple miRNAs using Bio-miRTa.

Flow chart of the first step for the identification of potential gene targets in the case of multiple miRNAs, e.g., miR-23b cluster. The algorithm first calculates the PS score for each individual miRNA as described in the case of a single miRNA and uses the same cut-off filter of PS ≥ 0.3. Subsequently, in a "Post-Filtering" step, to calculate the final PS of the cluster for each gene, the sum of the PS scores for each miRNA that passes the PS ≥ 0.3 bar is normalized according to the total number (n) of miRNAs in the cluster, such that the maximal PS = 1 (also see Methods). The second step of the algorithm remains the same as in the case of individual miRNAs.

A

miRNA	Target Prediction Algorithms				Experimentally Validated Targets		PS	Differential miRNA Expression			BS	FS
	TargetScan	miRanda	DIANA	PITA	TarBase	starBase		GOF/ LOF	Mouse PanNET	Human PanNET		
miR-130b	Inhßb	Inhßb	Inhßb	Inhßb	Inhßb	Inhßb	**1.00**	NA	16.67	NA	**1**	**2.00**
miR-301a/b	Inhßb	Inhßb	Inhßb	Inhßb	Inhßb	Inhßb	**1.00**	NA	16.18	NA	**1**	**2.00**
miR-130a	Inhßb	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.80**	NA	9.42	NA	**1**	**1.80**
miR-153	Inhßb	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.80**	NA	2.73	NA	**1**	**1.80**
miR-19a	Inhßb	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.80**	NA	5.74	NA	**1**	**1.80**
miR-19b	Inhßb	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.80**	NA	11.31	NA	**1**	**1.80**
miR-302b	Inhßb	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.80**	NA	1.52	NA	**1**	**1.80**
miR-9	Inhßb	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.80**	NA	3.37	NA	**1**	**1.80**
miR-10a	Inhßb	Inhßb	Inhßb	-	-	Inhßb	**0.65**	NA	1.81	NA	**1**	**1.65**
miR-143	-	Inhßb	Inhßb	Inhßb	-	Inhßb	**0.65**	NA	4.06	NA	**1**	**1.65**

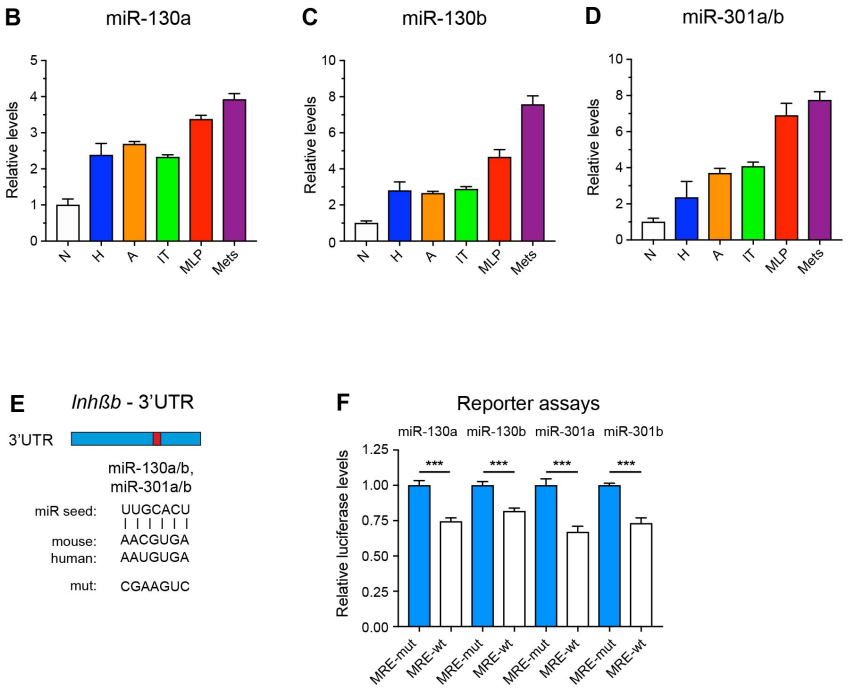


Fig. S5: Correlative evidence for regulation of activin B expression by a set of miRNAs during PanNET progression.

(A) Identification of potential miRNAs regulating *Inhßb* expression using Bio-miRTa.

The first column labeled as "miRNA" shows the candidate miRNA sorted according to the collective Final Score (FS) shown in the last column on the right. The columns with the headers "Target prediction algorithms" and "Experimentally validated targets" report the presence of an MRE in the *Inhßb* gene for the indicated miRNA. The columns with the header "Differential gene expression" show the differential miRNA expression in normal mouse islets versus islet tumors in RT2 mice

(Mouse PanNETs). There were no gain- or lost-of-function (GOF/LOF) datasets or human PanNET datasets available to be included in this analysis (NA; not-available). The intercalated columns show the PS (based on the presence of MRE according to the first step of the algorithm), BS (based on the differential gene expression), and FS (the sum of the PS and BS scores). The miRNAs with FS>1 are shown; Supplementary Tables S5 contains the complete list of ranked miRNAs.

(B-D) Expression of miR-130a/b and miR-301a/b during multi-stage tumorigenesis to invasive and metastatic PanNET in the RT2 mouse model.

(E) MREs for miR-130a/b and miR-301a/b in the 3'UTR of *Inhßb*. The miRNA-seed, the MREs for human and mouse *Inhßb*, as well as the mutated MRE used for the reporter assays, are shown.

(F) Luciferase reporter assays assessing functionality of the MREs. n = 6. Students t-test. *** p <0.001.

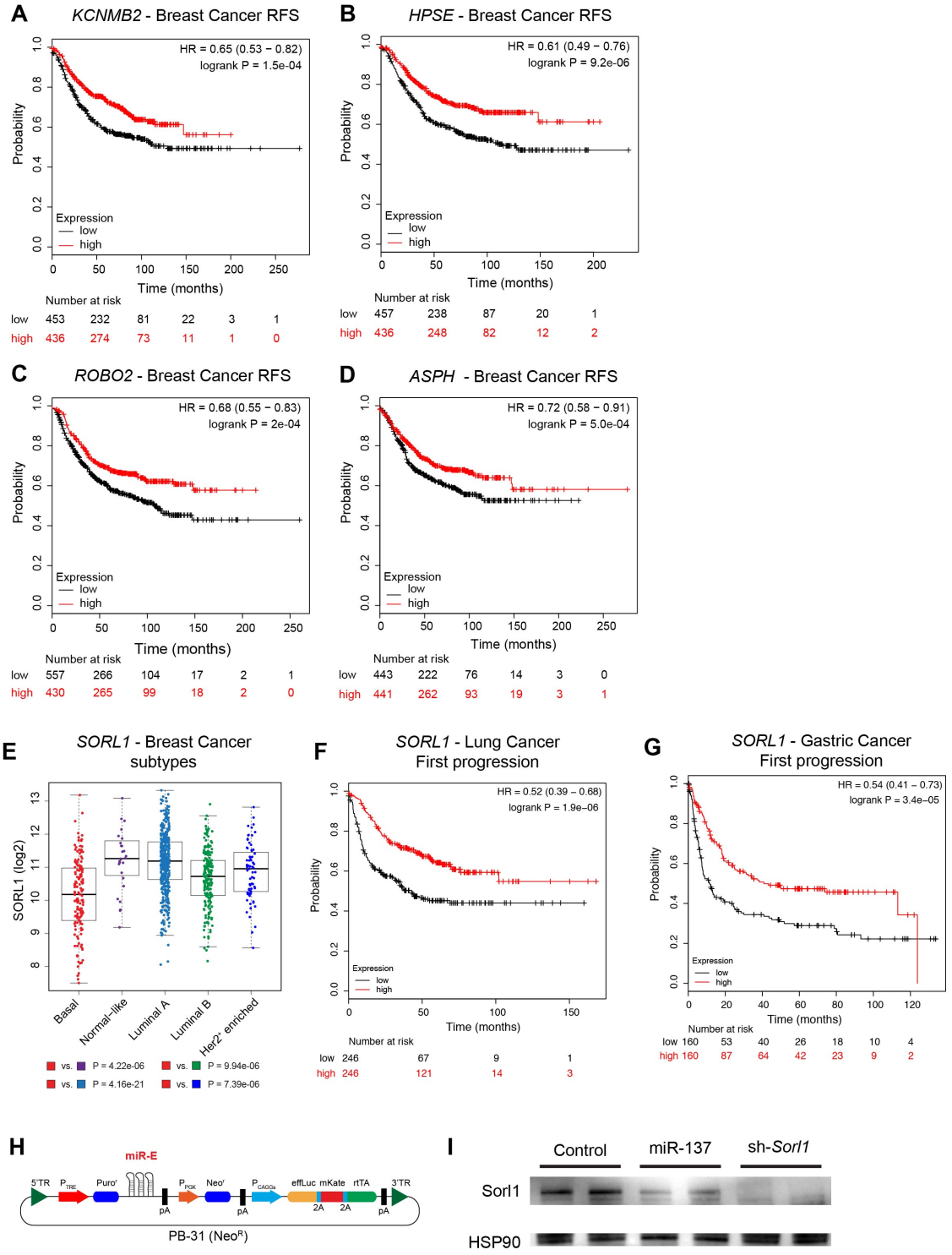


Fig. S6: Association of miR-137 candidate target genes with cancer prognosis.

(A-D) Association of miR-137 candidate target genes *KCNMB2* (A), *HPSE* (B), *ROBO2* (C), and *ASPH* (D) expression levels with time to progression of breast cancer patients including all subtypes.

(E) Boxplot comparison of *SORL1* mRNA expression across the five breast cancer PAM50 subtypes. P-values were computed using a two-sided ANOVA test. The number of samples for each breast cancer subtype are: Luminal A (n = 467), Luminal B (n = 198), Her2-enriched (n = 72), Normal-like (n = 28), Basal-like (n = 155). The thick central line of each box plot represents the median number of significant motifs, the bounding box corresponds to the 25th–75thpercentiles, and the whiskers extend up to 1.5 times the interquartile range.

(F and G) Association of *SORL1* expression levels with time to progression of TCGA cohorts of non-small cell lung (B) and gastric (C) cancer patients.

(H) *PiggyBac* transposon system for inducible gene knock-down expression.

(I) Sorl1 gene expression in control, miR-137 overexpressing, and *Sorl1* knock-down ßTC3 cells.

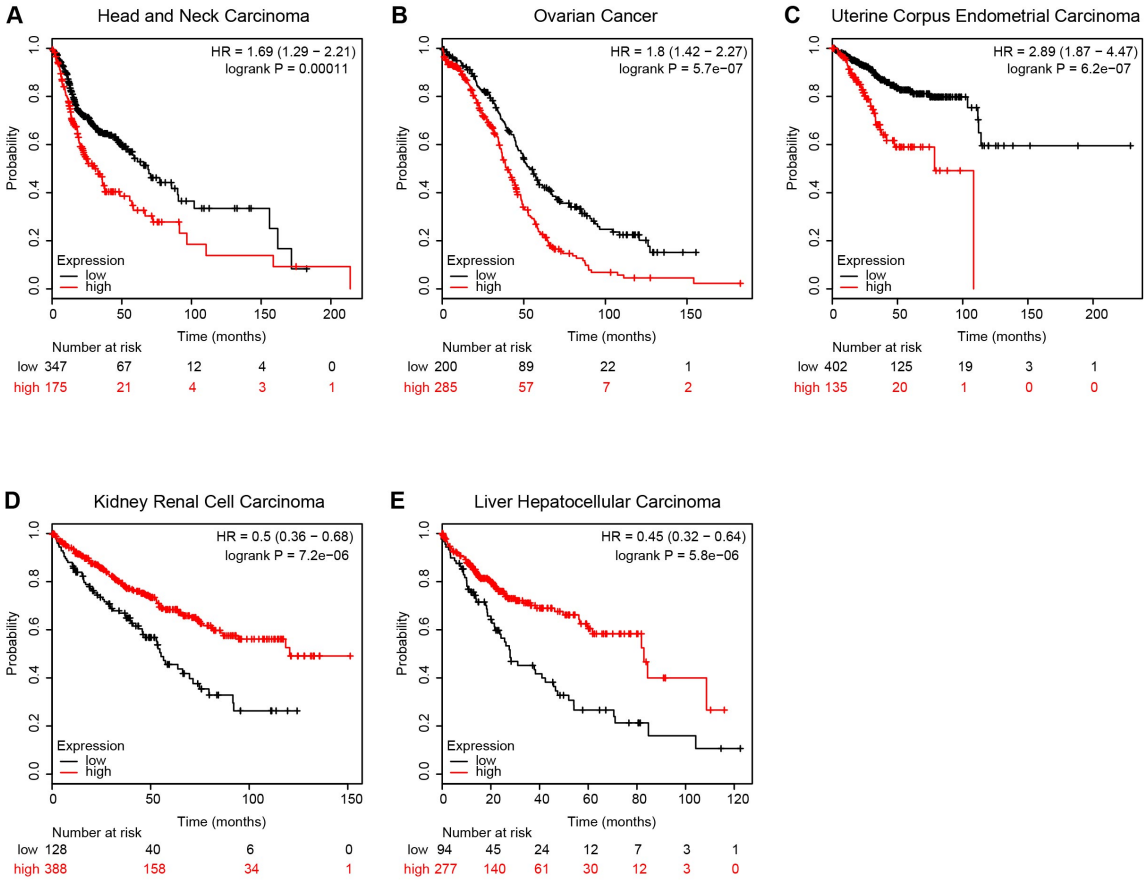


Fig. S7: Association of miR-137 expression levels with cancer prognosis.

Association of miR-137 expression levels with overall survival of patients with head and neck carcinoma (A), ovarian cancer (B), uterine corpus endometrial carcinoma (C), kidney renal cell carcinoma (D), and liver hepatocellular carcinoma (E).

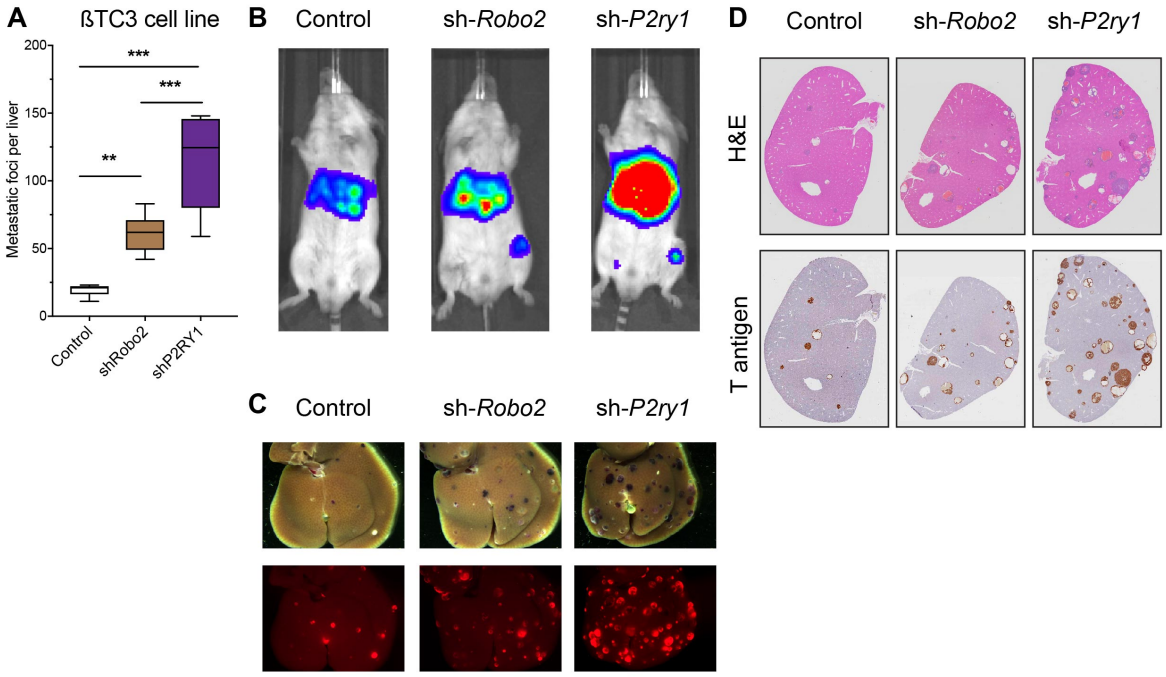


Fig. S8: Evaluation of Robo2 and P2ry1 as suppressors of PanNET liver metastasis.

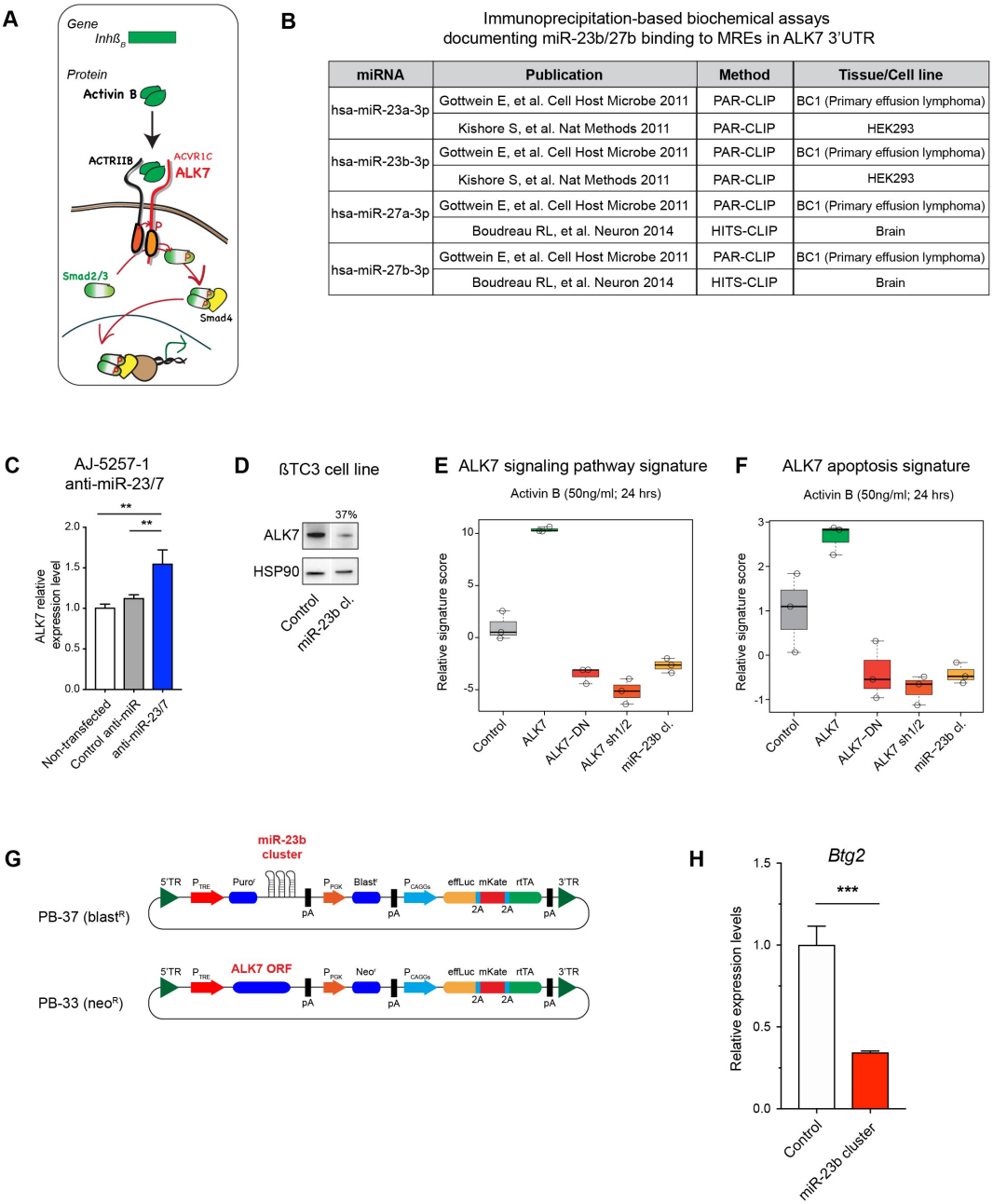
(A-D) Experimental liver metastasis assays, using control, and either *Robo2* or *P2ry1* knock-down ßTC3 cells.

(A) Quantification of metastatic foci. n = 6-7. One-way ANOVA followed by Holm-Sidak's multiple comparisons test. ** p <0.01, *** p <0.001.

(B) Bioluminescence imaging before euthanasia.

(C) mKate fluorescence imaging of excised livers.

(D) H&E staining and Tag oncoprotein immunostaining of liver sections.



miRNA	Publication	Method	Tissue/Cell line
hsa-miR-23a-3p	Gottwein E, et al. Cell Host Microbe 2011	PAR-CLIP	BC1 (Primary effusion lymphoma)
	Kishore S, et al. Nat Methods 2011	PAR-CLIP	HEK293
hsa-miR-23b-3p	Gottwein E, et al. Cell Host Microbe 2011	PAR-CLIP	BC1 (Primary effusion lymphoma)
	Kishore S, et al. Nat Methods 2011	PAR-CLIP	HEK293
hsa-miR-27a-3p	Gottwein E, et al. Cell Host Microbe 2011	PAR-CLIP	BC1 (Primary effusion lymphoma)
	Boudreau RL, et al. Neuron 2014	HITS-CLIP	Brain
hsa-miR-27b-3p	Gottwein E, et al. Cell Host Microbe 2011	PAR-CLIP	BC1 (Primary effusion lymphoma)
	Boudreau RL, et al. Neuron 2014	HITS-CLIP	Brain

Fig. S9: Characterization of miR-23b cluster mediated regulation of the ALK7 metastasis suppressor.

(A) Schematic diagram of the ALK7 signaling pathway.

(B) List of previous studies providing direct biochemical evidence via cross-linking immunoprecipitation assays for the interaction of miR-23a/b and miR-27a/b with *ALK7*.

(C) *ALK7* expression in MLP-like AJ-5257-1 cells treated with either control or miR23/7 anti-miRNAs. n = 6. Students t-test. ** p<0.01.

(D) *ALK7* expression in control and miR-23b cluster over-expressing ßTC3 cells.

(E, F) Relative score of generic ALK7 signaling associated (E) and ALK7 apoptosis-specific (F) gene expression signatures in ßTC3 control compared to either ALK7, ALK7-DN, ALK7-sh1/2, or miR-23b cluster overexpressing ßTC3 cells, after treatment with activin B (24hrs; 50ng/ml). (ALK7, ALK7-DN, ALK7-sh1/2 ßTC3 cell lines have been previously described (4).

(G) Dual *piggyBac* transposon system for concurrent transduction and inducible expression of the miR-23b cluster and *ALK7*.

(H) Relative expression levels of *Btg2* in ßTC3 IT control and miR-23b cluster expressing cells. n = 3. Students t-test. *** $p<0.01$.

Supplementary Datasets legends

Supplementary Dataset 1: Complete list of ranked miR-200 family target genes identified using Bio-miRTa.

Supplementary Dataset 2: Complete list of ranked miR-137 target genes identified using Bio-miRTa.

Supplementary Dataset 3: Complete list of ranked miR-23 cluster target genes identified using Bio-miRTa.

Supplementary Dataset 4: Complete list of Inhßb ranked target miRNAs identified using Bio-miRTa.

Supplementary Dataset 5: Transcriptomic ALK7 signaling pathway signatures.

Supplementary Table 1: Primers and oligos sequence.

Q-PCR primers for Genes		
Gene	Primer	Sequence
Mm_Acvr1c	Forward	5'-GTCTGGCTCACCTGCACAT-3'
	Reverse	5'-CAGCTATGGCACAAGTGTCAC-3'
Mm_Rpl13	Forward	5'-AGCCGGAATGGCATGATACTG-3'
	Reverse	5'-TATCTCACTGTAGGGCACCTC-3'
Q-PCR primers for miRNAs		
miRNA	Primer	Qiagen; Cat#
miR-23b	Mm_miR-23b_2	MS00032606
miR-27b	Mm_miR-27b_1	MS00001358
miR-24	Mm_miR-24_1	MS00005922
miR-137	Mm_miR-137_1	MS00001589
Primers for miR cloning		
miRNA	Sequence	
miR-137	Forward	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTGCTCCGC AGCAAGAGTTC-3'
	Reverse	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGGAAGATC CAGAACGAAACCAC-3'
miR-23b cluster	Forward	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGGAGAA CAGGGTGTGTCC-3'
	Reverse	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAGACA GGCATTCTCACTGC-3'
Primers for cloning of Acvr1c/ALK7 3'UTR mutated miR-23b cluster MREs		
Construct	Primer	Sequence
Fr-A_miR-23b MRE	Forward-1	5'-GCCAACTTTGTACAAAAAAGCAGGCTCTCGAGC-3'
	Reverse-1	5'-CCATTTCACCATAATATACAGTAGTagaaggTTTAAATAT GTCTTCATCTCATTTGTGG-3'
	Forward-2	5'-GAGATGAAGACATATTTAAAccttctACTACTGTATATTAT GGTGAAATGGAACTG-3'
	Reverse-2	5'-CCACTTGGGAAGAAACGCTCTATGGGAATTCAAG-3'
Fr-B_miR-23b MRE	Forward-1	5'-GGTTTTAATTTCCCCACTGTGTATAAGAGGCCTC-3'
	Reverse-1	5'-GCGTAACATGAGCAAagaaggTATGGGATATGTATTTCA CAGTC-3'
	Forward-2	5'-CTGTGAAATACATATCCCATAccttctTTGCTCATGTTACG CCTGTG-3'
	Reverse-2	5'-GGCCTACCCAAGCACCGGATATGCATATGG-3'

Fr-C_miR-27b MRE	Forward-1	5’-GAGAATCTGAGTAAAATATTAGACTCCAGTTCATTTC TAG-3’
	Reverse-1	5’-CTTTACAACAGTGATTCTGCagttccTAGTTCAAGAGTTC AAGGTGTG-3’
	Forward-2	5’-GAACTCTTGAACTAggaactGCAGAATCACTGTTGTAAAG-3’
	Reverse-2	5’-CAATGAAAACACAGCCGCACGACGATGCCCATCTGC-3’
Oligos for miR-E based shRNA		
ID	Sequence	
Sorl1.5287.t	5’- tcgagaaggtatatTGCTGTTGACAGTGAGCGATAGCAACTCTACAGAAATAAAT AGTGAAGCCACAGATGTATTTATTTCTGTAGAGTTGCTAGTGCCTACTGCCTCG GActtcaaggggctag-3’	
Sorl1.5287.b	5’- aattctagcccctt gaagTCCGAGGCAGTAGGCACTAGCAACTCTACAGAAATAAAT ACATCTGTGGCTTCACTATTTATTTCTGTAGAGTTGCTATCGCTCACTGTCAAC AGCAatataccttc-3’	
Sorl1.2328.t	5’- tcgagaaggtatatTGCTGTTGACAGTGAGCGATGGGTTCTTCATACAGGAGAAT AGTGAAGCCACAGATGTATTCTCCTGTATGAAGAACCCACTGCCTACTGCCTC GGActtcaaggggctag-3’	
Sorl1.2328.b	5’- aattctagcccctt gaagTCCGAGGCAGTAGGCAGTGGGTTCTTCATACAGGAGAAT ACATCTGTGGCTTCACTATTCTCCTGTATGAAGAACCCATCGCTCACTGTCAAC AGCAatataccttc-3’	
Sorl1.1493.t	5’- tcgagaaggtatatTGCTGTTGACAGTGAGCGATCGGTCATCACCTTTGACAAAT AGTGAAGCCACAGATGTATTTGTCAAAGGTGATGACCGACTGCCTACTGCCTC GGActtcaaggggctag-3’	
Sorl1.1493.b	5’- aattctagcccctt gaagTCCGAGGCAGTAGGCAGTCGGTCATCACCTTTGACAAAT ACATCTGTGGCTTCACTATTTGTCAAAGGTGATGACCGATCGCTCACTGTCAAC AGCAatataccttc-3’	
Sorl1.4822.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGACAGCGATGAGTTAACTGTATAT AGTGAAGCCACAGATGTATATACAGTTAACTCATCGCTGCTGCCTACTGCCTC GGActtcaaggggctag-3’	
Sorl1.4822.b	5’-aattctagcccctt gaagTCCGAGGCAGTAGGCAGCAGCGATGAGTTAACTGTATAT ACATCTGTGGCTTCACTATATACAGTTAACTCATCGCTGTCGCTCACTGTCAAC AGCAatataccttc-3’	
Sorl1.3382.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGCCACCTGTGTCAAAGAAGAGAAT AGTGAAGCCACAGATGTATTCTCTTCTTTGACACAGGTGTTGCCTACTGCCTC GGActtcaaggggctag-3’	
Sorl1.3382.b	5’-aattctagcccctt gaagTCCGAGGCAGTAGGCAACACCTGTGTCAAAGAAGAGAAT ACATCTGTGGCTTCACTATTCTCTTCTTTGACACAGGTGGCGCTCACTGTCAAC AGCAatataccttc-3’	
Sorl1.2638.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGAAGGGCAGGAGGTGATCATCAAT AGTGAAGCCACAGATGTATTGATGATCACCTCCTGCCCTGTGCCTACTGCCTC GGActtcaaggggctag-3’	
Sorl1.2638.b	5’-aattctagcccctt gaagTCCGAGGCAGTAGGCACAGGGCAGGAGGTGATCATCAAT ACATCTGTGGCTTCACTATTGATGATCACCTCCTGCCCTTCGCTCACTGTCAAC AGCAatataccttc-3’	
Robo2.2508.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGCAACGAAGTTGTCATTACTGAAT AGTGAAGCCACAGATGTATTCAGTAATGACAACTTCGTTATGCCTACTGCCTCG GActtcaaggggctag-3’	

Robo2.2508.b	5’-aattctagccccttgaagTCCGAGGCAGTAGGCATAACGAAGTTGTCATTACTGAAT ACATCTGTGGCTTCACTATTCAGTAATGACAACTTCGTTGCGCTCACTGTCAAC AGCAatataccttc-3’
Robo2.1298.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGATACAGCATTGTTGAAGTGTAAT AGTGAAGCCACAGATGTATTACACTTCAACAATGCTGTACTGCCTACTGCCTC GGActtcaaggggctag-3’
Robo2.1298.b	5’-aattctagccccttgaagTCCGAGGCAGTAGGCAGTACAGCATTGTTGAAGTGTAAT ACATCTGTGGCTTCACTATTACACTTCAACAATGCTGTATCGCTCACTGTCAAC AGCAatataccttc-3’
Robo2.5035.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGCAGGGAGTTTTCTGTACTGTAAT AGTGAAGCCACAGATGTATTACAGTACAGAAAACTCCCTTTGCCTACTGCCTC GGActtcaaggggctag-3’
Robo2.5035.b	5’-aattctagccccttgaagTCCGAGGCAGTAGGCAAAGGGAGTTTTCTGTACTGTAAT ACATCTGTGGCTTCACTATTACAGTACAGAAAACTCCCTGCGCTCACTGTCAAC AGCAatataccttc-3’
P2ry1.2464.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGCCAGGACGTAGAAGATGTGTTAT AGTGAAGCCACAGATGTATAACACATCTTCTACGTCCTGTTGCCTACTGCCTC GGActtcaaggggctag-3’
P2ry1.2464.b	5’-aattctagccccttgaagTCCGAGGCAGTAGGCAACAGGACGTAGAAGATGTGTTAT ACATCTGTGGCTTCACTATAACACATCTTCTACGTCCTGGCGCTCACTGTCAAC AGCAatataccttc-3’
P2ry1.3780.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGAAACCTTGTATCTGCTATTTAAT AGTGAAGCCACAGATGTATTAAATAGCAGATACAAGGTTGTGCCTACTGCCTC GGActtcaaggggctag-3’
P2ry1.3780.b	5’-aattctagccccttgaagTCCGAGGCAGTAGGCACAACCTTGTATCTGCTATTTAAT ACATCTGTGGCTTCACTATTAAATAGCAGATACAAGGTTTCGCTCACTGTCAAC AGCAatataccttc-3’
P2ry1.2737.t	5’-tcgagaaggtatatTGCTGTTGACAGTGAGCGCCCAGTTTGTATTTGTATCTAAT AGTGAAGCCACAGATGTATTAGATACAAATACAAACTGGTTGCCTACTGCCTC GGActtcaaggggctag-3’
P2ry1.2737.b	5’aattctagccccttgaagTCCGAGGCAGTAGGCAACCAGTTTGTATTTGTATCTAATA CATCTGTGGCTTCACTATTAGATACAAATACAAACTGGGCGCTCACTGTCAACA GCAatataccttc-3’

References

1. Pinzón N, et al. (2017) MicroRNA target prediction programs predict many false positives. *Genome Research* 27(2):234–245.

2. Fridrich A, Hazan Y, Moran Y (2019) Too Many False Targets for MicroRNAs: Challenges and Pitfalls in Prediction of miRNA Targets and Their Gene Ontology in Model and Non-model Organisms. *Bioessays* 41(4):e1800169.

3. Olson P, et al. (2009) MicroRNA dynamics in the stages of tumorigenesis correlate with hallmark capabilities of cancer. *Genes Dev* 23(18):2152–2165.

4. Michael IP, et al. (2019) ALK7 Signaling Manifests a Homeostatic Tissue Barrier That Is Abrogated during Tumorigenesis and Metastasis. *Developmental Cell* 49(3):409–424.e6.

5. Hendrikx S, et al. (2019) Endothelial Calcineurin Signaling Restrains Metastatic Outgrowth by Regulating Bmp2. *Cell Rep* 26(5):1227–1241.e6.

6. Micheletti R, et al. (2017) The long noncoding RNA Wisper controls cardiac fibrosis and remodeling. *Sci Transl Med* 9(395). doi:10.1126/scitranslmed.aai9118.

7. Li L, Hanahan D (2013) Hijacking the neuronal NMDAR signaling circuit to promote tumor growth and invasion. *Cell* 153(1):86–100.

8. Wang X, Spandidos A, Wang H, Seed B (2012) PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res* 40(D1):D1144–D1149.