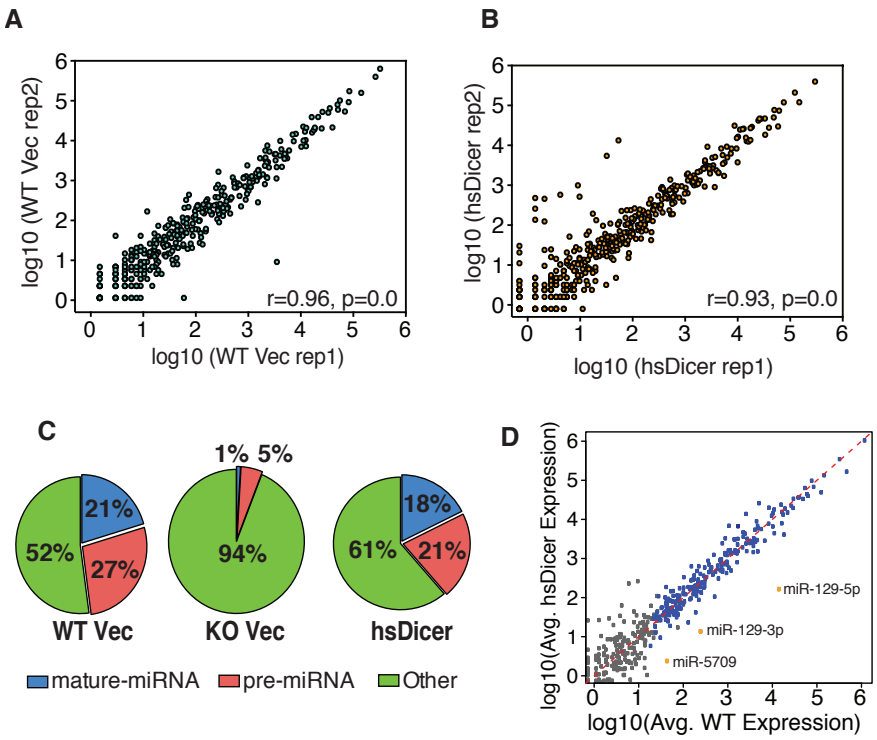
Supplemental Information

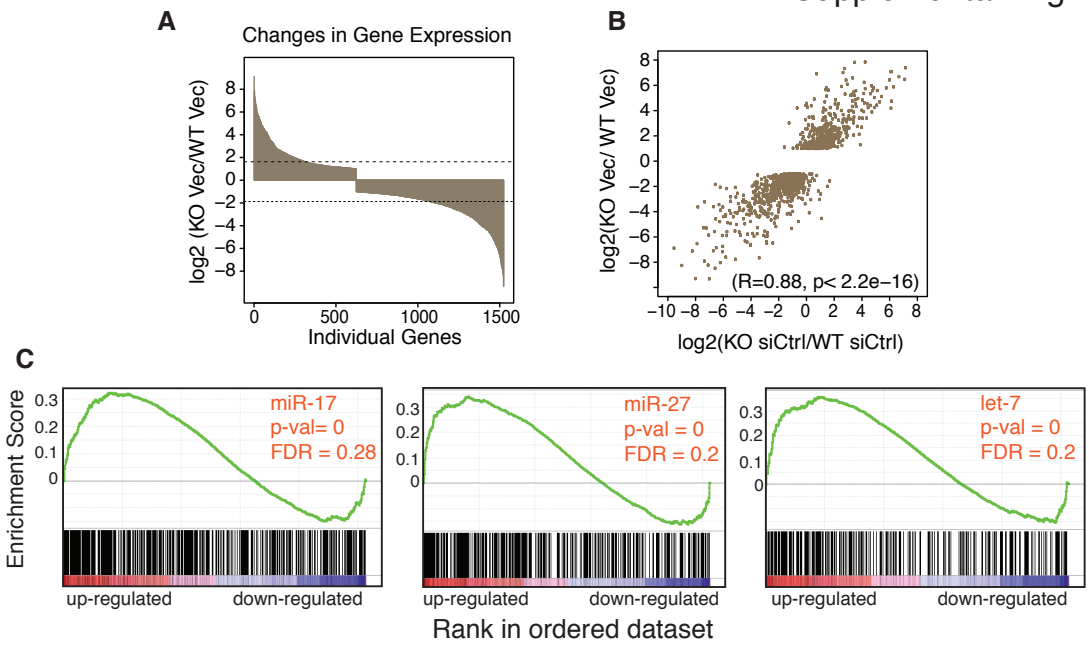
Dicer loss and recovery induces an oncogenic switch driven by transcriptional activation of the oncofetal Imp1-3 family

Courtney K. JnBaptiste, Allan M. Gurtan, Kevin K. Thai, Victoria Lu, Arjun Bhutkar, Mei-Ju Su, Asaf Rotem, Tyler Jacks, Phillip A. Sharp



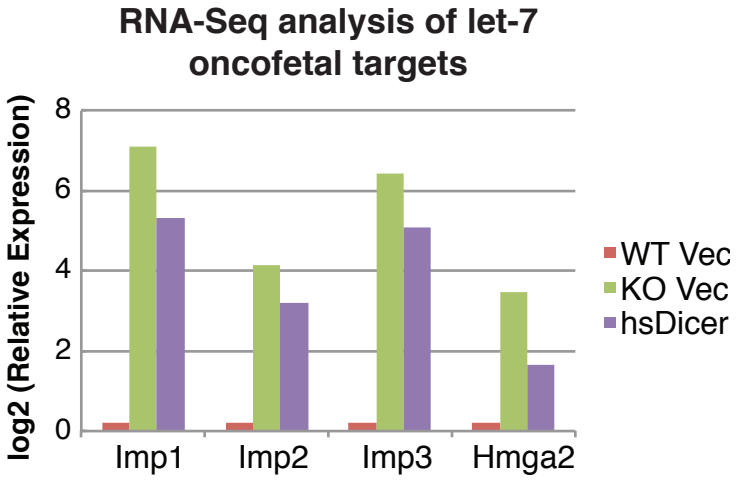
Supplemental Fig. 1. Restoration of miRNA expression through hsDicer. (A, B) Comparison of normalized miRNA counts between the 2 independently isolated clones each for WT vec and hsDicer. **(C)** Small RNA library composition for each cell type. Numbers indicated are averaged across the 2 libraries for each genotype. **(D)** Same as in Fig. 1C. Gray indicates non-expressed miRNAs (<5 RPM in WT Vec), yellow indicates the 3 differentially expressed miRNAs (Adj. p-value < 0.05), and blue indicates all other miRNAs. Red line: y=x.

Supplemental. Fig 2

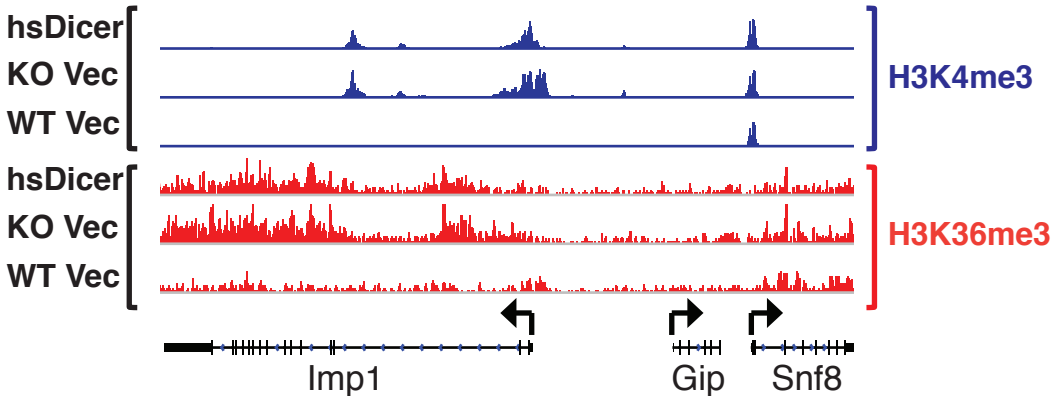


Supplemental Fig. 2. Effects of miRNA loss on gene expression. (A) Fold change waterfall plot for genes differentially expressed between WT Vec and KO Vec cells (FDR < 0.05; fold change ≥ 2 in either direction). Dotted horizontal lines represent median up and downregulation. **(B)** Scatter plot comparing fold changes reported in this study (y-axis) for differentially expressed genes from (A) to a previous report (Gurtan et al., 2013) (x-axis) with wild-type and knockout cells transfected with a non-targeting siRNA (siCtrl). Pearson correlation and p-value are indicated. **(C)** GSEA enrichment plots indicate that the most upregulated genes in KO Vec are enriched for targets of miR-17, miR-27 and let-7. Genes are ordered from left to right with decreasing fold change.

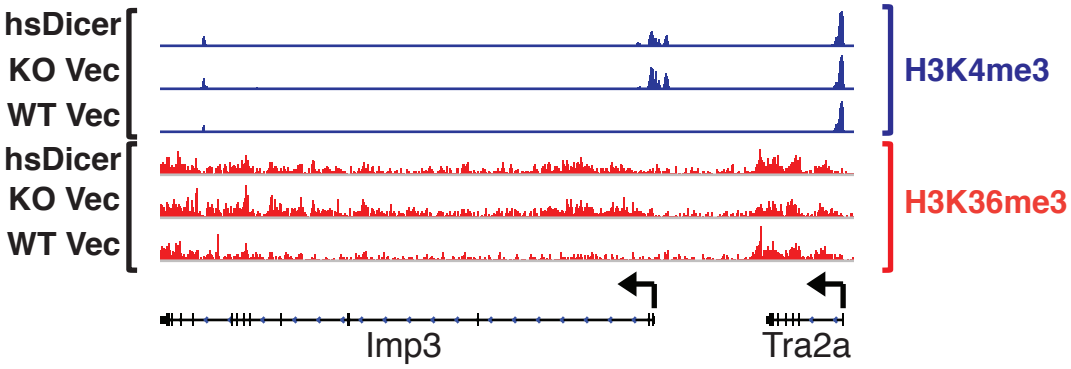
A


RNA-Seq analysis of let-7 oncofetal targets
log2 (Relative Expression)
8
6
4
2
0
Imp1
Imp2
Imp3
Hmga2
WT Vec
KO Vec
hsDicer

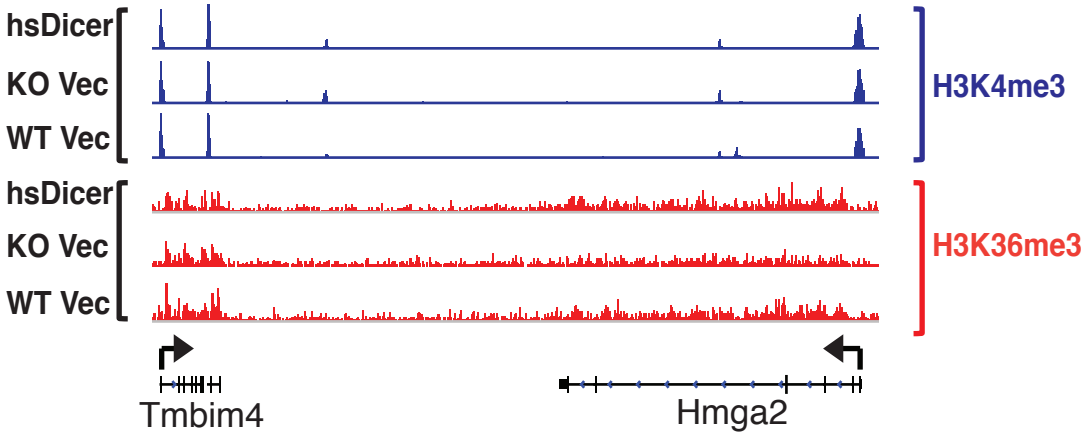
B


hsDicer
KO Vec
WT Vec
H3K4me3
hsDicer
KO Vec
WT Vec
H3K36me3
Imp1
Gip
Snf8

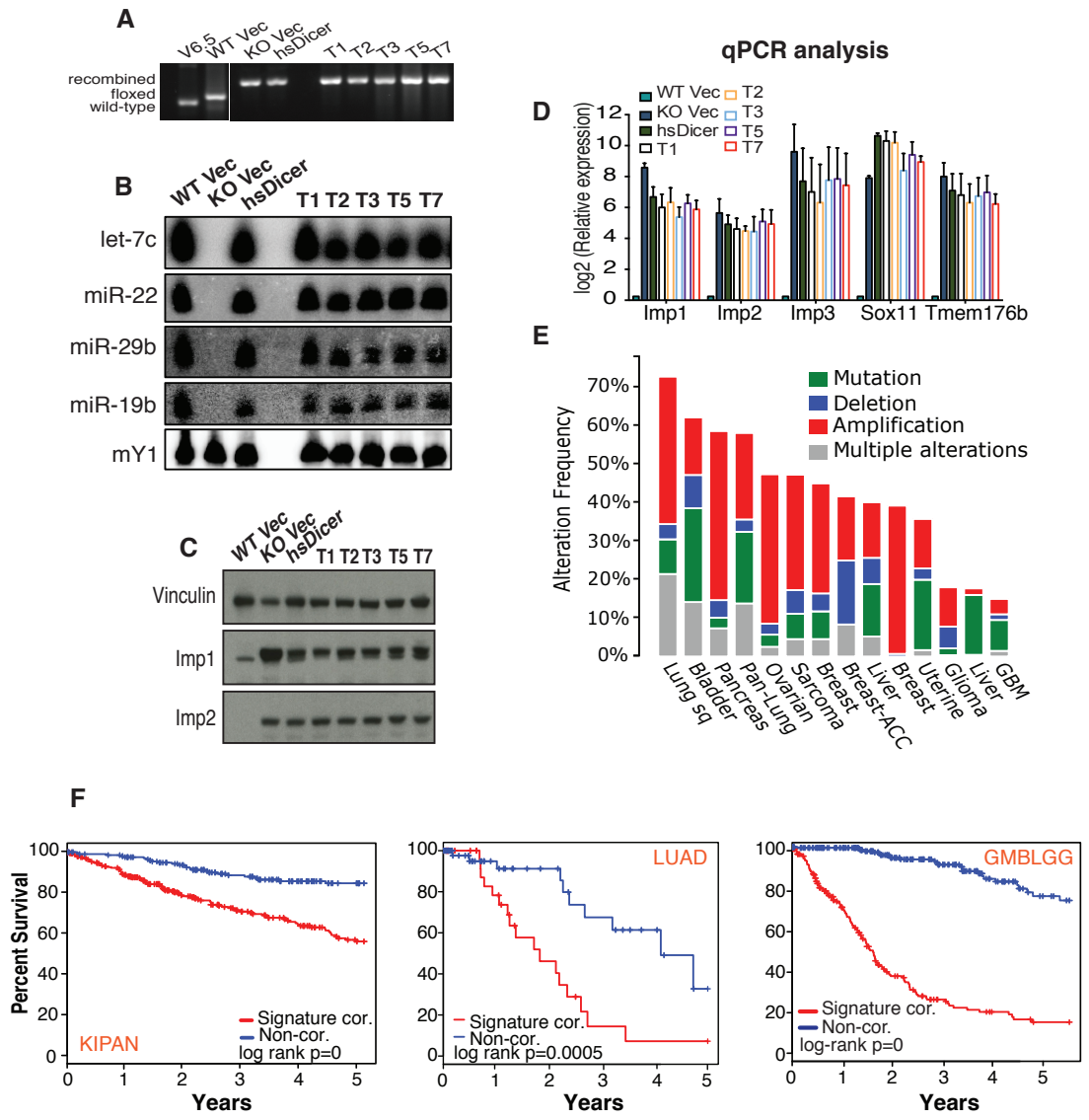
C


hsDicer
KO Vec
WT Vec
H3K4me3
hsDicer
KO Vec
WT Vec
H3K36me3
Imp3
Tra2a

D

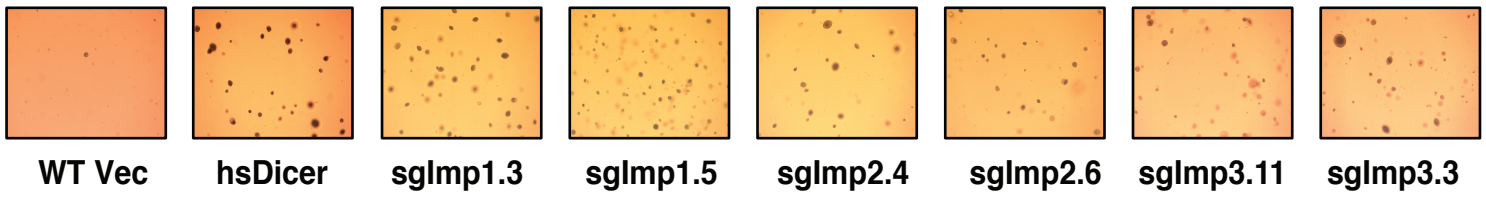

hsDicer
KO Vec
WT Vec
H3K4me3
hsDicer
KO Vec
WT Vec
H3K36me3
Tmbim4
Hmga2

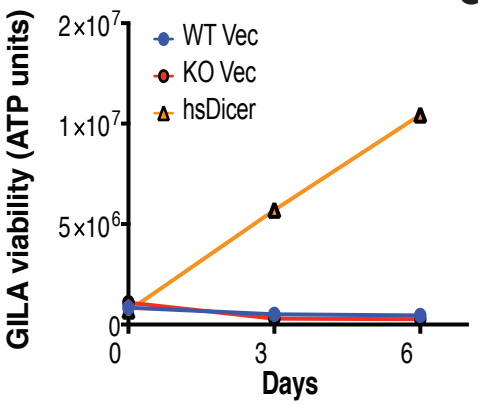
Supplemental Fig. 3. Transcriptional profiling of let-7 oncofetal targets. (A) Log2 RNA-Seq expression of four let-7 regulated oncofetal genes. For each gene, expression is relative to WT Vec levels. **(B-D)** Normalized read counts for H3K4me3 and H3K36me3 marks at the Imp1, Imp3 and Hmga2 loci. Within each chromatin mark, WT Vec, KO Vec and hsDicer are all set to the same scale. Flanking genes are shown as controls. Arrowheads indicate transcription start sites (TSS).



Supplemental Fig. 4. miRNA-resistant genes remain expressed in subcutaneous tumors – related to Figure 5. (A) PCR genotyping of hsDicer subcutaneous tumor derived cell lines shows recombination of the floxed alleles. V6.5: Embryonic stem cells not containing floxed alleles used as a control. **(B)** Northern blot analysis of tumor derived cell lines indicates robust expression of mature miRNAs. mY1 is used as a loading control. **(C)** western blot and **(D)** qPCR analyses also confirm high expression for a panel of irreversible genes. Bars represent mean and standard deviation of 3

independent experiments. In all cases, WT Vec, KO Vec and hsDicer are the parental cell lines, while T1-T7 are cell lines derived from hsDicer tumors. **(E)** Pan-cancer alteration profile of the high-confidence miRNA-resistant gene signature using cBioPortal. The cancer types are indicated. Lung sq: Lung Squamous Cell Carcinoma; Pan-Lung: pan non small cell lung cancer; Breast-ACC: Adenoid Cystic Carcinoma of the Breast; GBM: Glioblastoma Multiforme. **(F)** Kaplan-Meier survival analysis for patients with higher correlation (red) of the high-confidence miRNA-resistant signature in KIPAN (pan kidney, $|z| > 0.5$ extremes of score distribution), LUAD (lung adenocarcinoma, $|z| > 1.25$) and GBMLGG (pan glioma, , $|z| > 0.5$) cohorts compared to patients with lower signature correlation scores (blue).





C

Imp1

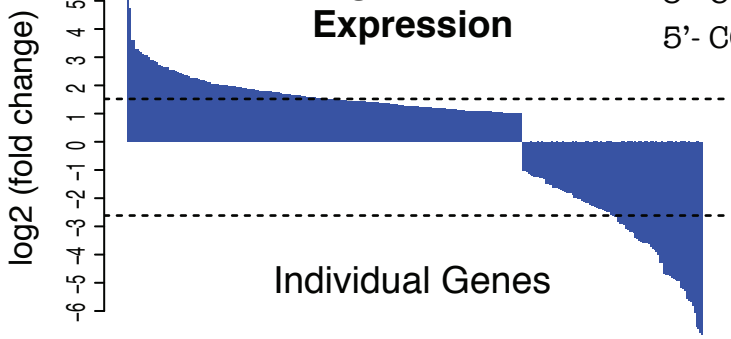
5'- TGCCCCGA CGAGCACTGGGCGATGAAGGC **WT**
5'- TGCCCC - - CGAGCACTGGGCGATGAAGGC **(both)**
5'- TGCCCCGAAGCGAGCACTGGGCGATGAAGGC **(clone 5)**
5'- TGCCCCGAC CGAGCACTGGGCGATGAAGGC **(clone 6)**

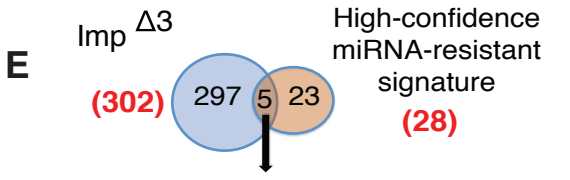
Imp2

5'- CGCCTTCGTGGACTACCCCGACCAGAACT **WT**
5'- CGCCTT - - TGGACTACCCCGACCAGAACT **(both)**
5'- CGCCTTCGT - - - - TACCCCGACCAGAACT **(both)**

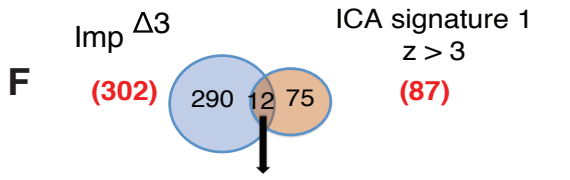
Imp3

5'- CCACGCCGGCCCCGCGG ACCTGGAAAGT **WT**
5'- CC - - - - - - - - - - - - CGG ACCTGGAAAGT **(clone 5)**
5'- CCACGCCGGCCCCGCGGGACCTGGAAAGT **(clone 5)**
5'- CCACGCCG - - - - - - CGG ACCTGGAAAGT **(clone 6)**

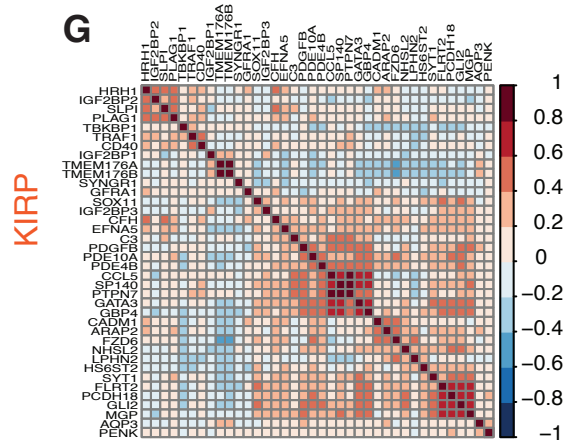


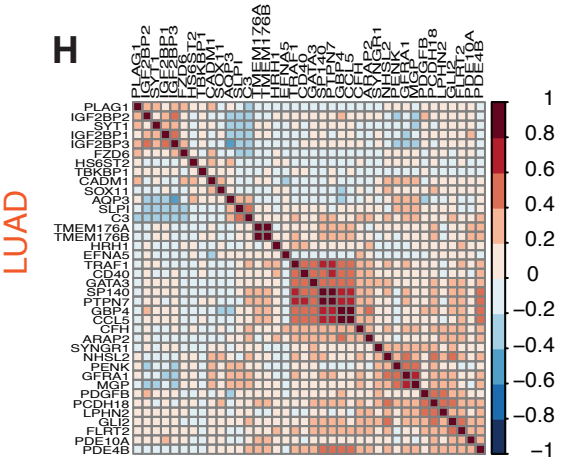


Msln, Traf1, Pde4b, Tmem176b, Cd40



Msln, Traf1, Pde4b, Tmem176b, Cd40
Cfh, Rbp2, Efna5, C3, Mgb, Fzd6, Ccl5





Supplemental Fig. 5. Simultaneous expression of Imp-3 drives the oncogenic switch – related to Figure 6. (A) Representative brightfield soft agar images at 4x magnification for each clone of Imp1, Imp2 and Imp3 single knockout cells. **(B)** GILA assay on the WT Vec, KO Vec and hsDicer cell lines. Only hsDicer cells are able to grow over a 6-day time course in GILA, consistent with their ability to grow and form colonies in soft agar. Data are plotted as the mean ± standard deviation (n=3). **(C)** Genomic PCR product sequencing at the Imp1-3 loci in $Imp^{\Delta 3}$ cells. WT designates the wild type sequence. The 20-nucleotide target sequence is indicated in green; and the PAM is shown in blue. Red dashes denote deletions, red nucleotides denote insertions, and parentheses indicate whether that particular allele is recovered from both or either $Imp^{\Delta 3}$ clone. **(D)** Waterfall plot of log2-fold changes ($Imp^{\Delta 3}$/ hsDicer) for differentially expressed genes (FDR $\leq$ 0.05 and fold change $\geq$ 2 in either direction). Dashed horizontal lines represent median up and downregulation. **(E)** Five genes belonging to the high-confidence miRNA-resistant signature (discovered in Fig. 4G) are significantly differentially expressed between $Imp^{\Delta 3}$ and hsDicer cells. **(F)** Overlap of $Imp^{\Delta 3}$ and hsDicer differentially expressed genes with ICA signature 1 associated genes. **(G-H)** Heatmap depicting gene-by-gene spearman correlation coefficients (of leading-edge genes) in tumor samples of **(G)** cervical kidney renal papillary cell carcinoma (KIRP) and **(H)** lung adenocarcinoma (LUAD) patients from TCGA.

Supplementary Material and Methods

Expression plasmid construction and cell line generation

Wild-type human Dicer (hsDicer) was PCR-amplified from pCAGGs-Flag-hsDicer (Addgene plasmid # 41584) using Forward: 5'-AAAAACATGTTATACCCATACGACGTACCAGATTACGCTGACTACAAAGACG ATGACGAC-3' and reverse: 5'-AAAAAAAAACATGTTCAGCTATTGGGAACC TGAGGTTG -3' primers. An HA-Flag tandem epitope was encoded within the 5' forward primer used for amplification. A single N-terminal leucine residue precedes the HA-Flag tag. Similarly, hsDicer is preceded by a single leucine residue after the flag epitope. HA-Flag-hsDicer was subcloned into the NcoI sites of the retroviral vector pMMP-Puro. Phoenix packaging cells were transfected with either an empty pMMP-Puro retroviral vector plasmid or the equivalent vector with the hsDicer insert using Lipofectamine 2000 (Invitrogen). Viral supernatants were cleared with a 0.45 um syringe filter and concentrated using Ultracel 100K centrifugal filter units (Millipore). MSCs at subconfluency were infected with viral supernantant supplemented with 8μg of polybrene (Sigma). Stably transduced lines were selected and maintained in alpha-MEM containing 2.5μg/ml puromycin (Life Technologies). After a heterogenous population of puromycin resistant cells grew out, cells were seeded at low density and two monoclonal lines were isolated per genotype.

Dual color miRNA reporter assay

The bidirectional pTRE-Tight-BI (Clontech) eYFP and mCherry reporter construct has been previously described (Mukherji et al. 2011). Oligonucleotides encoding target sites for miR-24, miR-29b, miR-31 and let-7c were synthesized by IDT and inserted into the 3' UTR of mCherry using HindIII and SalI cloning sites. The sequences are provided in Supplemental Table S7. Each miRNA target is imperfectly complementary across nucleotides 9-11 and consists of 2 tandem sites separated by a 6-nucleotide spacer. MSCs were seeded at 10,000 cells per well of a 96-well plate, and the next day were transfected with equivalent amounts (100ng each) of the reporter plasmid and rtTA using Lipofectamine 2000. At the time of transfection, the media was changed to complete Alpha-MEM supplemented with 2mg/ml doxycycline (Sigma). FACS measurements were taken on a LSR II HTS instrument (BD Biosciences) 24 hrs post-transfection and population averages of mCherry and eYFP intensities were used for determining the fold repression.

Doxorubicin and apoptosis assays

For basal apoptosis analysis, MSCs were seeded at 80,000 cells per well of a 12 well plate. Cells were harvested for analysis 4 days later. The cell culture media was harvested and after washing, was pooled with the PBS supernatant to collect floating cells. Wells were trypsinized, quenched with cell culture media and then cells were harvested by centrifugation and added to the pool of floating cells. The combined cell suspension was washed in cold PBS and harvested via centrifugation. Apoptosis was assayed using the Active Caspase3-FITC kit (BD Biosciences) following the manufacturer's protocol. The percentage of FITC stained cells was measured by an LSR

II machine (BD Biosciences).

For Doxorubicin induced stress analysis, MSCs were seeded at 10000 cells per well of a 96 well plate. The next day, cells were treated with Alpha-MEM supplemented with Doxorubicin at the indicated concentrations. Cells were harvested approximately 24 hours post drug treatment. Analysis of cell death was carried out as described above for the basal apoptosis assay.

PCR genotyping

Cells were genotyped for Dicer status as previously described (Harfe BD 2005). Briefly, genomic DNA was isolated using QuickExtract (catalogue #: QE0905T, Epicenter) following the manufacturer's protocol. Genomic DNA was then PCR amplified using the following primer mix in a 2:1:1 ratio respectively- MC93, MC94, MC113 with a mouse genotyping kit from KAPA Biosystems (catalogue #: KK5621). See Supplemental Table S7 for primer sequences. PCR products were analyzed on a 2% agarose gel. The endogenous wild-type allele yields a product of ~350bp, the floxed allele yields a product of ~420bp, and the deletion allele yields a product of ~470bp.

Doubling time

For calculating cell-doubling times, we used the xCELLigence system (Acea Biosciences) for the real-time monitoring of cell proliferation and viability. Cell index (CI- an arbitrary unit reflecting the electronic cell-sensor impedance) measurements were performed according to the manufacturer's instructions. Briefly, background impedance of 100ul of the cell culture media alone in the E-plate was first determined. Then, sub-confluent cells were trypsinized, counted by hematocytometer and resuspended at 25,000 cells/ml. 100ul of the cell suspension was added to each well. The plate was shaken gently to disperse the cells, then left to stand for 30 minutes in the tissue culture hood at room temperature. The E-plate was then transferred to the xCELLigence instrument and the CI (as a measure of cell proliferation) was monitored at 15-minute intervals over 4 days. Normalized cell index were plotted against incubation time and doubling times were determined based on a linear part of the cell index/ proliferation curve using the RTCA-integrated software of the xCELLigence system.

Subcutaneous injections & Histology

Subconfluent Dicer wild-type, Dicer knockout and hsDicer rescued cells were harvested with trypsin and washed in cold PBS. After centrifugation, cells were resuspended at $1x10^5$ cells/ 100ul PBS. 100ul of the cell suspension was subcutaneously injected into the flanks of six week old Nu/J nude female mice.

Eight weeks post injection, mice were sacrificed via CO_2 asphyxiation. Excised tumors were fixed in 4% paraformaldehyde, transferred to 70% ethanol, and then embedded in paraffin. Tumors were then sectioned and stained with hematoxylin and eosin, and analysis for tumor type was performed by a pathologist (R.T.B.). Some tumors were trypsinized and replated in Alpha-MEM to generate secondary cell lines.

For the tumorigenicity comparison of hsDicer and $Imp^{\Delta 3}$ cells, $1x10^5$ cells were injected into the left flank or the right flank respectively of the same nude mouse. A total of 10 mice were used in this cohort. Mice were monitored weekly for appearance of tumors. From initial appearance, tumors were measured manually using calipers by the

same investigator throughout the entire time course. Tumor volume was calculated as follows: volume = ½(length x width^2).

CRISPR/Cas9 knockouts

Imp1-3 targeting guide RNAs were designed using the MIT Optimized CRISPR design tool (http://crispr.mit.edu). Using protocols from http://crispr.genome-engineering.org, these guides were subsequently cloned into a Cas9/gRNA co-expressing construct (PX458, Addgene plasmid # 48138). hsDicer cells were transfected with the plasmid and ~72 hours later were single cell sorted into a 96 well plate based on GFP signal on an AriaII sorter. Single cell clones were expanded, screened via western blot and genomic PCR for successful knockouts, and two validated clones were subsequently used for experiments. For the Imp2 and Imp3 single knockout cells, the two clones used were derived from different sgRNAs. For the triple knockout cells, plasmids targeting all three Imp proteins were co-transfected into cells. In all knockouts, gRNAs were targeted to the first exons of the respective gene. Oligonucleotides used for gRNA cloning are listed in Supplemental Table S7.

For genomic characterization of the knockouts, we PCR amplified the locus of interest using primers listed in Supplemental Table S7. Blunt-end PCR products were either directly sequenced, or were first topocloned (ThermoFisher), propagated in bacteria, then mini-prepped and Sanger sequenced.

Survival Analyses

The high-confidence miRNA-resistant gene signature derived from RNA-seq analysis was used to score expression profiles of individual TCGA tumors (http://cancergenome.nih.gov/) in the indicated datasets using a single sample enrichment approach (Barbie et al. 2009). Tumors were stratified based on the standardized score and extremes of the score distribution (at various z-thresholds) were compared for differences in survival times using the Kaplan-Meier methodology. Additionally, the Cox proportional hazards regression model was used to analyze the prognostic value of the high-confidence miRNA-resistant signature across all patients in the TCGA pancreatic adenocarcinoma cohort, in the context of additional clinical covariates. Hazard ratio proportionality assumptions for the Cox regression model were validated by testing for all interactions simultaneously ($p = 0.611$). Interactions between the signature score and other significant covariates (age, number of lymph nodes) were tested using a likelihood ratio test (LRT) to contrast a model consisting of both covariates against another model consisting of both covariates plus an interaction term. All survival analyses were performed using the survival package in R.

Imp1-3 downstream gene expression correlations

RSEM RNASeqV2 Level3 normalized mRNA-Seq expression datasets were downloaded from the Broad Institute TCGA Genome Data Analysis Center (2016): Firehose 2016_01_28 run, Broad Institute of MIT and Harvard**.** doi:10.7908/C11G0KM9 via the URL: https://gdac.broadinstitute.org. The datasets were filtered to restrict it to tumor-derived expression values for the GSEA leading-edge genes from Figure 6K. Genes with an upper quartile value across all tumor samples of less than 10 were filtered

out, normalized expression was log transformed, and then gene-by-gene Spearman correlations were calculated in the R programming language. Heatmaps were generated using the corrplot package in R.

miRNA cloning, sequencing and quantitation

From MSCs, total RNA was isolated using Trizol (Life Technologies) and then 20μg was electrophoresed on a 10% denaturing polyacrylamide gel. Small RNAs in the range of 15-65nt were excised (deliberately avoiding the dominant tRNA band). DNA libraries from this size selected RNA population was then generated using the NEBNext Multiplex Small RNA Library Prep Set for Illumina, following the manufacturer's instructions (New England Biolabs).

Multiplexed sequencing reads from an Illumina HiSeq 2000 instrument were bucketed by sample barcode and processed using the FASTX-Toolkit (from the laboratory of Gregory Hannon: http://hannonlab.cshl.edu/fastx_toolkit/index.html) to strip adapters. Clipped reads less than 10bp were dropped form downstream analyses. Within each individual library, the remaining reads were aggregated by sequence identity. Processed reads were first mapped to mouse mature miRNA sequences annotated in miRBase release 19 (http://www.mirbase.org/), using the Bowtie short read alignment tool (Langmead et al. 2009) allowing for unique and repeat alignments with up to a single base pair mismatch per alignment. Reads that did not align to mature micro-RNA sequences were similarly mapped to micro-RNA hairpin sequences where possible, allowing for up to two mismatches per alignment. Individual miRNA mapped reads were summarized by TargetScan microRNA family (Lewis et al. 2005). Differential analysis for miRNA family counts was performed using DESeq (Anders and Huber 2010).

mRNA-Seq analysis

For the first mRNA sequencing dataset (across WT Vec, KO Vec and hsDicer cells), total RNA was isolated with an RNeasy Kit (Qiagen). Samples were DNase treated, and submitted to the BioMicroCenter at M.I.T. for library preparation (Illumina Tru-Seq protocol) and sequencing (Illumina HiSeq 2000; 51nt single end reads). For the second mRNA-Seq dataset (on triple knockout cells as well as the parental hsDicer cells), total RNA was isolated using Trizol and DNase treated with the TURBO DNA-free kit. RNA was then phenol chloroform extracted and analyzed for quality on an Agilent 2100 Bioanalyzer. 1 μg of total RNA with a RIN above 9, was then used as input for library preparation with poly(A) selection, using Illumina's TruSeq Stranded mRNA Library Preparation Kit and protocol. The final libraries were sequenced on a NextSeq 500 machine (75/75 paired end reads).

Illumina HiSeq 2000 or NextSeq 500 reads were mapped to the UCSC mm9 mouse genome build (http://genome.ucsc.edu/) using RSEM (Li and Dewey 2011). For consistency, reads from the second mRNA-Seq run were trimmed down to 51nts using the FASTX-Trimmer tool from the laboratory of Gregory Hannon. Raw estimated expression counts were upper-quartile normalized to a count of 1000 (Bullard et al., 2010) and log2 transformed. Independent Component Analysis (ICA) was used to identify statistically significant (Mann-Whitney-Wilcoxon test) and biologically relevant signatures that characterize the global gene expression profiles of these samples, as

described previously (Li et al. 2015; Dimitrova et al. 2016; Papagiannakopoulos et al. 2016). The R implementation of the core Joint Approximate Diagonalization of Eigenmatrices algorithm (JADE) (Nordhausen et al., 2012; Rutledge and Jouan-Rimbaud Bouveresse, 2013; Biton et al., 2013) was used along with custom R utilities.

For pairwise comparisons of RNA-seq samples, reads were mapped using RSEM as before, and differential expression analyses were performed using using EBSeq v1.12 (Leng et al. 2013). For EBSeq, we used the default settings of the EBTest function allowing the algorithm to filter out genes with 0 reads across all conditions. Genes were called differentially expressed if they changed atleast 2-fold in either direction and also had a PPDE of ≥0.95. Gene set enrichment analysis (GSEA) (Subramanian et al. 2005) was carried out using the pre-ranked mode filtering out gene sets larger than 2000 from the Hallmark or C2 Curated gene sets of the Molecular Signature Database Collection (http://software.broadinstitute.org/gsea/msigdb). For testing enrichment of our custom signatures, we added them to the larger set of C2 Curated or Hallmark gene sets and performed GSEA as before. All analyses were conducted in the R Statistical Programming language (http://www.r-project.org/).

ChIP-Seq library preparation

Chromatin immunoprecipitation (ChIP) was performed as described previously (Gurtan et al. 2013). Briefly, ~ 50 million MSCs were cross-linked for 10 minutes by the addition of 1/10 of the cell culture volume of 11% formaldehyde solution. The cross-linking was quenched by addition of 1/20 the volume of 2.5M glycine. Cross-linked cells were washed 2x with PBS, pelleted and stored at -80°C.

Cross-linked cells were lysed with LB1, and washed with LB2 then resuspended and sonicated in LB3. Samples were sonicated for 14-16 minutes total with a Qsonica Q700 sonicator using a 1/8" microtip and ½" coupler, in 2 minute cycles of 5 seconds on 5 seconds off. The initial 2 minutes were done at 30% amplitude and the remaining time was done in sets of 2 minutes at 65% amplitude, with 1 minute rests in between. Sonicated lysates were cleared and incubated overnight at 4°C with Dynal Protein G magnetic beads that had been previously conjugated with 5µg of the respective antibody (H3K4me3- Millipore 07-473; H3K36me3 -Abcam ab9050-100). Beads were then washed twice in LB3, once in LB3 supplemented with 500 mM NaCl, once in LiCl buffer, and once with TE + 50 mM NaCl. Bound complexes were eluted in elution buffer at 65°C for 15 minutes. Cross-links were reversed overnight at 65°C. RNA and protein were digested using RNAse A and Proteinase K respectively, and DNA was purified with phenol chloroform extraction and ethanol precipitation. Barcoded libraries were prepared and sequenced on Illumina HiSeq 2000 (single end, 45nt reads). Relevant buffer recipes are listed below and a detailed library preparation protocol is available upon request

LB1 (50mM HEPES-KOH pH 7.5, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100)
LB2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA)
LB3 (50mM Tris pH 8, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate)
LiCl Buffer (10mM Tris-HCl pH 8.0, 250mM LiCl, 1mM EDTA, 1% NP-40)
Elution Buffer (50mM Tris-HCl, pH 8.0, 10mM EDTA, 1% SDS)

ChIP-Seq alignments, peak calling and quantification

Reads were de-multiplexed based on library barcodes and mapped to mouse genome mm9 using bowtie (Langmead et al. 2009) requiring unique mapping with at most two mismatches (-n 2 -m 1 –best–strata). Mapped reads were collapsed, and for each histone ChIP the same number of reads was randomly sampled from the corresponding whole cell extract (WCE) input control libraries to match sequencing depth. For example: WT Vec H3K4me3 ChIP and the respective WT Vec WCE were reduced to the same read depth. For each ChIP in a particular cell type, peaks were called relative to the matching genotype's WCE using MACS (Zhang et al. 2008) with default settings and a P-value threshold of 1e-8. Significant H3K4me3 peaks were assigned to genes if they occurred within +/- 2Kb of the transcription start site and significant H3K36me3 peaks were assigned if they overlapped with the gene body (transcription start- transcription end).

For analysis of histone modifications across different cell types, we used MANorm (Shao et al. 2012), a method for quantitative comparison of ChIP-Seq datasets. MANorm normalizes the raw read counts across the two samples in a comparison, and the read density in common peaks are presumed to reflect the scaling relationship of ChIP-Seq signals between two samples. MANorm outputs an M-value, which represents the log2 fold change of normalized reads across the two conditions, as well as a p-value. The log2 fold change in H3K4me3 ChIP signal (for significant peaks called by MACS as described above) was calculated for each pair-wise comparison (KO Vec vs WT Vec and hsDicer vs WT Vec) using the default settings of the MANorm script. Since all peaks input into MANorm were significantly called by MACS (over the respective WCE), we disregarded MANorm's output p-values and instead filtered peaks and their associated genes solely based on M-value >=1.4. In case of any one gene having more than 1 associated H3K4me3 peak, we assigned the peak with the largest M-value to that gene.

Supplemental References:

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**: 108-112.

Biton, A., Zinovyev, A., Barillot, E., and Radvanyi, F. (2013). MineICA: Independent component analysis of transcriptomic data.

Dimitrova N, Gocheva V, Bhutkar A, Resnick R, Jong RM, Miller KM, Bendor J, Jacks T. 2016. Stromal Expression of miR-143/145 Promotes Neoangiogenesis in Lung Cancer Development. *Cancer Discov* **6**: 188-201.

Gurtan AM, Ravi A, Rahl PB, Bosson AD, JnBaptiste CK, Bhutkar A, Whittaker CA, Young RA, Sharp PA. 2013. Let-7 represses Nr6a1 and a mid-gestation developmental program in adult fibroblasts. *Genes Dev* **27**: 941-954.

Harfe BD MM, Mansfield JH, Hornstein E, Tabin CJ. 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proceedings of the National Academy of Sciences* **102**: 10898-10903.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, Haag JD, Gould MN, Stewart RM, Kendziorski C. 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**: 1035-1043.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li CM, Gocheva V, Oudin MJ, Bhutkar A, Wang SY, Date SR, Ng SR, Whittaker CA, Bronson RT, Snyder EL et al. 2015. Foxa2 and Cdx2 cooperate with Nkx2-1 to inhibit lung adenocarcinoma metastasis. *Genes Dev* **29**: 1850-1862.

Mukherji S, Ebert MS, Zheng GXY, Tsang JS, Sharp PA, van Oudenaarden A. 2011. MicroRNAs can generate thresholds in target gene expression. *Nature Genetics* **43**: 854-859.

Nordhausen, K., Cardoso, J.F., Miettinen, J., Oja, H., and Ollila, E. (2012). JADE: JADE and other BSS methods as well as some BSS performance criteria (R package version).

Papagiannakopoulos T, Bauer MR, Davidson SM, Heimann M, Subbaraj L, Bhutkar A, Bartlebaugh J, Vander Heiden MG, Jacks T. 2016. Circadian Rhythm Disruption Promotes Lung Tumorigenesis. *Cell Metab* **24**: 324-331.

Rutledge, D.N., and Jouan-Rimbaud Bouveresse, D. (2013). Independent Components Analysis with the JADE algorithm. TrAC Trends in Analytical Chemistry *50*, 22–32.

Shao Z, Zhang Y, Yuan GC, Orkin SH, Waxman DJ. 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol* **13**: R16.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15545-15550.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.