**Supplemental Information**

**Integrated genomic and functional microRNA analysis identifies decreased miR-30-5p as a** **tumor suppressor and potential therapeutic nanomedicine in head and neck cancer**

Anthony D. Saleh1,2, Hui Cheng1, Scott E. Martin3, Han Si4, Pinar Ormanoglu3, Sophie Carlson1, Paul E. Clavijo1, Xinping Yang1, Rita Das1, Shaleeka Cornelius1, Jamie Couper1, Douglas Chepeha5, Ludmila Danilova6,7, Thomas M. Harris8, Michael B. Prystowsky8, Geoffrey J. Childs8, Richard V. Smith9, A. Gordon Robertson10, Steven J. M. Jones10, Andrew D. Cherniack11, Sang S. Kim12, Antonina Rait12, Kathleen F. Pirollo12, Esther H. Chang12, Zhong Chen1\*, and Carter Van Waes1\*

**Supplementary Methods**

***microRNA Isolation, Library Preparation and Sequencing from HNSCC Samples***

Large and small RNA were purified using a mirVana miRNA isolation Kit (Life Technologies) following a modified manufacturer’s protocol. Twenty-five mg of frozen tissue was homogenized in 1 mL of trizol (Invitrogen) using a TissueLyser II (Qiagen). Following homogenization, extraction was performed using a standard phenol-chloroform method. To the extracted aqueous phase, 10% of volume of homogenate additive was added and then the standard manufacturer’s protocol for fractionating large and small RNA was performed. RNA concentration was determined using a Nanodrop spectrometer (Thermo Scientific), and total RNA integrity was verified on a Bioanalyzer 2100 instrument using an RNA 6000 Nano kit (Agilent Technologies). Sufficient presence of microRNA in small RNA enriched samples was verified by the Bioanalyzer using the small RNA kit (Agilent Technologies).

Small RNA sequencing libraries were constructed using the SOLiD™ Total RNA-Seq Kit (Life Technologies) by manufacturer’s protocol 4452437 Rev B. Briefly, 1 ug of enriched small RNA (<200 bases) was used for ligation into sequencing adaptors. cDNA libraries were reverse transcribed and then size selected by separation on denaturing urea 10% PAGE. Bands were excised that correspond to an insert size of 18-38 nucleotides. The library was then amplified and barcoded by in-gel PCR. Library construct size was verified using the DNA 1000 kit on the Bioanalyzer 2100. cDNA library concentration was determined by RT-PCR by the SOLiD™ library TaqMan® quantification kit. Equal parts of eight cDNA libraries were multiplexed together and 0.6 pmol of the multiplexed pool was used for emulsion PCR using the SOLiD™ EZ BeadTM system. E20 reagents. Emulsification was carried out using manufacturer’s protocol 4441486 Rev. E. Amplification was carried out using manufacturer’s protocol 4443494 Rev. E. Bead enrichment was carried out using manufacturer’s protocol 4443496 Rev. E. Enriched beads for each pool were 3’ labeled using the SOLiD™ pre-deposition plus kit and manufacturer’s protocol 4443496 Rev. E. 4 x 108 beads were deposited per lane of a 6-lane flow chip, and sequencing was performed on the SOLiD™ 5500 system next generation sequencer with protocol 4456991 Rev. J and the SOLiD™ Small RNA SP Kit (Life Technologies).

***miRNA Mapping, Expression Profiling Quantification, and Differential Abundance Analysis***

The Libraries were sequenced using SOLID5 (Life Technologies) and the color-space reads were mapped to human reference genome hg19 using the small RNA analysis module in LifeScope2 software (Life Technologies). The mapping program mapReads maps the small RNA reads to three different references in three steps: In the first step, the reads are mapped to filter sequences to eliminate the reads from irrelevant sources (such as tRNA, adaptor sequence, etc.); in the second step, the remaining reads from the first step are mapped to miRBase v16; in the third step, the unmapped reads from the second step are mapped to the genome reference sequence hg19 in order to find novel small RNAs. The mapped reads outputs from the second and third steps are merged as the primary output. The mapReads program also removes the F5-P2 adaptor sequence from the reads and to report only small RNA sequences in the BAM files. The downstream analysis steps were mainly performed using miRDeep2 package (1). Briefly, the mapping results in sam format were converted to the arf format used in miRDeep2 and in turn the miRDeep2.pl script was used to identify all known and novel miRs using default settings. Finally, all the identified miRs were quantified based on the reads numbers assigned to them and normalized using the total counts per million in that sample. We used SAMseq (samr v2.0, R 3.0.2) for two-class unpaired analyses with a read count input matrix and an FDR threshold of 0.05 to identify miRs that were differentially expressed. Each run generated a pair of files: genes “up” and “down”, then ranked the filtered results by a median-based fold change (2).

***Integrative Analysis to Identify miRNA-mRNA Pairs in HNSCC TCGA Data***

miR and mRNA abundance for 279 HNSCC tumor specimens were extracted from Level 3 data associated with the data freeze (3). Briefly, the miRNA sequence data are separated into individual samples based on the index read sequences, and the reads undergo an initial QC assessment. Adapter sequence is then trimmed off, and the trimmed reads for each sample are aligned to the NCBI GRCh37-lite reference genome.

Routine QC includes assessing a subset of raw sequences from each pooled lane for the abundance of reads from each indexed sample in the pool, the proportion of reads that possibly originate from adapter dimers (i.e. a 5’ adapter joined to a 3’ adapter with no intervening biological sequence), and for the proportion of reads that map to human miRNAs. Sequencing error is estimated by a method originally developed for SAGE.

Libraries that pass this QC stage are preprocessed for alignment. While the size-selected miRNAs vary somewhat in length, typically they are ~21 bp long, and so are shorter than the 31- bp read length. Given this, each read sequence extends some distance into the 3' sequencing adapter. Because this non-biological sequence can interfere with aligning the read to the reference genome, 3’ adapter sequence is identified and removed (trimmed) from a read. The adapter-trimming algorithm identifies as long an adapter sequence as possible, allowing a number of mismatches that depends on the adapter length found. A typical sequencing run yields several million reads; using only the first (5’) 15 bases of the 3’ adapter in trimming makes processing efficient, while minimizing the chance that a miRNA read will match the adapter sequence.

The algorithm first determines whether a read sequence should be discarded as an adapter dimer by checking whether the 3’ adapter sequence occurs at the start of the read. For reads passing this stage, the algorithm then tries to identify an exact 15-bp match anywhere within the read sequence. If it cannot, it then retries, starting from the 3' end, and allowing up to 2 mismatches. If the full 15bp is not found, decreasing lengths of adapter are checked, down to the first 8 bases, allowing one mismatch. If a match is still not found, from 7 bases down to 1 base is checked, with an exact match required. Finally, the algorithm will trim 1 base off the 3’ end of a read if it happens to match the first base of the adapter. This is based on two considerations. First, it is preferable to get a perfect alignment than an alignment that has a potential one-base mismatch. Second, if only 1 base of adapter was found in the read sequence, the read is likely too long to be from a miRNA and the effect of the trimming on its alignment would not affect this sample’s overall miRNA profiling result.

After each read has been processed, a summary report is generated containing the number of reads at each read length. Because the shortest mature miRNA in miRBase v16 is 15 bp, any trimmed read that is shorter than 15bp is discarded; remaining reads are submitted for alignment to the reference genome. BWA9 alignment(s) for each read are checked with a series of three filters. A read with more than 3 alignments is discarded as too ambiguous. For TCGA quantification reports, only perfect alignments with no mismatches are used. Based on comparing expression profiles of test libraries (data not shown), reads that fail the Illumina basecalling chastity filter are retained, while reads that have soft-clipped CIGAR strings are discarded.

For reads retained after filtering, each coordinate for each read alignment is annotated using the miRBase v16, requiring a minimum 3-bp overlap between the alignment and an annotation. In annotating reads we address two potential issues. First, a single read alignment can overlap feature annotations of different types; second, a read can have up to three alignment locations, and each alignment location can overlap a different type of feature annotation. By considering heuristically determined priorities ((1) Mature strand, (2) Star strand, (3) Precursor miRNA, (4) Stemloop, from 1 to 6 bases outside the mature strand, between the mature and star strands, (5) “unannotated”, any region other than the mature strand in miRNAs where no star strand is annotated.), we resolve the first issue by giving each alignment a single annotation. We resolve the second by collapsing multiple annotations to a single annotation, as follows.

If a read has more than one alignment location, and the annotations for these are different, we use the priorities from Table V.1 to assign a single annotation to the read, as long as only one alignment is to a miRNA. When there are multiple alignments to different miRNAs, the read is flagged as cross-mapped22, and all of its miRNA annotations are preserved, while all of its non-miRNA annotations are discarded. This ensures that all annotation information about ambiguously mapped miRNAs is retained and allows annotation ambiguity to be addressed in downstream analyses. Note that we consider miRNAs to be cross-mapped only if they map to different miRNAs, not to functionally identical miRNAs that are expressed from different locations in the genome. Such cases are indicated by miRNA miRBase names, which can have up to 4 separate sections separated by "-", e.g. hsa-mir-26a-1. A difference in the final (e.g. ‘-1’) section denotes functionally equivalent miRNAs expressed from different regions of the genome, and we consider only the first 3 sections (e.g. ‘hsa-mir-26a’) when comparing names. As long as a read maps to multiple miRNAs for which the first 3 sections of the name are identical (e.g. hsamir-26a-1 and hsa-mir-26a-2), it is treated as if it maps to only one miRNA, and is not flagged as cross-mapped.

From the profiling results for a tumor type, for a minimum of approximately 100 samples, we identify the depth of sequencing required to detect the miRNAs that are expressed in a sample by considering a graph of the number of miRNAs detected in a sample as a function of the number of reads aligned to miRNAs. For the current work, a library from a sequenced pool was required to have at least 750,000 reads mapped to miRBase annotations. For any sequencing run that fails to meet this threshold, we sequence the sample again to achieve at least the minimum number of miRNA-aligned reads.

Finally, for each sample, the reads that correspond to particular miRNAs are summed and normalized to a million miRNA-aligned reads to generate the quantification files that are submitted to the DCC. Quantification files include information on variable 5’ and 3’ read alignment locations, which can reflect isoforms, adapter trimming and RNA degradation.

Mature 5p and 3p strands were ranked by RPM variance across the samples, and the most variable 50% with a minimum expression of at least 50 RPM were used for integrated analysis. Gene expression was calculated from RNA-Seq data with RSEM v1.1.132 and zeros replaced with the minimum non-zero RSEM values (0.0033) (4). The most-variant 50% of genes were used for integrated analysis. Both miRNA and mRNA expression data were log2 transformed.

A multi-step approach was applied to identify miR-mRNA target relationships. We used linear regression to identify pair-wise negative correlations of miR and mRNA expression, in conjunction with available prediction tools from miR target databases. We generated a high confidence dataset of global miR-mRNA interactions. The details were documented in TCGA marker paper supplementary information S7.43 (3).

***Copy Number Variation (CNV) Data Analysis***

Copy number data for 279 HNSCC tumor specimens were extracted from Level 3 data associated with the data freeze. The CNV number associated with each gene was defined as the segmented

GISTIC2 value at the corresponding genomic location (5). The Integrative Genomics Viewer (IGV) was used to visualize copy number data (6). Linear regression was applied to assess the correlation between miRNA expression and CNV.

***TCGA DNA methylation Data Analysis***

For DNA methylation data analysis, we used Level 3 DNA methylation data for 279 HNSCC tumor specimens (3). The data were represented as beta values (β) from Illumina HumanMethylation 450k array. We found CpG probes in promoter regions of miRNAs from miR-30 family members using coordinates of transcription start sites (TSS) (7). The promoter region was specified as ± 1500bp from TSS. For every CpG probe, we estimated the difference of miRNA abundance between unmethylated (β < 0.1) and methylated (β > 0.3) samples using a t-test. BH corrected P-values (FDR) from the t-tests were used to find CpG probes that were significantly differential between unmmethylated and methylated groups using a 0.05 threshold (Supplementary Table S7). Then, we averaged methylation beta values across significant probes for each miR and correlated them with the corresponding miR expression using Spearman’s correlation test.

***Survival Analysis***

The survival analyses were performed with R package (8). The overall survival curves were obtained using Kaplan-Meier method and were compared using the log-rank test. The Cox proportional hazards model was used to estimate Hazard Ratios (HRs) with 95% Confidence Intervals (CIs). Subjects were dichotomized as low miR expression (< median) and high miR expression (≥ median), using the median expression of each miRNA as a cutoff. To compare overall survival time by CNV, subjects were categorized as having MIR30E/A deletion if their GISTIC copy number value was less than −0.1, otherwise they were considered to have no deletion.

**Associations of miR-30 Genetic Alterations and Expression with Stage, Site, Smoking and HPV**

***Status of HNSCC from TCGA Datasets***

Fisher’s exact tests were used to assess associations between miR-30a-5p expression/methylation and clinical characteristics, or between miR-30e-5p expression/copy number loss and clinical characteristics from 279 HNSCC tumor samples of TCGA datasets. Statistical analyses were performed using R version 3.2.2. Significance was defined as p < 0.05. Tumor site was classified as oral cavity if the tumor samples came from any of the following anatomic subdivisions: buccal mucosa, floor of mouth, hard palate, lip, oral cavity, oral tongue, and alveolar ridge; tumor site was classified as oropharynx if the tumor samples came from tonsil, base of tongue or oropharynx.

***Inverse Correlation of miR-30a Expression with Putative Target Genes***

Linear regression analysis was performed as describe previously (3). Briefly, inverse relationships between expressions of miR-30a-5p and its putative target genes were assessed using HNSCC TCGA datasets. P-values from linear regression measure the statistical significance of inverse relationship.

***HNSCC Cell Lines***

A panel of 10 HNSCC cell lines was obtained from the University of Michigan squamous cell carcinoma (UM-SCC) series from Dr. T.E. Carey. The origin of these UM-SCC cell lines was authenticated in 2010 by genotyping with 9 markers as described previously (9), and frozen stocks verified to be free of mycoplasma by RT-PCR were established. Preserved frozen stocks of lines were used within 3 months of culture. These HNSCC cell lines are also well characterized for TP53 status (10). UM-SCC cell lines were cultured in minimal essential medium supplemented with 10% fetal calf serum, penicillin and streptomycin (100 µg/mL), MEM Non-Essential Amino Acids. Human primary oral keratinocytes (HOK) from oral gingival mucosa were purchased from Lonza, and were used as a control cell line. The cells were cultured in serum free Oral Keratinocyte Medium with supplements (Science Cell) for less than five passages.

**In Vitro *microRNA Mimic Viability Screen***

Cells were maintained in MEM containing 10% heat inactivated fetal bovine serum (FBS) supplemented with non-essential amino acids and sodium pyruvate. Transfections were performed in 384-well plates (Corning 3570). Cell viability was measured using CellTiter Glo

(Promega). For transfections, 20 L of serum free media containing Lipofectamine RNAiMax (0.1L) was added to wells containing miR mimic (0.8 pmol). Lipid and miR mimic were allowed to complex for 45 min at ambient temperature before addition of 1500 cells in MEM, 20% FBS to yield final transfection mixtures containing 20 nM miR mimic in MEM, 10% FBS.

The screening campaign was conducted a miR mimic library (Qiagen) based on Sanger miRBase 13.0 and consisting of ~800 mimics. Viability (CellTiter Glo, Promega) was assayed 72 h post-transfection on a PerkinElmer Envision 2104 Multilabel plate reader. Ambion Silencer Select Negative Control #2 was incorporated on all screening plates for normalization (16 wells per plate; the median negative control value on each plate was used to normalize sample wells).

Qiagen’s AllStars Cell Death control was incorporated as a positive transfection control (16 wells per plate). All screen plates exhibited assay z’-factors greater than 0.6. Negative control normalized viability data was converted into robust z-scores using the median absolute deviation

(MAD) (11).

***RT-PCR Validation of mRNA Targets***

2x105 UM-SCC-46 cells were plated in each well of a 6-well plate. 15 nM of mirVana microRNA mimic or inhibitor (Life Technologies) was reverse transfected using 3.75 uL of Lipofectamine RNAiMAX by standard manufacturer’s protocol (Life Technologies) for 48-72 h. Then cells were washed with PBS, and collected into 0.5 mL Trizol reagent. Total RNA was purified using mirVana miR isolation Kit (Ambion). Two µg of total RNA was reverse transcribed using a high capacity cDNA reverse transcription kit (Applied Biosystems) following manufacturer’s instructions. mRNA expression levels were assessed by real time-PCR using TaqMan® gene expression assays (Applied Biosystems), with 40 ng of cDNA was used in each reaction. Reactions were run on an ABI 7900HT real-time PCR machine. Expression levels were normalized to 18S as an endogenous loading control. We verified that transfection resulted in an ~250 fold increase in total miR30a expression by RT-PCR, consistent with an estimated increase of ~12.5% miR30a available in the non-endosomal fraction (data not shown).

***RT-PCR for miRNA***

For miRNA qRT-PCR reverse transcription and real time primer probe sets (Applied Biosystems) were used with 10 ng of total RNA. Expression levels were normalized to RNU44 as an endogenous loading control.

***Western Blotting***

UM-SCC-46 cells were transfected as described above and then lysed into 100 uL of SDS lysis buffer [1% SDS, 50 mM Tris pH 8.0, 10 mM EDTA, Protease inhibiter (Roche), and Halt Phosphatase Inhibitor (Thermo Scientific)]. Samples were sonicated using a probe sonicator four times for 5 sec each on ice. Lysates were cleared by centrifugation at 14,000 xg for 10 min at 4 C.

Protein concentration was determined using a BCA Protein Assay (Thermo Scientific). 25 g of total protein was subjected to SDS-PAGE on a 4-12% gradient Bis-Tris gel (Invitrogen). Protein was transferred to a 0.45- m PVDF Immobilon-FL membrane (Millipore) using the XCell transfer system (Invitrogen). Primary antibodies used for probing are listed below. IRDye fluorescently labeled secondary antibodies were used for detection at a dilution of 1:5000 on an ODYSSEY® Quantitative Florescent imager (LI-COR) using the manufacturer’s protocol. Bands were quantitated using Odyssey imaging software version 3.0.30. Primary Antibodies: EGFR 1:1000 dilution (Cell Signaling Technology, #4405), FRZD2 1:500 dilution (Abcam, #52565), IRS1 1:1000 dilution (Cell Signaling Technology, #3407), ITGA6 1:1000 dilution (Cell Signaling Technology, #3750), IGF1R 1:1000 dilution (Cell Signaling Technology, #3018), MET 1:1000 dilution (Cell Signaling Technology, #8198), Erk1/2 1:1000 dilution (Cell Signaling Technology, #9102), pi-AKT Thr308 and Ser473 1:1000 dilution (Cell Signaling Technology, #????, 4060) Src 1:1000 dilution (Cell Signaling Technology, #2110), pi-Src Tyr416 1:1000 dilution (Cell Signaling Technology, #2101), Stat3 1:1000 dilution (Cell Signaling Technology, #9139), pi-Stat3 Ser727 1:1000 dilution (Cell Signaling Technology, #9134) and b-tubulin (Sigma-Aldrich, #T8328).

***Luciferase Reporter Assays***

Vectors encoding the wild-type or mutant 3’UTR of EGFR, IGF1R, MET, and IRS1, cloned behind *Renilla* luciferase, were purchased from Switchgear Genomics. Cells were seeded at 1x104perwell in 96-well plates with white well bottoms. The next day, 1.5 nmol/L of microRNA mimics was transfected using 0.15 L of RNAiMAX. After a 24 h incubation the microRNA transfection media was removed and 100 ng of vector was transfected using 0.1.5 L of DharmaFECT Duo transfection reagent (Thermo Scientific) in 100 uL of media. Cells were incubated for an additional 24 h. For normalization of cell number, 50 L of WST-1 reagent (Roche) was added to each well, and cells were incubated for 30 min at 37 C. The assay was performed as per the manufacturer’s protocol. Luciferase activity was detected using the Renilla-Glo® Luciferase Assay System (Promega) following the manufacturer’s instructions. Relative luciferase activity was normalized to florescence WST-1 viability readings for each well. All measurements represent the mean of 6 replicates in each experimental condition.

***XTT Proliferation Assay***

Cells were seeded at 2x103 /well in 96-well plates and were reverse transfected with 15 nM oligonucleotide for 48 h with 0.15 L of RNAiMAX as described above. Following transfection, 200 L of control or media containing 2 M cisplatin was placed on the cells for 3 h. Cells were washed with warm media, and then fresh media was added. Cell proliferation was assayed on the indicated days with a sodium 3’-[1-(phenylaminocarbonyl)-3,4-tetrazolium]-bis (4-methoxy-6-nitro) benzene sulfonic acid hydrate (XTT) Cell Proliferation Kit (Roche Diagnostics), following manufacturer’s instructions. XTT assay reagent was added for 4 h prior to assay. At each time point, absorbance was read at 450 nM and 655 nm, and absorbance was calculated. All time points represent the mean of 6 replicates in each experimental condition.

***Migration Assay***

Cells were seeded at 4x105 cells/ well in 6-well plates and reverse transfected with 15 uM oligonucleotide for 48 h as described above. After transfection, the media was replaced with media containing 1% serum to limit effects due to proliferation, and two scratchs devoid of cells was created in each well laterally and longitudinally with a p1000 tip. Four marked locations in each scratch were imaged at 100x magnification at time points 0- 20hrs as indicated, during which there were no significant differences attributable to proliferation in XTT assays. The area of the scratch was determined using ImageJ software (12), and the percent migration into the empty area over time was calculated.

***MATRIGEL Invasion Assay***

Cells were seeded into 6-well plates and were reverse transfected with 15 nM oligonucleotide for 48 h with RNAiMAX as described above. Following transfection, cells were trypsinized and suspended in DMEM without additives. Biocoat Growth Factor Reduced Invasion Chambers were prepared as per manufacturer’s instructions (Becton, Dickinson). 5 x 104 cells were placed in the top of each chamber. The bottom sides of chambers were placed in wells containing 100 ng/mL rEGF (Millipore) as a chemoattractant in DMEM. Chambers were incubated for 24 h at 370C. Non-invading cells were removed by scrubbing the top of invasion membranes, and invading cells were stained with 0.05% crystal violet solution in methanol (Sigma) for 1 min. Three replicate invasion membranes were mounted on glass slides and invading cells counted at 100x magnification.

***Colony Formation Assay***

Cells were seeded into 6-well plates and were reverse transfected with 15 nM oligonucleotide for 48 h with RNAiMAX as described above. Following transfection, cells were trypsinized and re-plated in 6-well plates at varying densities. Cells were incubated for 11 days and then stained with 0.1% crystal violet/methanol solution. Colonies with >50 cells were counted in three replicate wells, and the fraction of surviving cells was calculated.

***Immunofluorescence***

Fresh tumors were embedded in OCT and then frozen immediately on dry ice. Tumor tissues were sectioned into 5 m sections. Sections were fixed for 7 minutes at -200C with ice-cold methanol (EMD Millipore Corporation, Billerica, MA). Samples were then washed three times with PBS. Sections were blocked by incubation in a humidifying chamber at RT for one hour with blocking solution 1 (3% BSA + 0.05% Tween 20 in 1X PBS) followed by a one-hour incubation with blocking solution 2 (10% NGS in 1X PBS). Sections were then incubated with primary antibody diluted in dilution solution (1% BSA + 0.1% Tween 20 in 1X PBS) overnight at 4°C in a humidifying chamber. After washing the cells five times with 1X PBS, the slides were mounted with Vectashield mounting medium with DAPI (Vector Laboratories Inc, Burlingame, CA) in the dark. Samples were analyzed on a LSM 780 confocal microscope (Carl Zeiss Microimaging, Thronwood, NY). Confocal data was analyzed using Zen 2012 SP1 (black edition) software and the degree of color intensity was ascertained using Zen 2012 (blue edition) software.

***Development of miR-30a Nanocomplex Bearing an Anti-transferrin Receptor Single-chain***

***Antibody Fragment***

Fluorescent siRNA to test nanoparticle delivery *in vivo* was synthesized by Trilink Biotechnologies, and the formulation of oligonucleotides into a cationic liposomal nanodelivery system (scL) was performed as previously described (13,14). Briefly, 1:1 molar ratios of each single-stranded antisense and cognate sense oligonucleotide were annealed. Cationic liposome (dioleoyltrimethylammonium phosphate [DOTAP] and dioleoylphosphatidylethanolamine [DOPE][Avanti Polar Lipids, Alabaster, AL]) was prepared at a 1:1 molar ratio by ethanol injection. The anti-transferrin receptor single-chain antibody fragment (TfRscFv) was mixed with the liposomes at the previously established ratio of 1:30 (w/w) (15). The miR mimic were subsequently added to the admixture at a ratio of 1 μg siRNA to 7 nmol liposome, followed by sizing and confirmation of nanosize particle distributions in the final immunoliposome formulations by dynamic light scattering with a Malvern Zetasizer 3000 HS (Malvern, Worcestershire, UK). The control miR mimic had a guide strand sequence of 5’ UUGUACUACACAAAAGUACUG 3’ and a passenger strand sequence of 5’ CAGUACUUUUGUGUAGUACAA 3’, which is based on *C. elegans* cel-miR-239b, a miRNA that has minimal sequence identity with miRs in human, mouse, and rat. The miR-30a-5p double-stranded mature strand mimic contained a guide strand sequence of 5’ UGUAAACAUCCUCGACUGGAAGCU 3’ and a passenger strand sequence of 5’ AGCUUCCAGUCGGAUGUUUACACG 3’, and was synthesized by Trilink Biotechnologies. Following annealing the mimic was formulated as described above. The complexed miR-30a mimic is referred to as miR-30a-scL.

**In Vivo *Tumor Targeting and Growth Assays***

All animal experiments were carried out under protocols approved by the Animal Care and Use Committee of the NIDCD, and were in compliance with the Guide for the Care and Use of Laboratory Animal Resource (16), National Research Council. Six to eight week old athymic nu/nu female mice (obtained from the Frederick Cancer Research and Development Center, NCI) were injected subcutaneously (s.c.) with 2.5 x 106 UM-SCC-46 or UM-SCC-47 cells in 100 µL of 30% Type 3 BME Cultrex(Trevigen)/MEM media on the right leg. Once tumors reached ~150 mm3(approximately 1 - 2 weeks after injection), mice were randomized into treatment groups of (n=5 mice each); control 5% sucrose (vehicle), control miR-scL, and miR-30a-scL. Nine doses of 3 mg/kg miR-30a-scL was administered via tail vain injection on Monday, Wednesday, and Friday (MWF) over three weeks for a total of nine doses. Tumor size was measured on MWF with external calipers and volume calculated with the formula V = ½ L\*W2. Tumor growth was reported as mean volume with standard error of the mean. Kaplan-Meier survival analysis was performed in Graphpad PRISM software (v6.05). Survival statistics were performed using the Log-rank (Mantel-Cox) test, and Hazard ratio calculated via Log-rank test.

**Supplemental References**

1. Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S*, et al.* Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* **2008**;26:407–15.

2. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **2001**;98:5116–21.

3. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**;517:576–82.

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

5. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D*, et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* **2007**;104:20007–12.

6. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G*, et al.* Integrative genomics viewer. *Nat Biotechnol* **2011**;29:24–6.

7. Marsico A, Huska MR, Lasserre J, Hu H, Vucicevic D, Musahl A*, et al.* PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol* **2013**;14:R84.

8. Therneau TM. A package for survival analysis in S, R package version 2.38. 2015. <http://CRAN.R-project.org/package=survival>.

9. Brenner JC, Graham MP, Kumar B, Saunders LM, Kupfer R, Lyons RH*, et al.* Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines. *Head Neck* **2010**;32:417–26.

10. Friedman J, Nottingham L, Duggal P, Pernas FG, Yan B, Yang XP*, et al.* Deficient TP53 expression, function, and cisplatin sensitivity are restored by quinacrine in head and neck cancer. *Clin Cancer Res* **2007**;13:6568–78.

11. Chung N, Zhang XD, Kreamer A, Locco L, Kuan PF, Bartz S*, et al.* Median absolute deviation to improve hit selection for genome-scale RNAi screens. *J Biomol Screen* **2008**;13:149–58.

12. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **2012**;9:671–5.

13. Pirollo KF, Zon G, Rait A, Zhou Q, Yu W, Hogrefe R*, et al.* Tumor-targeting nanoimmunoliposome complex for short interfering RNA delivery. *Hum Gene Ther* **2006**;17:117–24.

14. Pirollo KF, Rait A, Zhou Q, Hwang SH, Dagata JA, Zon G*, et al.* Materializing the potential of small interfering RNA via a tumor-targeting nanodelivery system. *Cancer Res* **2007**;67:2938–43.

15. Yu W, Pirollo KF, Yu B, Rait A, Xiang L, Huang W*, et al.* Enhanced transfection efficiency of a systemically delivered tumor-targeting immunolipoplex by inclusion of a pH-sensitive histidylated oligolysine peptide. *Nucleic Acids Res* **2004**;32:e48.

16. Guide for the Care and Use of Laboratory Animal Resource. 1996.

**Supplementary Figure Legends**

**Supplementary Fig. S1. Overall workflow for Analysis and approach.**

miRNA expression profiling from TCGA was validated against miRNA from an independent set of University of Michigan squamous carcinomas (UMSC), and compared for overlap with miRNA genome-wide anti-proliferation screen, as described in Methods. This uncovered decreased miR30-a, e family expression in tumors compared to normal mucosa. mRNAs inversely correlated with the most differentially expressed family member miR-30a were determined and filtered through miR databases, identifying 94mRNAs with predicted miR30 binding sites. Selected targets were validated by functional assays in vitro, and miR30a mimic nanoparticles were tested in xenograft models in vivo. The role of gene copy number and methylation in expression of decreased miR30-a/e was determined from TCGA data, and clinical correlation of expression with disease specific survival was determined from an independent dataset as described.

**Supplementary Fig. S2. Differentially expressed microRNAs in HNSCC.** Genome-wide microRNA profiling wasperformed using small RNA high throughput sequencing of human HNSCC tissues. miRs that were differentially abundant between tumor and adjacent normal samples in TCGA and UMSC samples were determined using SAMseq v2.0. Results are from an unpaired analysis, samples from TCGA: tumor=244, mucosa=16; samples from UMSC: tumor=13, mucosa=9. Thirty three microRNAs that were identified as differentially expressed by SAMseq in both the TCGA **(**A) and UMSC **(**B**)** HNSCC tumor cohorts are presented when compared with mucosa controls. Left: fold change of median expression between tumor and mucosa is displayed, and presented by linear scale. Right: box and whisker plot of median expression distribution of mucosa and tumor presented on right as log10 RPM (reads per million base pairs). Medians are represented by the thick black lines in the middle, bars represent 25th and 75th percentile, and outliers are displayed as individual points. FDR cutoff is 0.2.

**Supplementary Fig. S3. miR-30 family effects on proliferation and relative expression in UM-SCC lines, and anti-correlation with selected mRNAs in HNSCC tumors. (A)** Proliferation was measured in UM-SCC-1 or UM-SCC-46 cells by an XTT assay in 6 replicates at day 5 following transfection with control miR or a miR-30-5p family members a-e. Similar inhibition of proliferation is observed between family members sharing the same seed sequence. Values represent the mean of six replicate transfections and error bars represent SEM. (\*) denotes a p-value < 0.05 by Student’s t-test. **(B)** Basal level of miR-30a-5p and miR-30e-5p expression in HOK, UM-SCC-1 and UM-SCC-46 cell lines measured by qRT-PCR. Total RNA was collected in three replicate experiments from HOK, UM-SCC-1 and UM-SCC-46 cell lines in log-growth phase.  Similar to HNSCC tumor samples, miR30a-5p is the more highly expressed miR30 family member in HOK and UM-SCC cell lines and is significantly reduced in UM-SCC-46 compared to HOK and UM-SCC-1 cells. Values represent mean expression, and error bars represent SEM of three independent experiments, and (\*) denotes a p-value < 0.05 by Student’s t-test. **(C)** Significant anti-correlations between expression of miR-30a-5p and putative mRNA targets in HNSCC tumors. Expression of miR-30a-5p (log2 RPM) was inversely correlated with mRNA expression [log10 RSEM (RNA-Seq by Expectation Maximization)] from 279 HNSCC in the TCGA dataset, and filtered for mRNAs containing predicted miR-30 binding sites. Linear regression scatterplots are presented for selected mRNAs with p-values . Shaded bands show 95% confidence intervals on the best-fit lines.

**Supplementary Fig. S4. Ectopic expression of miR-30a-5p inhibits direct target mRNA, protein expression, and downstream signaling. (A)** Selected miR-30-5p target genes were validated by qRT-PCR measurement in UM-SCC-47 cells transfected with miRNA negative control (neg Con), miR-30a-5p, or anti-miR-30a-5p oligonucleotide for 72 hrs. The mean of three independent experiments, ±SEM; \* denotes p< 0.05 by Student’s t-test. Protein levels of miR-30a-5p target genes were analyzed from triplicate experiments by western blots with transfection of 15 nM control oligo, miR-30a-5p, or anti-miR-30a-5p into UM-SCC-46 or UM-SCC-47 cells. Relative band intensity was quantitated and normalized to β-tubulin levels for direct targets **(B)** or phosphorylation levels of downstream signaling proteins **(C)**. Values represent mean expression, and error bars represent SEM of three independent experiments, and (\*) denotes p-value < 0.05 by Student’s t-test.

**Supplementary Fig. S5. Ectopic expression of miR-30a-5p inhibits proliferation of HNSCC. (A and B)** Proliferation was measured in UM-SCC-46 cells by an XTT assay in 6 replicates at days 0, 1, 3 and 5 following transfection with control, miR-30a-5p, or its anti-miR, or in combination with cisplatin treatment at the IC50 dose**.** (**C)** UM-SCC-1 cells were transfected with miR-30a-5p mimic for 48 hrs, and treated with 3.5 µM cisplatin for 3 h and then washed with warm media and cell density was measured by XTT assay 72 h after cisplatin treatment. The mean of at least three experiments ±SEM, \* denotes p< 0.05 by a Student’s t-test. (**D)** Representative images of colony formation assays with control, miR-30a-5p, or anti-30a transfections. **(E)**A UM-SCC-46 subline that stably overexpresses EGFR-GFP fusion protein without a 3’UTR was created to assess compensation for the anti-proliferative activity of miR-30-5p. EGFR protein expression was determined by Western blots. **(F)** Proliferation by XTT assays on day 5 following transfection of a miR-30a-5p mimic. The proliferation rate was statistically significantly higher in the EGFR stable following miR-30-5p transfection. Error bars represent SEM, and (\*) denotes p-value < 0.05 by Student’s T-test.

**Supplementary Fig. S6. *In vivo* knockdown of miR-30-5p target genes with a therapeutic mimic.** Mice were implanted with UM-SCC-46 xenograft tumors. When tumors reached ~150 mm3, mice were randomized and injected IV with four doses of 60 µg of control miR-ScL or miR-30a-ScL on MWF schedule. On day 19, 36 hrs after the fourth treatment, mice were sacrificed and tumor tissues were collected for total RNA extraction. **(A)** Gene expression of miR-30a-5p targets by real-time PCR. Data represents the mean of 3 animals, error bars represent SEM, and (\*) denotes p-value < 0.05 by student’s t-test. **(B, C)** Representative images and quantification of UM-SCC-46 xenograft tumors stained for Ki-67 by immunohistochemistry. Values represent mean intensity quantified from six independent 20x fields and error bars represent ±SEM, (\*) denotes p< 0.05 by a student’s t-test.

**Supplementary Figure S7**. **Decreased expression of miR-30a-5p target genes in tumor specimens after treatment by miR-30a-5p nanoparticles *in vivo*.** Immunofluorescent staining of EGFR and MET were performed in frozen sections harvested from xenograft tumors after control miR-scL or miR-30a-scL treatment *in vivo*. On day 25, 36 h after the last treatment, tumor tissue was collected for molecular analysis. Representative microscopy images with decreased immunofluorescent staining intensities of EGFR and MET were observed in miR-30a-scL treated **(A)** UM-SCC-46 and **(B)** UM-SCC-47 samples. Scale bars, 20 µm. **(C)** Mean florescence intensity was quantified from six independent 40x fields and error bars represent ±SEM, (\*) denotes p< 0.05 by a student’s t-test. **(C)**

**Supplementary Figure S8**. **Ingenuity Pathway Analysis networks constructed miR-30-5p target mRNAs with reported interactions and function in relation to proliferation and migration**. Molecules denoted in red are validated miR-30a-5p target genes with inverse relationships to miR-30a-5p expression in HNSCC tumors. Molecules denoted in blue are direct binding or signaling intermediaries.

**Supplementary Fig. S9. Analysis of miR-30a/e-5p expression in relation to survival.** Kaplan-Meier plots and log rank test p-values comparing overall survival. **(A)** for miR-30a-5p expression segregated into high and low by median expression from the TCGA dataset. **(B)** for miR-30e-5p. Left, expression segregated into high and low by median expression; middle, copy number deletion of MIR30E gene locus versus non-deletion; right, median low vs high miR-30e-5p expression in patients with oropharynx cancer from the TCGA dataset. **(C)** Median low vs high miR-30e-5p expression in patients with oropharynx cancer from the Einstein dataset.

**Supplementary Table S2. Tumor, treatment, and outcome characteristics of human HNSCC specimens.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Specimens Gender Age Primary sites Stage/TNM Differentiation Tobacco/pack Alcohol/Quit**

----------------------------------------------------------------------------------------------------------------------------------------------------------------------------

UMSC 2900 M 57 Lateral Tongue T2N0M0 Moderate NA NA

UMSC 3100 M 75 Anterior Tongue T1N0M0 Poor MD MD

UMSC 3300 F 60 Lateral Tongue T3N1M0 Moderate NA NA

UMSC 4300 F 47 Lateral Tongue T3N0M0 Well Y/14 NA

UMSC 4500 F 25 Anterior Tongue T4N2cM0 Moderate NA NA

UMSC 8200 M 72 Tonsil T4N0M0 Well Y/150 Y

UMSC 8400 M 44 Lateral Tongue T2N0M0 Well Y/20 Y/Y

UMSC 8500 F 40 Lateral Tongue T2N0M0 Well NA NA

UMSC 8800 M 47 Floor of Mouth T4N2bM0 Moderate Y/45 Y

UMSC 4400 F 41 Floor of Mouth T1N0M0 Well Y/60 Y/Y

UMSC 7300 M 55 Floor of Mouth T4N2cM0 Well Y/30 Y

UMSC 7500 F 71 Hard Palate T4N0M0 Moderate NA NA

UMSC 7800 M 55 Lateral Tongue T4N2bM0 Poor Y/60 Y/Y

UMSC 8300 F 50 Lateral Tongue T2N0M0 Well Y/28 NA

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

HNSCC tumor specimens from oral cavity were obtained from University of Michigan and designated as UMSC. Clinical information was provided by Dr. Douglas Chepeha. Primary sites, the origin of the primary tumor; TNM, tumor-node-metastasis (staging system). Y: Yes; NA: not available.