Am J Transl Res 2014;6(5):580-592
www.ajtr.org /ISSN:1943-8141/AJTR0001908

Original Article

Prognostic value of miR-375 and miR-214-3p in early stage oral squamous cell carcinoma

Angela J Yoon[1], Shuang Wang[2], Jing Shen[3], Nicolas Robine[4], Elizabeth Philipone[1], Martin W Oster[5], Albert Nam[6], Regina M Santella[3]

[1]Division of Oral and Maxillofacial Pathology, Columbia University Medical Center and College of Dental Medicine, New York, NY, USA; Departments of [2]Biostatistics, [3]Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY, USA; [4]New York Genome Center, New York, NY, USA; [5]Division of Hematology/Oncology, Columbia University Medical Center, New York, NY, USA; [6]University of Chicago, Chicago, IL, USA

Received August 15, 2014; Accepted September 4, 2014; Epub October 11, 2014; Published October 15, 2014

Abstract: MicroRNAs (miRs) control cell growth, apoptosis and differentiation, and thus play a key role in carcinogenesis. Identification of a set of miRs that demonstrate differential expression in oral squamous cell carcinoma (OSCC) patients with poor prognosis has potential for utility as a prognostic marker. A retrospective study of miR expression was conducted in 20 tissue samples from early stage (Stages I & II) OSCC patients with known clinical outcome (10 from those who had 5-year disease free survival and 10 who died of disease within 5 years) using genome-wide deep sequencing analysis. The promising miR candidates were then validated in 80 tissue samples using quantitative real-time PCR (qRT-PCR). The deep sequencing and qRT-PCR analysis identified two promising miRs, miR-375 and miR-214-3p. Combining the two miRs as a panel with age and gender had a predictive value for the area under the curve (AUC) of 0.932, with a sensitivity of 87.5% and a specificity of 87.2% ($p<0.0001$) to identify patients with poor prognosis. A miR-based prognostic risk score model was constructed, which included the miR-214-3p, miR-375, age and gender, each weighed by relative contribution. The risk score model was able to identify high-risk individuals who had significantly shorter time to relapse ($p<0.001$) and time to death ($p<0.001$). The model consisting of a two-miR panel with age and gender may be useful in prognostication of early stage OSCC patients, which can aid in identifying patients with poor prognosis who will benefit from a subsequent aggressive treatment regimen.

Keywords: microRNA, deep sequencing, qRT-PCR, prognosis, risk score, oral squamous cell carcinoma

Introduction

An estimated 30,000 people in the United States (US) are diagnosed with oral squamous cell carcinoma (OSCC) each year [1-4]. Leading etiologic factors include tobacco, alcohol and high-risk human papillomavirus (HPV types 16 and 18) infection [2, 4, 5]. Half of those newly diagnosed with OSCC die, accounting for 1 death every hour in the US [1-4].

Of the newly diagnosed oral cancer cases, ~80% are TNM Stages I & II without regional lymph node involvement or distant metastasis [4]. Surgical excision of the tumor with or without neck dissection is the treatment of choice for Stage I/II oral cancers [4, 6]. Following surgery, adjuvant radiation therapy and/or chemotherapy are administered if the surgical margins were positive, if it had high-grade histologic features, if there was bone invasion, and/or if perineural/angiolymphatic invasion was evident histologically [7, 8]. However, the aforementioned criteria may be inadequate in the decision-making process regarding adjuvant radio/chemotherapy since the survival rate has not improved significantly over the last several decades. Death occurs in 32-47% of Stage I/II individuals within a 5-year period [4]. Although some may die of the disease as many as 10 years after the initial treatment, the great majority of deaths occur within the first 5 years [4].

It is imperative to discover reliable and accurate prognostic indicators of patient survival so that those with poor predicted clinical outcome with surgery alone may receive adjuvant radio/che-

Table 1. Demographic and clinical characteristic of patients

Variables	Deep Sequencing (n=20)		qRT-PCR Validation (n=80)		
	Favorable-prognosis Group (n=10)	Poor-prognosis Group (n=10)	Favorable-prognosis Group (n=40)	Poor-prognosis Group (n=40)	*p*-value*
Age	61.3 (range: 44-75)	67.9 (range: 51-83)	62.0 (range: 42-82)	65.5 (range: 30-83)	0.2293
Gender					0.2481
Male	6	4	28	23	
Female	4	6	12	17	
TNM Stage					0.6525
I	8	5	24	21	
II	2	5	16	19	
Histologic Grading					0.0011
Well-differentiated	6	2	23	9	
Moderately-differentiated	3	7	12	29	
Poorly-differentiated	1	1	5	3	
HPV Status					0.3271
Positive	0	0	3	0	
Negative	3	2	6	6	
Unknown	7	8	31	34	
Neck Dissection					0.0069
Yes	6	6	27	13	
No	4	4	13	27	
Smoking Status					0.0481
Current	3	4	11	13	
Past	1	2	0	5	
Never	6	4	29	22	
Alcohol Abuse					0.0047
Current	1	3	1	8	
Past	0	1	0	5	
Never	9	6	39	28	

*p-value is for the 80 qRT-PCR validation set.

Table 2. Mean expression levels of 4 miRNAs and the fold change between the two groups by deep sequencing (n=19) and qRT-PCR (n=79)

		miRNA expression levels[a]			
miRNAs	Methods	Favorable Prognosis	Poor Prognosis	Fold-change[b]	*p*-value[c]
miR-375	Deep Seq	3003.9	397.3	-2.92	0.001
	qRT-PCR	-3.77	-1.63	-4.4	0.001
miR-214-3p	Deep Seq	238.9	42.3	-2.50	0.018
	qRT-PCR	-5.52	-3.08	-5.4	0.0001
miR-199a-5p	Deep Seq	1694.5	422.7	-2.00	0.045
	qRT-PCR	-1.68	-0.37	-2.8	0.0001
miR-1269a	Deep Seq	7.27	49.5	2.77	0.012
	qRT-PCR	2.23	2.66	-1.3	0.462

[a]Mean count for deep sequencing and mean normalized Ct value for qRT-PCR. [b]Log-fold change for deep sequencing and fold change ($-1/2^{-\Delta\Delta Ct}$) for qRT-PCR. [c]*p*-value is based on the two-sample *t*-test comparing the mean normalized miR expression levels.

motherapy in order to improve the overall survival. Because 80% of oral cancer patients are in early stages (Stage I/II) at the time of diagnosis, a window of opportunity exists in which proper prognostication and subsequent decisions for additional treatment may dramatically

improve the survival rate associated with this deadly disease.

MicroRNAs (miRs) are small, 18-24 nucleotide long, non-coding RNA molecules that regulate expression of the targeted genes either by facilitating mRNA degradation or by repressing translation [9, 10]. One microRNA is capable of binding over 100 different mRNAs with different binding efficiencies and plays a crucial role in their post-transcriptional regulation [11]. miRs control cell growth, apoptosis and differentiation, and various types of cancer have demonstrated distinct miR expression profiles [9, 12]. Thus far, a number of miRs associated with clinical outcome have been reported for lung, breast, gastric, and pancreatic cancers, as well as OSCC/head and neck cancers (HNSCC, 60% of which is OSCC) [9-34]. Discovery of signature miRs for cancer diagnosis, prognosis and pharmacogenomics is currently an emerging avenue of investigation, as these small molecules have potent regulatory potential in various molecular pathways underlying cancer [9, 12].

A limited number of studies evaluated the profiles of differential miR expression in OSCC tissues and/or cell lines using microarrays or quantitative reverse transcription-polymerase chain reaction (qRT-PCR) [10, 14, 16, 17, 19-24]. These profiling studies demonstrated specific microRNA expression profiles correlating with pathogenesis, metastasis, and chemoresistance of OSCC. Next-generating sequencing technology can detect all miRs present in samples, thereby increasing testing coverage [35]. Moreover, the predictive accuracy of a single biomarker can be greatly enhanced when grouped into a panel with additional markers [36].

The aim of our study was to discover a miR marker panel prognostic of 5-year survival in early stage OSCC patients that may be utilized in parallel with current clinical covariates. We assessed differential expression of microRNAs genome-wide via deep sequencing in tumor tissue samples. We also attempted to identify deregulated miR expression signatures that may serve as a prognostic marker of cancer survival. A miR marker-based panel may serve as a guide for selection of appropriate follow-up chemo/radiation treatments, significantly improving the clinical management of OSCC and overall survival. To the best of our knowledge, this is the first study to assess genome-wide miR expression to identify a panel of miRs that may be utilized as a cancer survival prognostication modality in early stage OSCC.

Materials and methods

Subjects

Following institutional review board approval at the Columbia University Medical Center, 100 patients ≥21 years old, newly-diagnosed with primary OSCC Stage I or II and with a minimum of 5-year clinical outcomes information were identified. Based on the medical record, only those subjects who underwent surgical treatment with curative intent without adjuvant chemo or radiation therapy were included. The cases were placed into the following two groups; Group 1 (OSCC patients with 5-year disease-free survival following surgical treatment) and Group 2 (OSCC patients who died of the disease within 5-years following surgical treatment).

We obtained demographic information including age, gender, tobacco use (current, past, never smokers) and alcohol intake (current, past, never drinker) from the medical charts. Clinical information including the initial stage of OSCC (Stage I or II), site of tumor (limited to oral cavity; tongue, floor of mouth, gingiva, etc.), HPV status, pathologic grade (well/moderately/poorly-differentiated), history of neck dissection, the presence of occult lymph node metastasis and history of other medical problems were elicited from existing medical records for all patients. For the subjects in the favorable prognosis group, a thorough medical record review was conducted to ensure that these individuals had a minimum of 5-year survival free of disease from the time of initial surgical treatment. For the poor prognosis group, time from initial surgical treatment to death was recorded. Information on relapse, whether locally within the oral cavity, in regional lymph nodes, or at a distant site was recorded and the subsequent treatment received was noted. Archived formalin-fixed paraffin-embedded (FFPE) tissue blocks were retrieved for the identified subjects. In case the subject had recurrent and/or second primary OSCC, the initial OSCC surgical tissue sample was utilized for the analysis.

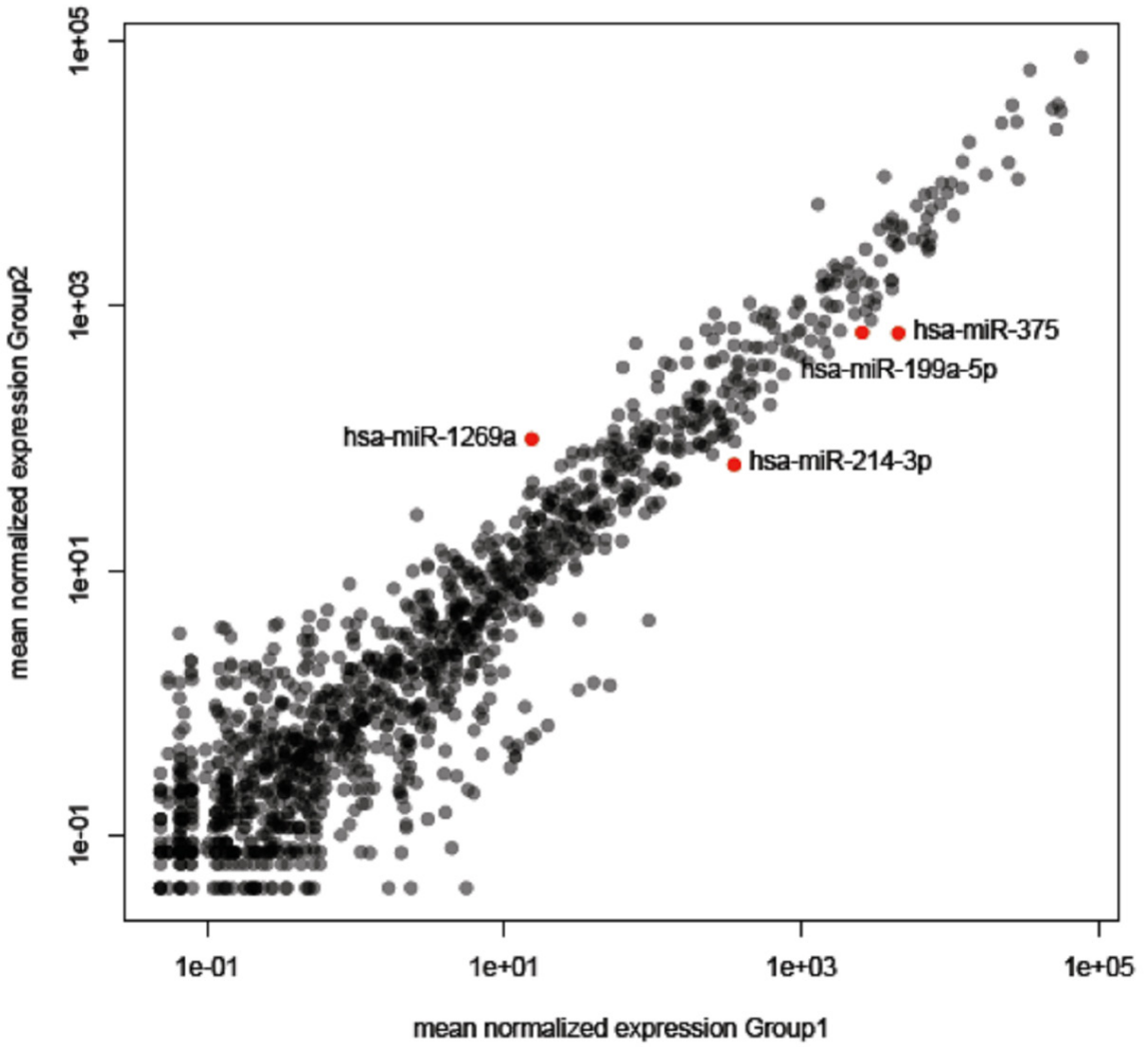

Figure 1. Scatterplot representing the average expression (post-normalization) of Group 1 (favorable 5-year survival) and Group 2 (poor 5-year survival).

Total RNA extraction

Ten 10-μm sections were obtained from archived FFPE tumor tissue samples for all subjects. For each sample, a representative section was stained with H&E and reviewed by a pathologist to identify regions containing >90% malignant epithelial cells for macrodissection. Total RNA was isolated from tissues using RNeasy FFPE kit (Qiagen Inc., Valencia, CA) following the manufacturer's protocol, yield was quantitated by Nanodrop, and samples were stored at -80°C.

microRNA quality control, library preparation and sequencing

Deep sequencing analysis was carried out on 20 OSCC tissue samples, 10 from the favorable prognosis group and 10 from the poor prognosis group. For quantification and quality control, total RNAs were tested using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and the Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY). Total RNA (minimum of 2 μg) ranging from 18 to 30 nt were gel-purified and ligated to 3'- and 5'-adaptors. Ligation products were reverse transcribed, then amplified for 16 cycles using the adaptor primers, and the fragments around 150 bp were isolated from PAGE-gel using TruSeq Small RNA Sample Prep Kit (Illumina, San Diego, CA). Libraries were sequenced on an Illumina HiSeq 2500 platform-50 SR that allows for 1 x 50 base-pair single-end reads. The data was deposited in the publicly available Gene

Expression Omnibus (GEO) database (GSE5-2633).

microRNA mapping and differential expression analysis

Adaptors were removed and low quality tags were filtered with FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were processed with the pipeline miraligner, which mapped them to the miRBase v.20 sequences. The number of reads per miR was first assessed and analyzed using the bioconductor package DESeq to compare miR expression between the favorable prognosis group and the poor prognosis group [36]. In addition to normalizing between the samples, DESeq performed a statistical test of differential expression under the hypothesis of a negative binomial distribution of the reads. To select the top miRs with prognostic value, those with significant ($p<0.05$) ≥2 log2-fold change of the different mean normalized expression levels between the two groups were identified. The association between selected miRs and cancer prognosis was also investigated through a literature search.

qRT-PCR for microRNAs

The expression levels of the top four miRs (miR-375, miR-214-3p, miR-199a-5p, miR-1269a) selected from the deep sequencing analysis were quantified using TaqMan MicroRNA Reverse Transcription Kits (Life Technologies, Grand Island, NY) following the manufacturer's protocols in an additional 80 OSCC tissue samples (validation set; 40 in the favorable prognosis group and 40 in the poor prognosis group). In brief, RNAs were eluted with 100 μl of RNase-free water, concentrated in a final volume of 20 μl and quantified by NanoDrop ND-1000. Input RNA was reverse transcribed using the TaqMan miR Reverse Transcription Kit and miR-specific stem-loop primers for the selected four miRs in a small-scale RT reaction. For quantification, diluted RT product was combined with PCR assay reagents and real-time PCR carried out on an ABI7900HT thermocycler. The endogenous controls (U6 snRNA and RNU48) initially used showed significantly different expression levels between the two groups of interest. Based on the literature [37], and also based on the deep sequencing data searching for a miR that showed the most consistent expression across all samples in the two groups of interest, miR-16 was selected as the endogenous control and was used to normalize the relative expression of target microRNAs. The assays were validated by repeating the assay on three different days for a minimum of 10 samples. The coefficient of variation was calculated and values <5% was considered acceptable. Test samples were assayed in duplicate with the laboratory blinded to survival status and with 5% triplication after relabeling. Data was analyzed with SDS Relative Quantification Software version 2.2.2 (Applied BioSystems, Foster City, CA) to determine the threshold cycle (Ct). The fold change was determined by the $2^{-\Delta\Delta Ct}$ method [38]. A two-sample *t* test was used to compare the normalized expression levels between the two groups.

microRNA target prediction and pathway analysis

Gene ontology analysis (www.pantherdb.org) was performed by the PANTHER classification system to compare potential target genes affected by the top four miRs with the NCBI reference (human genome build 36). The binomial test was used to identify significantly enriched pathways, biologic processes, molecular functions, cellular components, and protein class terms after Bonferroni correction for multiple comparisons with a cutoff of $p\leq0.001$.

Statistical analysis

In the qRT-PCR validation set, univariate logistic regression was first used to test if each of the selected microRNAs was an independent prognostic factor for 5-year survival status. Multiple logistic regression was then used to build a prediction model with a panel of miRs adjusting for age and gender. Using the same multiple logistic regression model, we constructed receiver-operating characteristic (ROC) curve and calculated the maximum sensitivity and specificity (using 0.5 probability of poor prognosis status as the cutoff point) and the area under the curve (AUC); $p<0.05$ was considered statistically significant.

We performed multivariate Cox regression analyses separately for the time to relapse and the time to death, where the models include the selected miRs with age and gender. We then calculated mortality risk score and relapse risk score for every patient in the validation set

Table 3. Univariate analysis of microRNAs associated with poor prognosis in 79 OSCC patients in the validation set

miRNAs	OR (95%CI)	*p*-value[a]	Chromosomal Location	Putative function	Expression in poor prognosis group
miR-375	1.28 (1.08-1.51)	0.038	2q35	Protective	down
miR-214-3p	1.85 (1.37-2.49)	<0.0001	1q32.1	Protective	down
miR-199a-5p	2.63 (1.62-4.28)	<0.0001	2p21	Protective	down
miR-1269a	1.07(0.88-1.31)	0.456	4	High risk	up

[a]$p<0.05$ is considered to be significant.

from the expression values of the selected miRs, age and gender weighted by the regression coefficients obtained from the multivariate Cox regression analyses [39]. Based on the individual mortality risk score and the disease relapse risk score, the patients were stratified into high mortality risk vs. low mortality risk groups, and high relapse risk vs. low relapse risk groups using the median score as the cut-off. Kaplan-Meier curves were generated for the high vs. low mortality risk groups for the time to disease related death. Kaplan-Meier curves were also generated for the high vs. low relapse-risk groups for the time to relapse (local, regional lymph nodes, or distant metastasis). A log-rank test was used to compare two survival curves. To assess the effectiveness of neck dissection, we examined the differences in time to relapse/time to death between the high and low relapse/mortality risk groups using the Kaplan-Meier curves and the log-rank tests. A subgroup analysis was further performed for the high relapse/mortality risk groups only, comparing the time to relapse/the time to death between those with and without neck dissection. Statistical analyses were conducted using SAS 9.3 (SAS Institute).

Results

Deep sequencing, mapping of miRs and selection of prognostic microRNAs

The demographic and clinicopathologic characteristics of the 20 subjects whose tissues were analyzed by deep sequencing and the 80 analyzed by qRT-PCR are shown in **Table 1**. For deep sequencing, one sample in the favorable prognosis group did not pass the quality control and was subsequently eliminated from the future analysis. The total number of reads obtained from sequencing ranged from 3, 229, 855 to 39, 216, 964 for the 19 tissue samples. After removing adaptors and filtering out low quality tags, 973, 424-31, 446, 445 clean reads were obtained (~30.14-80.19%). From these reads, the mapping rate to miR ranged from 0.89 to 12.71, and the rate to miRaligner, which distinguishes -3p and -5p sequences, ranged from 1.48 to 14.75. The length distribution analysis revealed a peak at 22 nt, which is the size of most known microRNAs.

For each sample, ~30 thousand to 11 million sequence reads that mapped to the human genome were obtained and included miRNA, rRNA, Mt_rRNA, snoRNA, snRNA, tRNA. A total of 1910 miRs were detected in the samples using miRaligner and the number of reads of each miR ranged from 0 to 328,368. Four miRs had at least a two log2 fold change with $p<0.05$ (**Table 2**). Three out of the 4 miRs, including miR-375, miR-214-3p, and miR-199a-5p, were underexpressed in the poor prognosis group. One miR, miR-1269a, was overexpressed in the poor prognosis group (**Figure 1**).

Putative target genes for the selected miRs

The biologic characteristics of the 852 conserved genes potentially targeted by the four miRs (miR-375, miR-214-3p, miR-199a-5p, and miR-1269a) were evaluated using PANTHER ontology analysis (Supplemental Table 1). Enriched genes were significantly associated with 31 biologic pathways ($p<0.001$), including pathways involving regulation of transcription, GTPase activity, I-kappaB kinase/NF-kappaB signaling, ERK1/2 cascade, MAPK cascade, JNK cascade, Wnt signaling pathway, cell proliferation and apoptotic process, cell migration, and cell cycle. Our data confirm the broad range of targets for the four miRs potentially involved in the progression of OSCC.

Validation of microRNA expression profiles by RT-qPCR

To further confirm the miR-375, miR-214-3p, and miR-199a-5p underexpression and the miR-1269a overexpression in the poor prognosis group, these four miRs were quantified using qRT-PCR in an additional 40 OSCC tissue samples from the favorable prognosis group and 40 from the poor prognosis group. In one sample from the favorable prognosis group, the endogenous control was undetectable. Since the normalization could not be performed for the miR expression levels, this sample was eliminated from the future analyses (n=39 for the favorable prognosis group; n=40 for the poor prognosis group). Consistent with the deep sequencing analysis, three of the four miRs (miR-375, miR-214-3p, and miR-199a-5p) showed significantly decreased expression in the poor prognosis group compared to the favorable prognosis group (**Table 2**). The fold changes for miRs-375, 214-3p and 199a-5p were -4.4, -5.4, and -2.8, respectively, suggesting a potentially important role for downregulation of these putative tumor suppressors in the poor prognosis group. Univariate logistic regression analyses showed that these three miRs were also significantly associated with increased risk of death due to disease within 5 years compared with the patients in the favorable prognosis group (OR of 1.28, 1.85 and 2.63 for miR-375, miR-214-3p, and miR-199a-5p, respectively; **Table 3**). miR-1269a was shown to be a promising candidate in the deep sequencing analysis with a 2.7 fold increase in the poor prognosis group (p=0.012 from the t-test). However, in the subsequent qRT-PCR analysis, miR-1269a showed underexpression in the poor-prognosis group, which was statistically insignificant using the t-test to compare normalized expression levels between the two prognostic groups (p=0.462) or using the univariate logistic model (p=0.574).

Multiple logistic regression was used to construct a predictive model of poor prognostic status using miRs, adjusting for age and gender. Three miRs that are significantly associated with poor prognostic status in the univariate analyses were assessed together with age and gender. While miR-375 and miR-214-3p were still significantly associated with poor prognostic status, miR-199a-5p was no longer statistically significant (p=0.156). The AUC of the ROC curve with the 3 miR panel together with age and gender was 0.935, with a sensitivity of 87.5% and a specificity of 84.6% to identify individuals with poor prognosis (p<0.0001 for the 5 predictors combined).

Since miR-199a-5p was not significant in the multiple logistic model, we further examined the predictability of the two miR panel, miRs-375 and 214-3p, together with age and gender. The AUC of the ROC curve was 0.932, with a sensitivity of 87.5% and a specificity of 87.2% (p<0.0001 for the 4 predictors combined; **Figure 2**). Thus the final predictive model included miR-214-3p (OR=3.18, CI=1.88-5.38, p< 0.0001), miR-375 (OR=1.73, CI=1.25-2.39, p= 0.0008), age (OR=1.08, CI=1.02-1.15, p= 0.007) and gender (OR=1.69, CI=0.40-7.12, p=0.471). We noted that this predictive model with age and gender was a much improved model over the two miR-panel alone, which has the AUC of the ROC curve of 0.903, a sensitivity of 77.5% and a specificity of 79.5% (p<0.0001 for the 2 predictors combined). Older age and female gender were associated with worse prognosis. Moreover, combining two microRNAs provided higher sensitivity and specificity over individual miRs (miR-214-3p alone: AUC= 0.823, sensitivity=72.5%, specificity=82.1%, p<0.0001; miR-375 alone: AUC=0.674, sensitivity=62.5%, specificity=56.4%, p=0.0038).

Using the aforementioned final model (39), a prognostic risk score model was constructed, which included the two miRs (miR-214-3p and miR-375), age and gender, each weighed by relative contribution (Supplemental Table 2):

Mortality risk score=(0.182 x expression value of miR-375) + (0.370 x expression value of miR-214-3p) + (0.035 x age) + (0.570 x gender);

Disease relapse risk score=(0.167 x expression value of miR-375) + (0.289 x expression value of miR-214-3p) + (0.035 x age) + (0.394 x gender).

For gender, females were entered as 1 and male as 0. Based on this formula, the risk score was calculated for all 79 patients in the qRT-PCR validation set. The median risk scores were used as the cutoff to stratify patients into high and low risk groups.

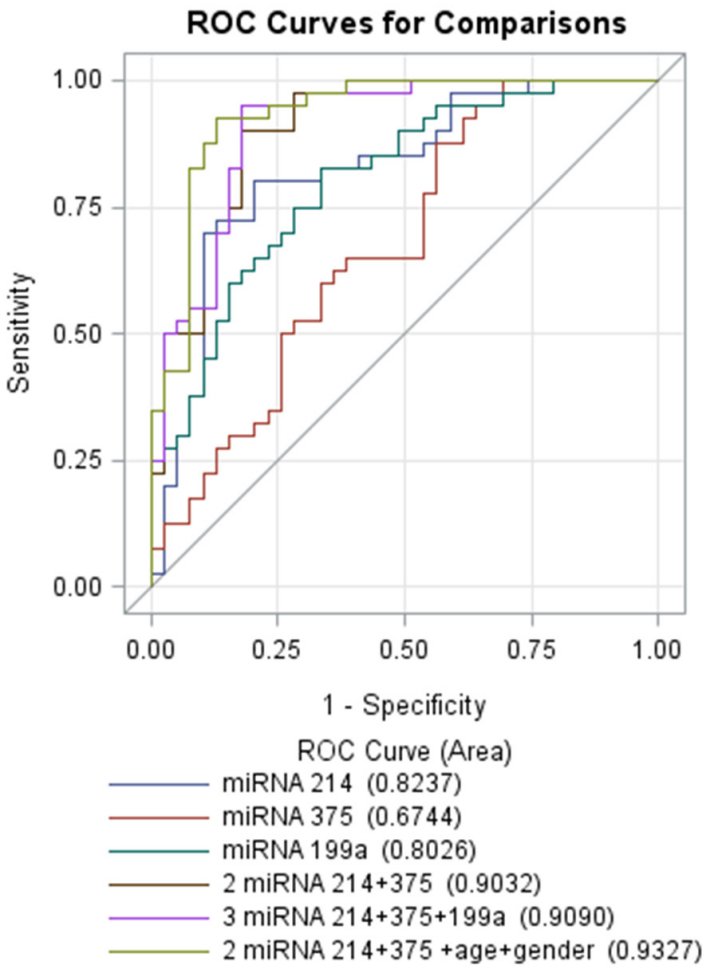

Figure 2. ROC curves for miR levels that can identify OSCC patients with poor-prognosis.

The mortality risk score ranged from -3.31 to 3.02 (median=0.49) and the disease relapse risk score ranged from -2.42 to 2.93 (median=0.77). Higher score was considered to carry greater risk for mortality and disease relapse. Using the median as the cutoff point, the morality risk score model was able to identify 34 out of 39 patients (87.2%) in the favorable prognosis group (those who had 5-year disease free survival) as the low-risk group and 36 out of 40 patients (90.0%) in the poor prognosis group (those who died of disease within 5 years) as the high-risk group. Similarly, using the median as the cutoff point, the disease relapse risk score model was able to identify 87.2% of the patients in the favorable prognosis group as low risk and 90.0% of the patients in the poor prognosis group as high risk. The high risk patients identified using the mortality risk score or disease relapse risk score did not overlap completely (One patient from the favorable prognosis group had a low mortality risk score but high disease relapse risk score).

To compare the time to disease relapse and the time to death between high and low risk groups, Kaplan-Meier curves were generated. For the patients who had 5-year disease free survival (favorable prognosis group), the time to last follow-up was used for the analysis, and, thus, those patients were censored. The patients stratified into the high risk group had a significantly shorter time to disease relapse ($p<0.001$) and time to death ($p<0.001$) (**Figure 3A**).

Additional analyses were performed to examine if the risk score model could provide some insight into selection of appropriate treatment. Considering only the patients who had concurrent neck dissection with the initial surgery, we compared the time to disease relapse and time to death between the high and low risk groups (**Figure 3B**). The patients in the high-risk group still had a significantly shorter time to relapse and time to death after neck dissection, compared to those in the low-risk group who also had neck dissection. This finding suggests that the neck dissection does not improve prognosis of those in the high-risk group. We then compared the time to relapse and time to death between those who had neck dissections and those who did not, only within the high risk group (**Figure 3C**). Although not statistically significant, the median time to relapse was longer for high-risk individuals with neck dissection (24 months vs 12 months for high-risk individuals without neck dissection), but the median time to death was the same for those with or without neck dissection (24 months for both subgroups), which further suggests that neck dissection has a minimum influence in terms of prognosis for those patients assigned the high risk scores (Supplemental Table 3).

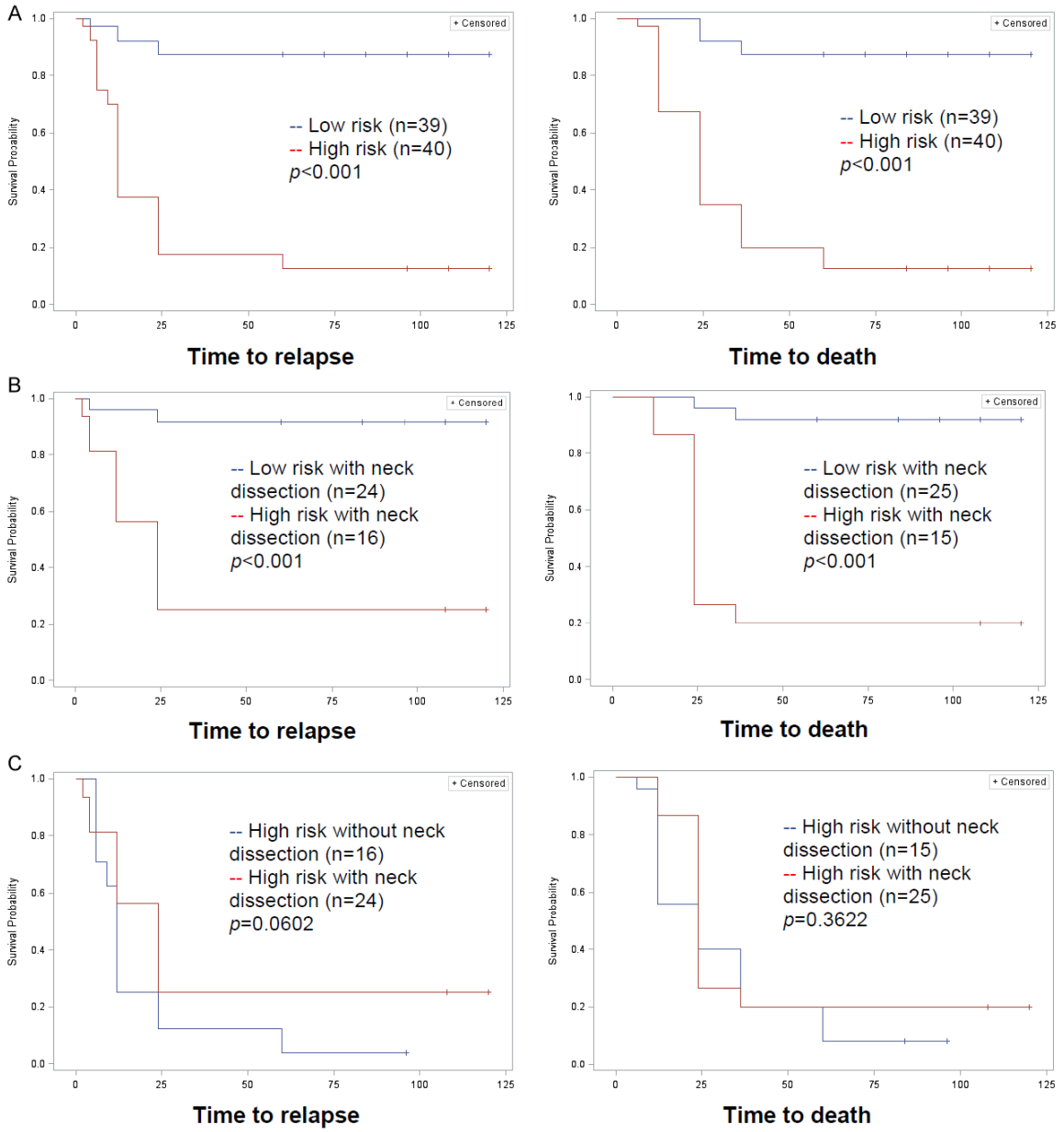

Figure 3. A. Kaplan-Meier curves of time to relapse and time to death in high risk and low risk patients based on the risk score calculation. B. Kaplan-Meier curves of time to relapse and time to death in the subgroup of patients who had neck dissection stratified into high and low risk groups. C. Kaplan-Meier curves of time to relapse and time to death only in the high risk patients with and without concurrent neck dissection.

Discussion

Of newly diagnosed oral cancer cases, ~80% present in an early TNM Stages (I & II) [4]. The prognosis is generally fair for these early stage OSCC patients with a five-year survival rate of 53-68%. Thus, 32-47% of patients presenting in an early stage will die of the disease within a 5-year period [4]. Such variation in clinical outcome suggests that the present staging system is not adequate for prognosis. An additional prognostication modality is crucial to further guide the treatment of patients presenting with an early stage oral cancer. We performed a deep sequencing analysis and qRT-PCR validation assay and discovered a two-miRNA panel that is predictive of survival of patients with early TNM stage OSCC. The miR prognostic

panel might be useful in identifying subset of patients who may benefit from additional therapy to surgery alone.

From the next-generation sequencing analysis, we detected 1910 miRs, and found four miRs that were consistently expressed across all samples with a more than a twofold change with $p<0.05$ between the favorable vs poor prognosis groups. In the validation study, three out of four miRs were significantly associated with a risk of poor clinical outcome (miRs-375, 214-3p, and 199a-5p). While the deep sequencing analysis required a pre-amplification step, the qRT-PCR assays do not need amplification, which may explain the inconsistent results for one miR (miR-1269a) in the initial discovery and later validation studies.

A panel of miRs has greater prognostic value over a single miR [30, 36]. The ROC analysis illustrates that the two-miR panel (miRs-375 and 214-3p) has a predictive value with the area under the ROC to be 0.903, which is much higher than the single miR (the area under the ROC of 0.674 for miR-375 and 0.823 for miR-214-3p). The predictive power is further enhanced by adding age and gender to the model, and the final model consisting of the two miR-panel plus age and gender has a predictive power of 0.932, a sensitivity of 87.5%, and a specificity of 87.2% ($p<0.0001$). In this multiple logistic regression analysis step, miR-119a-5p showed an insignificant OR and was eliminated from the final model.

miR-375 functions as a tumor suppressor and targets oncogenes such as AEG-1/MTDH, lactate dehydrogenase B (LDHB) and 3'-phosphoinositide-dependent protein kinase 1 (PDK1) [40]. miR-375 was shown to inhibit cell proliferation and regulate cell survival and its lower expression levels correlated with poor clinical outcome in OSCC as well as in head and neck cancer [40, 41]. miR-214-3p is reported to be a negative controller of fibroblast growth factor receptor 1 (FGFR1), the FGFR tyrosine kinase family that is upregulated in various tumors and implicated in tumor metastasis in prostate cancer, non-small cell lung cancer, colorectal cancer and breast cancer [42-45]. Expression of miR-199a-5p along with miR-199a-3p and miR-214 is regulated by the same transcription activator, Egr1 mRNA [46]. miR-199a targets and increases expression of Brm during carcinogenesis, which in turn negatively regulates Egr1, resulting in decreased expression of miR-199a [46]. Having this double-negative feedback loop in a variety of human cancers partially explains the up and downregulation of miR-199a that has been reported. It may also explain the reason for the non-significant contribution of miR-199a-5p found in our multivariate analysis.

Based on the final ROC model, the prognostic risk scores were calculated. The risk scores take into account the relative contribution of patients' age and gender, and the expression levels of miR-214-3p and miR-375 in their tumor tissue samples. The patients with early stage (stages I & II) OSCC can be assigned a mortality or disease-relapse 'risk score' for stratification into either high or low risk groups. TNM staging is the key in assessing prognosis and establishing a treatment strategy for OSCC patients. However, for patients in the same TNM stage (Stages I or II in our study), a miR-based risk score model allows consideration of the biological characteristics of OSCC and provides a quantifiable risk assessment modality that can readily be applied in the clinical setting. The patients in the high risk group may be directed to more aggressive treatment to improve clinical outcomes.

The miR-based risk-score stratification strategy also permits assessment of the patient's response to different types of treatment. In this study, we first compared the clinical outcome of patients in the high risk group who had neck dissections with that of patients in the low risk group who also had neck dissections. There was no significant improvement of prognosis in terms of time to disease relapse or time to death for the individuals assigned a high risk score after having neck dissection. We then compared the time to disease relapse and time to death of patients with and without neck dissections, within the high risk group. There was improvement in terms of time to relapse, although not significant, but the time to death was similar for the high risk patients with and without neck-dissections. This is consistent with the finding that neck dissection reduces regional recurrence but does not improve overall survival in early stage oral cancer [47]. Similar approaches can be taken in future studies to assess the effectiveness of adjuvant chemo/radiotherapy to improve survival rates

in patients with high-risk scores, thereby leading to a more personalized treatment for patients with oral cancer.

This study is limited by small sample size. Also there are limitations in terms of generalizability as this is a single institutional study. However, our study is unique in that it assesses genome-wide miR expression to identify a panel of miRs that may be utilized as a cancer survival prognostication modality in OSCC. We have plans to perform an internal validation and test the repeatability of the final risk score model in a larger set of samples. We will then test the model in a multicenter setting prior to conducting a large-scale prospective study. If validated, we will have obtained a useful prognostic modality that can be applied in the clinic, which will guide appropriate treatment for OSCC patients.

Acknowledgements

This work was supported by the NIH/NIDCR R03-DE-0022797 (to A. Yoon) and the NIH/NIEHS P30 ES009089 (to R. Santella). We thank Irina Gurvich and Iryna Sirosh for technical assistance.

Disclosure of conflicts of interest

None.

Address correspondence to: Dr. Angela Yoon, Division of Oral and Maxillofacial Pathology, Columbia University Medical Center, 630 West 168th Street, PH15W-1562, New York, NY 10032, USA. Tel: 212-305-7676; Fax: 212-305-5958; E-mail: ajk55@cumc.columbia.edu

References

[1] Barasch A, Safford M, Eisenberg E. Oral cancer and oral effects of anticancer therapy. Mt Sinai J Med 1998; 65: 370-377.
[2] Kademani D. Oral cancer. Mayo Clin Proc 2007; 82: 878-887.
[3] Rodu B, Cole P. Oral cavity and pharynx-throat cancer in the United States, 1973-2003. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2007; 104: 653-658.
[4] Sciubba JJ. Oral cancer: the importance of early diagnosis and treatment. Am J Clin Dermatol 2001; 2: 239-251.
[5] McDowell JD. An overview of epidemiology and common risk factors for oral squamous cell carcinoma. Otolaryngol Clin North Am 2006; 39: 277-294.
[6] Specenier PM, Vermorken JB. Current concepts for the management of head and neck cancer: chemotherapy. Oral Oncol 2009; 45: 409-415.
[7] Brandwein-Gensler M, Teixeira MS, Lewis CM, Lee B, Rolnitzky L, Jille JJ, Genden E, Urken ML, Wang BY. Oral squamous cell carcinoma: histologic risk assessment, but not margin status, is strongly predictive of local disease-free and overall survival. Am J Surg Pathol 2005; 29: 167-178.
[8] Massano J, Regateiro FS, Januario G, Ferreira A. Oral squamous cell carcinoma: review of prognostic and predictive factors. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2006; 102: 67-76.
[9] Bandres E, Bitarte N, Arias F, Agorreta J, Fortes P, Agirre X, Zarate R, Diaz-Gonzalez D, Ramirez N, Sola JJ, Jimenez P, Rodriguez J, Garcia-Foncillas J. microRNA-451 regulates macrophage migration inhibitory factor production and proliferation of gastrointestinal cancer cells. Clin Cancer Res 2009; 15: 2281-2290.
[10] Bhatti I, Lee A, Lund J, Larvin M. Small RNA: a large contributor to carcinogenesis? J Gastrointest Surg 2009; 13: 1379-1388.
[11] Park NJ, Zhou H, Elashoff D, Henson BS, Kastratovic DA, Abemayor E, Wong DT. Salivary microRNA: discovery, characterization and clinical utility for oral cancer detection. Clin Cancer Res 2009; 15: 5473-5477.
[12] Myers J. Potential role of micro-RNAs in head and neck tumorigenesis. Head & Neck 2010; 32: 1099-1011.
[13] Avissar M, Christensen BC, Kelsey KT, Marsit CJ. MicroRNA expression ratio is predictive of head and neck squamous cell carcinoma. Clin Cancer Res 2009; 15: 2850-2855.
[14] Liu X, Yu J, Jiang L, Wang A, Shi F, Ye H, Zhou X. MicroRNA-222 regulates cell invasion by targeting matrix metalloproteinase 1 (MMP1) and manganese superoxide dismutase 2 (SOD2) in tongue squamous cell carcinoma cell lines. Cancer Genomics Proteomics 2009; 6: 131-139.
[15] Zhang X, Yang H, Lee JJ, Kim E, Lippman SM, Khuri FR, Spitz MR, Lotan R, Hong WK, Wu X. MicroRNA-related genetic variations as predictors for risk of second primary tumor and/or recurrence in patients with early-stage head and neck cancer. Carcinogenesis 2010; 31: 2118-2123.
[16] Wu BH, Xion XP, Jia J, Zhan WF. MicroRNAs: new actors in the oral cancer scene. Oral Oncol 2011; 47: 314-319.
[17] Gomes CC, Gomez RS. MicroRNA and oral cancer: future perspectives. Oral Oncol 2008; 44: 910-914.
[18] Mahfouz ME, Rodrigo JP, Takes RP, Elsheikh MN, Rinaldo A, Brakenhoff RH, Ferlito A. Cur-

rent potential and limitations of molecular diagnostic methods in head and neck cancer. Eur Arch Otorhinolaryngol 2010; 267: 851-860.
[19] Wong TS, Liu XB, Wong BY, Ng RW, Yuen AP, Wei WI. Mature miR-184 as potential oncogenic microRNA of squamous cell carcinoma of tongue. Clin Cancer Res 2008; 14: 2588-2592.
[20] Kozaki K, Imoto I, Mogi S, Omura K, Inazawa J. Exploration of tumor suppressive microRNAs silenced by DNA hypermethylation in oral cancer. Cancer Res 2008; 68: 2094-2105.
[21] Li J, Huang H, Sun L, Yang M, Pan C, Chen W, Wu D, Lin Z, Zeng C, Yao Y, Zhang P, Song E. MiR-21 indicates poor prognosis in tongue squamous cell carcinomas as a apoptosis inhibitor. Clin Cancer Res 2009; 15: 3998-4008.
[22] Yu ZW, Zhong LP, Ji T, Zhang P, Chen WT, Zhang CP. MicroRNAs contribute to the chemoresistance of cisplatin in tongue squamous cell carcinoma lines. Oral Oncol 2010; 46: 317-322.
[23] Cervigne NK, Reis PP, Machado J, Sadikovic B, Bradley G, Galloni NN, Pintillie M, Jurisica I, Perez-Ordonez B, Gilbert R, Gullane P, Irish J, Kamel-Reid S. Identification of a microRNA signature associated with progression of leukoplakia to oral carcinoma. Hum Mol Genet 2009; 18: 4818-4829.
[24] Chang KW, Liu CJ, Chu TH, Cheng HW, Hung PS, Hu WY, Lin SC. Association between high miR-211 microRNA expression and the poor prognosis of oral carcinoma. J Dent Res 2008; 87: 1063-1068.
[25] Barker EV, Cervigne NK, Reis PP, Goswami RS, Xu W, Weinreb I, Irish JC, Kamel-Reid S. microRNA evaluation of unknown primary lesions in the head and neck. Molecular Cancer 2009; 8: 127.
[26] Chang SS, Jiang WW, Smith I, Poeta LM, Begum S, Glazer C, Shan S, Westra W, Sidransky D, Califano JA. MicroRNA alterations in head and neck squamous cell carcinoma. Int J Cancer 2008; 123: 2791-2797.
[27] Kimur S, Naganuma S, Susuki D, Hirono Y, Yamaguchi A, Fujieda S, Sano K, Itoh H. Expression of microRNAs in squamous cell carcinoma of human head and neck and the esophagus: miR-205 and miR-21 are specific markers for HNSCC and ESCC. Oncol Rep 2010; 23: 1625-1633.
[28] Tran N, McLean T, Zhang X, Zhao CJ, Thomson JM, O'Brien C, Rose B. MicroRNA expression profiles in head and neck cancer cell lines. Biochem Biophys Res Commun 2007; 358: 12-17.
[29] Ramdas L, Giri U, Ashorn CL, Coombes KR, El-Naggar A, Ang KK, Story MD. MiRNA expression profiles in head and neck squamous cell carcinoma and adjacent normal tissue. Head Neck 2009; 31: 642-654.
[30] Liu X, Jiang L, Wang A, Yu J, Shi F, Zhou X. MicroRNA-138 suppresses invasion and promotes apoptosis in head and neck squamous cell carcinoma cell lines. Cancer Lett 2009; 286: 217-222.
[31] Hui AB, Lenarduzzi M, Krushel T, Waldron L, Pintilie M, Shi W, Perez-Ordonez B, Jurisica I, O'Sullivan B, Waldron J, Gullane P, Cummings B, Liu FF. Comprehensive microRNA profiling for head and neck squamous cell carcinomas. Clin Cancer Res 2010; 16: 1129-1139.
[32] Avissar M, McClean MD, Kelsey KT, Marsit CJ. MicroRNA expression in head and neck cancer associates with alcohol consumption and survival. Carcinogenesis 2009; 30: 2059-2063.
[33] Childs G, Fazzari M, Kung G, Kawachi N, Brandwein-Gensler M, McLemore M, Chen Q, Burk RD, Smith RV, Prystowsky MB, Belbin TJ, Schlecht NF. Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma. Am J Pathol 2009; 174: 736-745.
[34] Wu X, Somlo G, Yu Y, Palomares MR, Li AX, Zhou W, Chow A, Yen Y, Rossi JJ, Gao H, Wang J, Yuan YC, Frankel P, Li S, Ashing-Giwa KT, Sun G, Wang Y, Smith R, Robinson K, Ren X, Wang SE. De novo sequencing of circulating miRNAs identifies novel markers predicting clinical outcome of locally advanced breast cancer. J Translational Med 2012; 10: 42.
[35] Weng L, Wu X, Gao H, Mu B, Li X, Wang JH, Guo C, Jin JM, Chen Z, Covarrubias M, Yuan YC, Weiss LM, Wu H. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffin-embedded tissue specimens. J Pathol 2010; 222: 41-51.
[36] Pantano L, Estivill X, Martí E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res 2010; 38: e34.
[37] McDermott AM, Kerin MJ, Miller N. Identification and validation of miRNAs as endogenous controls for RQ-PCR in blood specimens for breast cancer studies. PLoS One 2013; 8: e83718.
[38] Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative Ct method. Nat Protoc 2008; 3: 1101-1108.
[39] Liu N, Chen NY, Cui RX, Li WF, Li Y, Wei RR, Zhang MY, Sun Y, Huang BJ, Chen M, He QM, Jiang N, Chen L, Cho WC, Yun JP, Zeng J, Liu LZ, Li L, Guo Y, Wang HY, Ma J. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. Lancet Oncol 2012; 13: 633-641.
[40] Chen D, Cabay RJ, Jin Y, Wang A, Lu Y, Shah-Khan M, Zhou X. MicroRNA deregulations in head and neck squamous cell carcinoma. J Oral Maxillofac Res 2013; 4: e2.

[41] Wiklund ED, Gao S, Hulf T, Sibbritt T, Nair S, Costea DE, Villadsen SB, Bakholdt V, Bramsen JB, Sørensen JA, Krogdahl A, Clark SJ, Kjems J. MicroRNA alterations and associated aberrant DNA methylation patterns across multiple sample types in oral squamous cell carcinoma. PLoS One 2011; 6: e27840.

[42] Yang F, Zhang Y, Ressler SJ, Ittmann MM, Ayala GE, Dang TD, Wang F, Rowley DR. FGFR1 is essential for prostate cancer progression and metastasis. Cancer Res 2013; 73: 3716-3724.

[43] Kim HR, Kim DJ, Kang DR, Lee JG, Lim SM, Lee CY, Rha SY, Bae MK, Lee YJ, Kim SH, Ha SJ, Soo RA, Chung KY, Kim JH, Lee JH, Shim HS, Cho BC. Fibroblast growth factor receptor 1 gene amplification is associated with poor survival and cigarette smoking dosage in patients with resected squamous cell lung cancer. J Clin Oncol 2013; 31: 731-737.

[44] Chen DL, Wang ZQ, Zeng ZL, Wu WJ, Zhang DS, Luo HY, Wang F, Qiu MZ, Wang DS, Ren C, Wang FH, Chiao LJ, Pelicano H, Huang P, Li YH, Xu RH. Identification of microRNA-214 as a negative regulator of colorectal cancer liver metastasis by way of regulation of fibroblast growth factor receptor 1 expression. Hepatology 2014; 60: 598-609.

[45] André F, Bachelot T, Campone M, Dalenc F, Perez-Garcia JM, Hurvitz SA, Turner N, Rugo H, Smith JW, Deudon S, Shi M, Zhang Y, Kay A, Porta DG, Yovine A, Baselga J. Targeting FGFR with dovitinib (TKI258): preclinical and clinical data in breast cancer. Clin Cancer Res 2013; 19: 3693-3702.

[46] Sakurai K, Furukawa C, Haraguchi T, Inada K, Shiogama K, Tagawa T, Fujita S, Ueno Y, Ogata A, Ito M, Tsutsumi Y, Iba H. MicrRNAs miR-199a-5p and -3p target the brm subunit of SWI/SNF to generate a double-negative feedback loop ina variety of human cancers. Cancer Res 2011; 71: 1680-1689.

[47] Duvvuri U, Simental AA, D'Angelo G, Johnson JT, Ferris RL, Goodinig W, Myers EN. Elective neck dissection and survival in patients with squamous cell carcinoma of the oral cavity and oropharynx. Laryngoscope 2004; 114: 2228-2234.

Supplemental Table 1. Biologic characteristics of conserved genes targeted by miR-375, miR-214-3p, miR-199a-5p, and miR-1269a

PANTHER classification category	Background frequency[a]	Sample frequency[b]	+/-[c]	*P*-value[d]
biological_process (GO:0008150)	14768	680	+	1.25E-21
cellular process (GO:0009987)	12397	593	+	7.03E-18
regulation of cellular process (GO:0050794)	8209	427	+	4.29E-14
regulation of biological process (GO:0050789)	8671	442	+	2.45E-13
biological regulation (GO:0065007)	9162	457	+	2.23E-12
regulation of cellular metabolic process (GO:0031323)	4817	272	+	7.94E-10
regulation of metabolic process (GO:0019222)	5336	288	+	2.30E-08
regulation of primary metabolic process (GO:0080090)	4817	266	+	2.42E-08
regulation of biosynthetic process (GO:0009889)	3536	209	+	4.73E-08
regulation of cellular biosynthetic process (GO:0031326)	3491	205	+	1.53E-07
regulation of macromolecule biosynthetic process (GO:0010556)	3348	197	+	3.90E-07
cellular macromolecule metabolic process (GO:0044260)	5774	299	+	9.24E-07
regulation of nitrogen compound metabolic process (GO:0051171)	3769	213	+	1.72E-06
regulation of macromolecule metabolic process (GO:0060255)	4561	247	+	1.85E-06
single-organism process (GO:0044699)	11277	509	+	2.11E-06
regulation of RNA biosynthetic process (GO:2001141)	2992	178	+	2.19E-06
regulation of cellular macromolecule biosynthetic process (GO:2000112)	3238	189	+	2.28E-06
regulation of transcription, DNA-templated (GO:0006355)	2955	176	+	2.64E-06
organic substance metabolic process (GO:0071704)	7967	384	+	2.97E-06
regulation of nucleobase-containing compound metabolic process (GO:0019219)	3679	208	+	3.06E-06
macromolecule metabolic process (GO:0043170)	6234	315	+	3.47E-06
regulation of RNA metabolic process (GO:0051252)	3077	181	+	3.55E-06
cellular metabolic process (GO:0044237)	7725	374	+	3.86E-06
primary metabolic process (GO:0044238)	7690	372	+	5.15E-06
single-organism cellular process (GO:0044763)	9883	455	+	7.16E-06
metabolic process (GO:0008152)	8430	399	+	1.09E-05
positive regulation of biological process (GO:0048518)	3945	216	+	1.98E-05
regulation of gene expression (GO:0010468)	3748	205	+	7.00E-05
positive regulation of cellular process (GO:0048522)	3532	195	+	9.62E-05
single organism signaling (GO:0044700)	4267	222	+	8.04E-04
signaling (GO:0023052)	4267	222	+	8.04E-04

[a]The number of genes in the reference list that map to PANTHER classification category. [b]The number of genes targeted by four selected miRs that map to PANRHER classification category. [c]+: over-representation of this category in this experiment compared to the reference list. [d]$p \leq 0.001$ as a cutoff

Supplemental Table 2. Multivariable Cox regression analysis of miRs, age and gender for time to relapse and time to death.

	Time to Relapse		Time to Death	
	HR (95%CL)	*p*-value	HR (95% CL)	*p*-value
miR-214-3p	1.33 (1.18-1.50)	<0.0001	1.44 (1.26-1.65)	< 0.0001
miR-375	1.18 (1.06-1.31)	0.0020	1.20 (1.07-1.33)	0.0010
Older Age	1.03 (1.01-1.06)	0.0149	1.03 (1.01-1.06)	0.0128
Female Gender	1.484 (0.78-2.80)	0.2228	1.769 (0.92-3.37)	0.0838

[a]$p<0.05$ is considered to be significant.

Supplemental Table 3. Median time to relapse and time to death in the patients with high risk scores with and without concurrent neck dissection

	Neck dissection		No neck dissection
Mortality Number of subjects	25		15
Median time to death (mths)	24.0		24.0
p-value	0.3622		
Disease Relapse Number of subjects	24		16
Median time to relapse (mths)	24.0		12.0
p-value		0.0602	

[a]p<0.05 is considered to be significant.