HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2020 January 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2019 January ; 28(1): 127–136. doi: 10.1158/1055-9965.EPI-18-0002.

MicroRNA-related genetic variants associated with survival of head and neck squamous cell carcinoma

Owen M. Wilkins[1], **Alexander J. Titus**[1], **Lucas A. Salas**[1], **Jiang Gui**[2], **Melissa Eliot**[3], **Rondi A. Butler**[3], **Erich M. Sturgis**[4,5], **Guojun Li**[4,5], **Karl T. Kelsey**[3,6], and **Brock C. Christensen**[1,7,8]

[1]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH, 03755, USA

[2]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756

[3]Department of Epidemiology, Brown University, Providence, RI 02912 USA

[4]Department of Head and Neck Surgery, The University of Texas M.D. Anderson Cancer Center, Houston TX 77030

[5]Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston TX 77030

[6]Department of Pathology and Laboratory Medicine, Brown University, Providence, RI 02912 USA

[7]Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756

[8]Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756

Abstract

Head and neck squamous cell carcinoma (HNSCC) is commonly diagnosed at an advanced stage and prognosis for such patients is poor. There remains a gap in our understanding of genetic variants related with HNSCC prognosis. MicroRNA-related single nucleotide polymorphisms (miR-SNPs) are a class of genetic variants with gene regulatory potential. We used a genome-scale approach and independent patient populations in a two-stage approach to test 40,286 common miR-SNPs for association with HNSCC survival in the discovery population (n = 847), and selected the strongest associations for replication in validation phase cases (n = 1236). Further, we leveraged miRNA interaction databases and miRNA expression data from The Cancer Genome Atlas (TCGA), to provide functional insight for the identified and replicated associations. Joint population analyses identified novel miR-SNPs associated with overall survival in oral and laryngeal cancers. rs1816158, located within long-non-coding RNA *MIR100HG*, was associated with overall survival in oral cavity cancer (HR; 1.56, 95% CI; 1.21–2.00). In addition, expression

Correspondence: Brock C. Christensen, Geisel School of Medicine at Dartmouth, 660 Williamson Translation Research Building, Dartmouth Hitchcock Medical Center, Lebanon NH 03756, Brock.C.Christensen@dartmouth.edu.

Conflict of Interests: Jiang Gui serves as a statistical consultant to the Veterans Affairs Medical Center (White River Junction VT 05009, USA) and Celdara Medical (Lebanon NH 03766, USA). Otherwise, the authors declare no conflict of interests.

of *MIR100HG*-embedded microRNA, miR-100, was significantly associated with overall survival in an independent cohort of HNSCC cases (HR; 1.25, 95% CI; 1.06–1.49). A SNP in the 3′UTR of *SH3BP4* (rs56161233), that overlaps predicted miRNA binding sites, and is predicted to disrupt several miRNA-mRNA interactions was associated with overall survival of laryngeal cancer (HR; 2.57, 95% CI; 1.71–3.86). Our findings reveal novel miR-SNPs associated with HNSCC survival, and extend our understanding of how genetic variation contributes to HNSCC prognosis.

Keywords

microRNA-related genetic variation; HEAD; NECK & ORAL CANCERS; EPIDEMIOLOGY; GENETICS OF RISK, OUTCOME AND PREVENTION/Polymorphisms in genes related to cell growth, differentiation, metastatic potential, and apoptosis in cancer; PRECLINICAL PREVENTION STUDIES/Biomarkers and intervention studies

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide (1). Treatment can be effective in early stage disease, but is much less so for advanced disease, and approximately two thirds of patients are diagnosed with advanced stage disease (2). With the exception of recently approved immune checkpoint inhibitors as second-line therapy (3), treatment paradigms have improved little over the last three decades (4), and involve a combination of radiation, chemotherapy, and disfiguring surgery with severe morbidity (5). Alcohol consumption, tobacco use and infection with high risk subtypes of human papilloma virus (HPV) are well established risk factors for HNSCC (6,7). Several genetic risk factors have been identified (8–10), however prognostic factors of HNSCC survival are less well studied. Recognizing this fact, in 2017 the International Head and Neck Consortium initiated a survival working group to address gaps in outcome research for HNSCC. Current models of HNSCC survival are based on histopathologic and staging criteria (11). Alcohol consumption, tobacco use, and HPV infection have also been associated with overall survival of HNSCC (12–14). In addition, genetic predictors of HNSCC survival have been identified, uncovering variants involved in DNA repair (15–17), cell cycle control (18), and angiogenesis (19). Single Nucleotide Polymorphisms (SNPs) in miRNA target sites, miRNA seed sites and miRNA processing proteins represent a class of genetic variation with gene regulatory potential, and have been associated with survival of several cancers (20), including HNSCC (21–24). Such genetic variation can alter miRNA dependent gene regulation resulting in changes in cellular behavior (20). Although predictive utility of miRNA expression for HNSCC survival has been extensively studied (25,26), genome-scale studies of the association between miR-SNPs and HNSCC survival are lacking. To date, associations between miR-SNPs and outcomes in HNSCC have used candidate SNP–approaches (21,22,24,27,28). To address this gap in knowledge, we conducted a genome-scale investigation of the association between miR-SNPs and survival in HNSCC.

Methods

Study participants

DNA was collected from study participants of two population-based case-control studies of HNSCC. The Massachusetts (*n,* 904 cases, *n*, 1051 controls) and M.D. Anderson study populations (*n*, 1338 cases, *n*, 1356 controls) have been described previously (23,29,30). For the Massachusetts study, incident cases of HNSCC were identified at nine medical institutions in the Boston, MA, metropolitan area between 1999–2003 (Phase I) and 2006–2011 (Phase II). An independent study pathologist confirmed pathology report histology. Clinical information was obtained through medical chart review while demographic and exposure data were collected using self-administered questionnaires, reviewed by trained study interviewers. Publically available databases were used to obtain overall survival times for cases in the Massachusetts study. In the M.D. Anderson study, incident cases of HNSCC were recruited at The University of Texas M.D. Anderson Cancer Center between 1996 and 2008. All subjects were treated for curative intent at this institution. Demographic and exposure data was also collected through completion of a self-administered questionnaire. While both categorical and continuous measures of tobacco and alcohol use were collected, continuous measures were not available for most subjects from the M.D. Anderson study. Medical record review was used to obtain overall survival times, defined as the time between first appointment and death from any cause, or last contact date. Date of first appointment was used to define survival times as patients are referred to the cancer center from across the United States after initial diagnosis. Importantly, subjects remain untreated until after their first appointment at the M.D. Anderson Cancer Center. International Classification of Disease, Ninth Revision (ICD-9) codings and pathological analyses were used to assign cases to anatomical subdivisions within the head and neck oral cavity, pharyngeal or laryngeal. These groups were used to conduct analyses stratified by tumor site. HNSCC cases were classified as ICD9 codes 141, 143–146, 148, 149 and 161. Site-specific ICD9 codings used are as follows: oral cavity; ICD9 codes 141.1–141.5, 141.8, 141.9, 143–145.2, 145.5–145.9, 149.8 and 149.9, pharyngeal; ICD-9 codes 141.0, 141.6, 145.3, 145.4, 146, 148, 149.0, 149.1, laryngeal; ICD9 code 161.

Genotyping

Whole blood or buccal cell DNA from discovery phase (Massachusetts study) was extracted using the QIAamp DNA mini kit (Qiagen) and genotyped using the standard Axiom miRNA Target Site Genotyping Array (Affymetrix). This array is specifically designed to interrogate miRNA-related genetic variation, and has previously been described (23). The array contains SNPs and insertions/deletions (indels) in ~238,000 miRNA-related loci including miRNAs, miRNA regulatory regions (such as promoters), miRNA processing proteins, and miRNA target sites. Five online databases containing data regarding miRNA-related genetic variation were used to construct the majority of content on the array; PolymiRTS (31), dPORE (32), Patrocles (33), miRNASNP (34), and microRNA.org (35). We performed quality control (QC) according the Axiom Best Practice Genotyping Analysis Workflow using data from all available cases and controls. Twenty-six samples with either a call rate <97% or a Dish QC (DCQ) metric below a threshold of 0.82 were excluded from the analysis. Minor allele frequencies (MAFs) were calculated using all available discovery phase study subjects.

Analysis was restricted to variants with MAF ≥5%, leaving 40,286 markers. Genotyping of validation phase (M.D. Anderson study) subjects was performed using the MassARRAY iPLEX gold assay (Sequenom) for 114 variants selected from discovery analyses. Validation phase subjects with a call rate of <95% were removed from the analysis. Two variants that were monomorphic in the validation population and six variants with a call rate <95% were removed during QC. An additional six variants deviating from Hardy-Weinberg equilibrium (HWE) with $P<1\times10^{-3}$ in healthy Caucasian control subjects were removed. Two heterozygote genotypes (AG, $n=4$; CG, $n=856$) were detected for rs2450137 in M.D. Anderson study subjects, in addition to CC ($n=113$) and GG ($n=1991$) genotypes. Given the lack of subjects homozygous for rs2450137 with the A allele, subjects with AG genotypes were omitted from the final analysis for this SNP. All genotypes provided are from the + strand of hg19.

Statistical analysis

Multivariable cox proportional hazards regression was used to test the association between miR-SNP genotype and overall survival. R Statistical Software version 3.3.0 was used to calculate hazard ratios (HRs), 95% confidence intervals (CIs), and *P*-values, assuming a dominant mode of inheritance. Regression models for discovery phase analyses were adjusted for potential confounders including age (≤50, >50 to ≤60, >60 to ≤70, >70), sex (male, female), race (Caucasian, other), HPV serology (positive/negative for listed serological subtypes), alcohol consumption (lifetime average number of drinks per week), tobacco use (lifetime pack-years), and tumor stage (low-stage; I & II vs high-stage; III & IV, due to the clinical and prognostic similarity among early and late stage disease). Data from all available study subjects (cases & controls) was used to calculate population quartile distributions for alcohol-consumption (≤1, >1 to ≤6, >6 to ≤31, >31) and tobacco (≤2, >2 to ≤6, >6 to ≤14, >14). Cases were ordered into discrete groups for the analysis, based on these quartile distributions. Complete covariate data were available in 847 subjects from the Massachusetts study (Table 1), and were used in the discovery phase analyses. Evidence of population stratification was assessed through calculation of genomic inflation factor λ using the *GenABEL* R-package. Markers were selected for validation genotyping based on *P*-values from adjusted analyses, and were pruned based on pairwise LD patterns (from all available study subjects) to select the two SNPs in lowest pairwise LD in each gene. In total, 123 associations (35 for overall HNSCC, 35 for oral cavity cancer, 29 for pharyngeal cancer, and 24 for laryngeal cancer) across 114 variants (due to overlap between some variants selected in the overall and site-specific analyses) were selected for replication testing. Of the markers selected for validation genotyping, only one variant, rs1971475 deviated from HWE in control subjects from the discovery population, and was removed from the final joint population analyses. After the QC steps described in the *Genotyping* section, 31, 20, 28 and 19 variants were left to be tested for replication of effect in the overall HNSCC, oral cavity cancer, pharyngeal cancer, and laryngeal cancer analyses, respectively. Analyses of validation phase subjects and the joint population analyses were adjusted for age, sex, race, smoking status (current, former, never), and tumor stage (as above). Complete data on these covariates was available for 1236 cases from the M.D. Anderson population (Table 1), and were used in validation phase analyses. In the joint population analysis, genotyping data across studies were pooled and adjusted for age, sex, race, smoking status (current, former,

never), and tumor stage. 37 cases from the Massachusetts study were missing smoking status (current, former, never) and were excluded from the joint analysis, leaving 2046 subjects for the joint population analysis. To correct for inflated type I error resulting from multiple testing, we determined replication as those variants with $P<1\times10^{-3}$ in the joint analysis, a lower joint analysis *P* than discovery phase *P*, and a consistent predicted direction of effect across both populations.

In silico analyses of miRNA target site disruption

To identify miRNA-mRNA interactions potentially disrupted by genetic variants associated with overall survival in HNSCC, we utilized data from several publicly available databases containing predicted miRNA target sites or predicted effects of genetic variation upon miRNA-mRNA interaction. We have previously described the analytical framework used for this approach (23). Briefly, to identify high confidence predicted miRNA target sites, we intersected all available miRNA target site predictions from the microRNA.org database (August 2010 Release) with UCSC Genome Assembly hg19 coordinates for survival associated SNPs. miRNA target site predictions in the microRNA.org database were made using the miRanda algorithm and scored for meaningful downregulation of mRNA targets by the mirSVR algorithm (more negative scores suggest greater potential for meaningful downregulation of its target). High confidence predictions were considered those with a mirSVR score of < -0.1. Percentiles for mirSVR scores were calculated to help interpret the potential impact of a predicted miRNA target site. It should be noted that since more negative mirSVR scores are suggestive of greater potential for downregulation of a target mRNA, those with the lowest percentile ranks (i.e. 1st, 2nd, 3rd etc.) represent those predictions with the most likely functional impact. To identify miRNA-mRNA interactions potentially disrupted by survival associated genetic variants, we mined data from three online databases; PolymiRTSv3.0 (31), miRNASNP (34), and MirSNP (36). PolymiRTS (http://compbio.uthsc.edu/miRSNP/) uses context+ scores (generated by the TargetScan algorithm) to rank the predicted effects of genetic variation upon miRNA target site binding. TargetScan context+ scores predict binding of a miRNA over a 3′UTR. Greater likelihood of miRNA-mRNA disruption for a given variant is indicated by more negative context+ scores. The difference between context+ scores predicted for sequences containing each allele of a genetic variant provides a measure of disruption to predicted miRNA binding. We calculated percentile ranks from all available context+ score differences available from the PolymiRTS database. Lower percentile ranks indicate miRNA-mRNA interactions that are more likely to be disrupted by genetic variation. miRNASNP (http://www.bioguo.org/miRNASNP/index.php) utilizes TargetScan and miRanda algorithms to identify genetic variants likely to impact miRNA-mRNA interaction. MirSNP (http://bioinfo.bjmu.edu.cn/mirsnp/search/) determines predicted miRNA target sites for sequences containing each allele of a given genetic variant using the miRanda algorithm, and assigns a predicted impact of genetic variation (enhance, decrease, create or break) at this locus based on the results.

TCGA miRNA & mRNA expression analyses

Clinical/demographic data and miRNA-seq data for subjects from The Cancer Genome Atlas (TCGA) HNSCC project were downloaded from at the Genomic Data Commons Data Portal (accessed 05-29-17) (37,38). Excluding subjects whose samples were from recurrent

tumors or received neoadjuvant treatment, 523 primary HNSCCs (oral cavity cancer; n = 314, pharyngeal cancer; n = 86, laryngeal cancer; n = 119, other; n = 4) and 44 adjacent normal tissue samples were available for analysis. ICD-10 codes available in the TCGA data set were converted to ICD-9 codings to assign TCGA subjects to oral cavity, pharyngeal, and laryngeal cancer groups, for concordance with the subsite classification used in our study, outlined in the *Study participants* section. 4 subjects were removed from the TCGA data set as they did not align with the subgroup classification criteria of our study. Conversions for ICD-10 to ICD-9 codes are summarized in Supplementary Table 1. To assess expression of miRNAs predicted to target regions containing miR-SNPs, average mapped reads/million from isoform-specific miRNA-seq data were calculated across all available subjects for either tumor or normal tissue samples, $\log_2$ transformed, and converted to percentile ranks. For survival analyses, $\log_2$ transformed FPKM (Fragments Per Kilobase Million) expression values were calculated using the overall miRNA expression (not isoform-specific) and RNA-seq data. 303 oral cavity cancer cases with available overall miRNA expression and complete covariate data for the variables of interest were used to test the association between survival time and *MIR100HG*-derived microRNAs. For mRNA expression analyses, RNA-seq data were available for 478 subjects with complete covariate data. *BLID* expression was undetectable in the majority of subjects with tumors of the oral cavity. Measurable *BLID* expression and complete covariate data was only available in 37 (12%) of oral cavity cancer subjects, therefore the association between *BLID* expression and overall survival was not tested. *SH3BP4* expression and complete covariate data, was available for 108 laryngeal tumors. Cox proportional hazards regression models adjusting for age at diagnosis, sex, race (white vs non-white), and tumor stage (low grade; I+II vs high-grade; III+IV). Gene expression was modelled as a continuous variable.

Code and Data availability

R Code used to perform the *in silico* and TCGA analyses has been deposited in the "HNSCC-miRSNP-Survival" repository on GitHub (https://github.com/Christensen-Lab-Dartmouth). Database of Genotypes and Phenotypes (dbGAP) accession numbers for the deposited data will be provided as they become available, and will accompany the final version of this manuscript.

Results

Discovery population associations of miR-SNPs with HNSCC survival

To identify miR-SNPs associated with HNSCC-specific survival, HNSCC cases from a population-based case-control study of HNSCC were genotyped using the Affymetrix Axiom miRNA Target Site Array, which interrogates ~238,000 genetic variants in miRNA target sites, miRNA genes, and genes in the miRNA biogenesis pathway. 40,286 genetic variants with a minor allele frequency ≥5% (in available cases and controls) were tested for association with overall survival in all HNSCC cases, or tumor site-specific (oral cavity, pharyngeal, laryngeal) analyses. Demographic and clinical characteristics of the 847 cases from the discovery population are in Table 1. We applied Cox proportional hazards regression adjusted for age, sex, race, HPV seropositivity, tobacco consumption, alcohol consumption and tumor stage. The strongest associations were observed for rs4880198

(overall HNSCC, *UAP1L1*, HR; 1.54, 95% CI; 1.24–1.92, Supplementary Figure 1A), rs2292842 (oral cavity cancer, *PRDM8*, HR; 3.40, 95% CI; 2.06–5.60, Supplementary Figure 1B), rs72765149 (pharyngeal cancer, *ZBED3-AS1*, HR; 2.90, 95% CI; 1.79–4.71, Supplementary Figure 1C), rs1568391 (laryngeal cancer, *IRF8*, HR; 0.25, 95% CI; 0.13–0.47 Supplementary Figure 1D). Generally, the strongest associations were observed in the laryngeal cancer analysis, with 74 variants observed with an adjusted P<1.0E-03 (Supplementary Figure 1D). In the analyses of overall HNSCC, oral cavity cancer, and pharyngeal cancer, 39, 57 and 45 variants respectively, had adjusted P<1.0E-03 (Supplementary Figures 1A–C). We observed minimal evidence of population stratification (Supplementary Figures 2A–D). Association strength by genomic location for all variants is show in Supplementary Figure 1, where variants selected for genotyping in the validation phase are highlighted in green.

Validation phase associations of miR-SNPs with HNSCC risk

To validate our findings, we genotyped selected variants in an independent study population of incident HNSCC from The University of Texas M.D. Anderson Cancer Center. Pairwise linkage patterns between the strongest associated variants in the discovery phase was used to prune and select variants for validation phase genotyping. After quality control, 31, 30, 28, and 19 variants were tested for replication of effect with overall HNSCC, oral cavity cancer, pharyngeal cancer and laryngeal cancer respectively. By pooling genotyping data across studies, we performed a joint analysis adjusting for age, sex, race, smoking status, and tumor stage, with a total of 2046 cases (with complete covariate data) of HNSCC. In site-specific analyses 648, 1081, and, 317 cases were included in the oral cavity, pharyngeal, and laryngeal analyses, respectively. Summary statistics for all variants tested for replication are presented in Supplementary Table 2. Two of these variants - rs1816158 (oral cavity cancer survival) and rs56161233 (laryngeal cancer survival) – had P<1.0E-03 in the joint analysis, a lower joint analysis P than discovery phase P, and a consistent direction of effect (Table 2). Carriers of the minor allele for rs1816158 and rs56161233 had worse probability of survival than those homozygous for the major allele (Figure 1A & Figure 1B). rs1816158 is within intron 1 of the long non-coding RNA (lncRNA) *MIR100HG* was significantly associated with oral cancer survival (joint analysis oral cavity cancer HR; 1.56, 95% CI; 1.21 – 2.00, Table 2). In addition, rs1816158 overlaps with promoter and enhancer histone marks, as well as DNase hypersensitivity sites, in several tissues (Supplementary Table 3). No nearby genotyped variants were in strong linkage with rs1816158 (Supplementary Figure 3A). rs56161233 is located within the 3′UTR of *SH3BP4* and was validated for its association with laryngeal cancer survival (joint analysis HR; 2.87, 95% CI; 1.73 – 4.75, Table 2). rs4589703 was the only genotyped variant in strong linkage (r^2 > 0.8) with rs56161233, and had a modest association with laryngeal cancer survival in the discovery phase (Supplementary Figure 3B) which was dependent upon rs56161233 in conditional analyses (Supplementary Figure 4). Neither rs56161233 or rs56161233 showed evidence of an association in the analysis of overall HNSCC (Supplementary Table 4), suggesting their effects are site-specific. While no variants from the overall HNSCC and pharyngeal cancer analyses met our replication criteria, several variants approached replication (Table 2), including *RPL28* variants rs56312243 and rs45494801, both of which have been identified as eQTL loci for *RPL28* in multiple tissues(Supplementary Table 5). Using the schoenfield

residual test to assess the proportional hazards assumption, we observed no evidence for violation of this assumption in these analyses. Generally, similar results were obtained in a sensitivity analysis restricted to Caucasian subjects (Supplementary Table 6). Additionally, we conducted a sensitivity analysis limiting follow up to five-year survival and observed consistent effect sizes (Supplementary Table 7).

Survival analysis of gene expression levels from HNSCC-associated variants

As genetic variation at HNSCC-associated miR-SNPs may result in gene expression changes that modify cellular behavior and ultimately survival times, we tested the association between expression of genes containing miR-SNPs that met our replication criteria (rs1816158, *MIR100HG*; rs56161233, *SH3BP4*), and overall survival, in HNSCC subjects from the TCGA HNSCC project (Table 3). To determine the potential effects of *MIR100HG*-derived microRNAs on overall survival in oral cavity cancer, we utilized miRNA-seq data from available oral cavity cancer cases (n = 303). Expression of miR-100 was significantly associated with oral cancer survival (HR; 1.25, 95% CI; 1.06 – 1.49), in multivariable Cox proportional hazards regression models adjusting for age at diagnosis, sex, race, and tumor stage (Table 3). These findings suggest increased miR-100 expression may alter cellular behavior in such a way that modifies survival times. We did not observe a significant association between *SH3BP4* expression and overall survival in laryngeal cancer cases (n = 108, HR; 1.13, 95% CI; 0.65 – 1.97, Table 3).

HNSCC survival associated miR-SNPs effect miRNA-mRNA interactions

Genetic variation within miRNA target sites can create, destroy or alter the binding affinity for complimentary miRNAs. To identify miRNA-mRNA interactions potentially disrupted by genetic variation we intersected predicted miRNA target sites made using the miRanda algorithm with genomic coordinates of SNPs outlined in Table 2. Primary tumor miRNA expression data from the TCGA HNSCC project was also used to assess the potential functionality of implicated miRNAs. We identified multiple high confidence miRNA-mRNA interactions overlapping each of rs16988668 (*ZSCAN22*), rs3831960 (*ZSWIM5*), rs56312243 (*RPL28*), rs77506493 (*ZNF766*), and rs56161233 (*SH3BP4*) (Table 4). Several of the identified miRNAs are highly expressed in normal or tumor tissue from HNSCC subjects (Table 4). The highest confidence miRNA target site observed in this analysis was the interaction between miR-2110 and *SH3BP4* overlapping laryngeal cancer associated miR-SNP rs56161233, which was predicted with a mirSVR score in amongst the top 1% of all mirSVR predictions genome-wide. In addition, miR-548m, miR-593-3p, and miR-3150a-3p are also predicted to target *SH3BP4* within the region of the 3′UTR containing rs56161233 (all among the top quartile of mirSVR predictions). Multiple miRNA target sites overlapped each of miR-SNPs, rs3831960 and rs16988668, that approached replication in the association analysis of overall HNSCC survival (Table 4). miR-96-3p was predicted to target *ZSWIM5* with notably high confidence (mirSVR score 6th percentile), while *ZSCAN22* was predicted to be targeted by miR-339-5p which is abundantly expressed in both normal and tumor tissue (90th and 92nd percentiles, respectively) from HNSCC subjects. Predicted miRNA target sites were also identified within *RPL28* and *ZNF766* that overlap rs56312243 and rs77506493, respectively.

We next sought to identify miRNA target sites whose binding is disrupted by genetic variation at these loci. We mined three publically available databases (PolymiRTSv3.0, miRNASNP and MirSNP) that contain predicted effects of genetic variation at miRNA-related loci. As above, for each miRNA-mRNA relationship identified, miRNA expression was calculated using normal and tumor tissue samples from TCGA HNSCC subjects. Details regarding the predicted effects of genetic variation upon the miRNA-mRNA interactions from each database are provided in Supplementary Tables 8–10. TargetScan context+ score differences between each allele of a specified variant are provided in the PolymiRTS database, and provide a likelihood measure that the variant will affect miRNA binding. miRNA-mRNA interactions with context+ score differences in at least the top 20% of all available scores were identified for rs4127682 (*RGS3*), rs16988668 (*ZSCAN22*), rs56312243 (*RPL28*), rs77506493 (*ZNF766*), and rs56161233 (*SH3BP4)* (Supplementary Table 8). Several miRNAs whose binding is predicted to be disrupted by genetic variation at these loci are expressed at high levels in normal and tumor tissue from TCGA HNSCC subjects. Contrastingly, predictions stored in the MIRNASNP database are made using the TargetScan and miRanda prediction algorithms simultaneously to identify miRNA target sites created or destroyed by genetic variation. From the miRNASNP database, we identified target sites created or destroyed by genetic variation at rs4127682 (*RGS3*), rs16988668 (*ZSCAN22*), rs75820821 (*CCDC97*), rs56312243 (*RPL28*), and rs56161233 (*SH3BP4)* (Supplementary Table 9). Notably, six miRNA target sites are predicted to be created by presence of the T allele for rs56312243 (*RPL28*), while six target sites are predicted to be destroyed by the T allele. miR-486-3p, one of the target sites predicted to be created by presence of the T allele of rs56312243, was expressed in the 83rd and 73rd percentiles of normal and tumor tissue samples, respectively. Finally, for predictions in the MirSNP database, the miRanda algorithm was used to identify miRNA binding sites created, destroyed, enhanced or decreased by genetic variation. Predicted effects for several of the target sites predicted to be created or destroyed by rs4127682 (*RGS3*), rs16988668 (*ZSCAN22*), rs75820821 (*CCDC97*), rs56312243 (*RPL28*), rs45494801 (*RPL28*), and rs56161233 (*SH3BP4)* were similar between the MIRNASNP and MirSNP databases (Supplementary Tables 9 & 10). Presence of the G allele for rs4127682 was predicted to enhance and decrease binding of miR-1909-3p and miR-4763-3p to *RGS3*, respectively, across five *RGS3* transcript variants (Supplementary Table 10). Many of these miRNAs were expressed at high levels in normal and tumor tissue from TCGA HNSCC subjects. miR-2355-5p, predicted to be disrupted by genetic variation of rs45494801, was among the most abundantly expressed miRNAs, in the 89th percentile in normal tissue, and 93rd percentile in tumor tissue. These results highlight miRNA-mRNA interactions with potential to be disrupted by the miR-SNPs tested for association with overall survival of HNSCC.

Discussion

Studies of the association between germline genetic variation and clinical outcomes in cancer can reveal insights into the complex regulatory networks related with disease progression, as well as identify biomarkers that may possess clinical utility. Previous studies of miR-SNPs and overall survival in HNSCC (21,22,24), have used candidate-based approaches. We present a genome-scale evaluation of the contribution of common miR-

SNPs to survival in overall HNSCC, and site-specific disease, validating associations of *MIR100HG* variant rs1816158 with overall survival in oral cavity cancer, and *SH3BP4* 3′UTR variant rs56161233 with overall survival of laryngeal cancer. Moreover, providing mechanistic insight into the association of rs1816158 with oral cavity cancer, we demonstrated that expression of *MIR100HG* derived miR-100 is associated with head and neck cancer survival in TCGA tumors. We also identify multiple high confidence miRNA target sites predicted to bind the 3′UTR of *SH3BP4* in the region containing rs56161233. Additionally, we describe miRNA target sites located in the 3′UTR of *SH3BP4* that are likely to be disrupted by genetic variation at rs56161233. Finally, we identify several additional miR-SNPs with putative survival associations for head and neck cancer where there is also high confidence for miRNA target sites and disruption of miRNA-mRNA interaction introduced by the miR-SNP: oral cavity cancer, rs3831960 (*ZSWIM5*, 3′UTR), rs16988668 (*ZSCAN22*, 3′UTR); laryngeal cancer, rs3777710 (*FAM8A1*, 3′UTR), rs12280753 (*BUD13*, 3′UTR). Together, these findings demonstrate miR-SNPs are associated with overall survival in HNSCC, and provide evidence that genetic variation at these loci result in functional changes to gene regulation.

MIR100HG is a long non-coding RNA from which 3 miRNAs (miR-100, let-7A2, miR-125B1) and one coding gene (*BLID*) are derived. Multiple studies suggest that *MIR100HG* and its derived genes contribute to various cancers. *MIR100HG* has been identified as an oncogene in acute megakaryoblastic leukemia (39) and implicated in regulation of lymph-node metastases in early stage cervical cancer (40). mir-100 induces epithelial-mesenchymal transition in mammary epithelial cells, while suppressing tumorigenesis, migration and invasion (41). Further evidence of potential tumor-suppressive effects exists in the context of adult stem cells (42), though the effects of *MIR100HG* and its derived genes in cancer are likely highly complex and tissue dependent, warranting further study. Most recently, miR-100 and miR-125b were found to be overexpressed in HNSCC cell lines, and to mediate resistance to cetuximab (43). Given these data, we would expect the C allele of rs1816158, associated with increased risk of death in our study (Table 2), to increase miR-100 expression, however this will require validation in future studies. Let-7A-2, one of three isoforms that produce mature let-7a, is a member of the highly conserved let-7 microRNA family. Let-7a expression is downregulated in several cancers and negatively regulates oncogenes *RAS* (44), *HMGA2*, (45) and *MYC* (46). Expression of let-7A and its family members are tumor suppressive across a range of cancer types (47). Although let-7a-2 expression was not associated with overall survival of HNSCC in this study, given the known functions of let-7a, let-7A-2 cannot be discounted as a potential mediator of the observed association between rs1816158 overall survival in HNSCC. Overall, our findings add to the growing body of data supporting a complex role for *MIR100HG* in tumor biology.

SH3BP4 was among the first proteins identified to mediate cargo-specific control of clathrin-mediated endocytosis (48). Further studies have indicated that *SH3BP4* negatively regulates amino acid-Rag GTPase-mTORC1 signaling (49). As hyperactivation of *mTORC1* has been observed in several cancer types (50), it has been postulated *SH3BP4* exerts tumor suppressive effects over *mTORC1* signaling (49). *SH3BP4* has also been identified as a required component of a molecular complex including *FGFR2b* and *PI3K* that controls

receptor recycling and cellular outcomes in response to FGF10 stimulation (51). The diverse functions of *SH3BP4* provide a rationale for how microRNA-related genetic variation in *SH3BP4* may modify survival in HNSCC. Although we did not observe an association between laryngeal cancer survival and *SH3BP4* expression, this does not preclude the possibility that rs56161233 affects survival through modification of microRNA-mediated gene regulation. MicroRNA binding can result in decreased mRNA translation rather than promoting enzymatic cleavage of the target (20). Future studies involving proteomic data collection and analysis will be required to delineate such instances. Furthermore, in recent years development of the competing endogenous RNA (ceRNA) hypothesis has posited that competition for miRNA binding between transcripts is a substantial determinant of gene regulation (52). Disruption of microRNA binding through genetic variation within a microRNA target site may therefore alter regulation of a gene that shares this target site, while potentially leaving expression of the gene containing the variant mostly unchanged. Such phenomena have not been largely explored and deserve study.

Recent technological advances have facilitated the identification of predicted and experimentally validated miR target sites throughout the genome. Here, we leverage a novel genotyping platform to assay miR-related genetic variation at the genome-scale level. Our study represents a substantial improvement upon the existing candidate-based approaches. Furthermore, our study leverages detailed follow up information on subject specific survival to identify miR-SNPs associated with overall survival of HNSCC in two population-based studies. However, our study is limited by modest somewhat sample size for the detection of germline genetic variants associated with outcomes of a complex trait such as overall survival of HNSCC. Collection of detailed follow up data in additional population-based studies of HNSCC are needed to increase statistical power in future work, particularly for tumor-site-specific analyses. Additionally, detailed exposure information regarding HPV infection and alcohol consumption was not available in the validation population, preventing adjustment for these factors in joint population analysis. HPV infection is an important predictor of overall survival in oropharyngeal cancers and alcohol consumption is a prognostic factor for overall survival in HNSCC (53). Lack of available data likely contributed limited replication of miR-SNPs with putative survival associations observed in the discovery population. Furthermore, use of HPV infection as a prognostic factor in HNSCC relates to tumor cell HPV infection, therefore potential discordance between the HPV seropositivity detection methods used here and tumor HPV status could introduce bias (54). Although presence of HPV-specific antibodies has been associated with HPV infection in tumor cells and improved prognosis in multiple studies (55,56), future studies should take this into consideration in both study design and interpretation of results. Cross-study differences in variable structure may have also contributed to lack of replication. Additionally, continuous tobacco exposure was available for discovery phase subjects, though only categorical exposure data (current, former, never smokers) was available in validation phase subjects. Our results highlight the importance of adjustment for clinical and demographic factors that strongly predict survival in genetic association studies. Although treatment data were also not considered in these analyses, future studies capable of collecting such data should aim to evaluate the potential effects of different treatment regimens on the hazards conferred by miR-SNPs. Finally, with continued development of

miRNA-prediction algorithms in recent years, and generation of new data sets of experimentally validated miRNA target sites, more up-to-date lists of miRNA target sites may be available for future studies aiming to evaluate genome-wide miRNA-related genetic variation. Detailed collection of specific exposures in future studies will improve statistical power for the detection of genetic variants associated with survival in diseases with major environmental contributors, such as HNSCC. Overall, our genome-scale analysis of miR-related genetic variation with survival in HNSCC, highlights the power of using functional annotation to guide genetic association studies, as well as the need for more powerful evaluation of the association between miR-SNPs and survival in HNSCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: This work was supported by National Institutes of Health grants R01DE022772 and R01CA216265 to B.C. Christensen, R01CA078609 to K.T. Kelsey. We acknowledge funding contributions from The University of Texas MD Anderson Christopher and Susan Damico Chair in Viral Associated Malignancies, National Institute of Environmental Health Sciences grant R01ES11740 and R01CA131274 (to Q. Wei); and NIH grant P30CA016672 (to The University of Texas MD Anderson Cancer Center). A. Titus. was supported in part by T32LM012204. We would like to thank Dr. Qingyi Wei for agreeing to collaborate on R01DE022772 before he moved on from The University of Texas M.D. Anderson Cancer Center.

We are grateful to Kevin C. Johnson for discussions that improved this manuscript.

Abbreviations

HNSCC	Head and neck squamous cell carcinoma
HWE	Hardy-Weinberg equilibriu
HPV	Human papillomavirus
miR-SNP	miRNA-related single nucleotide polymorphism
MAF	Minor allele frequency

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015; 136:E359–86. [PubMed: 25220842]
2. Worsham MJ. Identifying the risk factors for late-stage head and neck cancer. Expert Rev Anticanc. 2011; 11:1321–5.
3. Cho J, Johnson DE, Grandis JR. Therapeutic Implications of the Genetic Landscape of Head and Neck Cancer. Semin Radiat Oncol. 2018; 28:2–11. [PubMed: 29173752]
4. Chin D, Boyle GM, Porceddu S, Theile DR, Parsons PG, Coman WB. Head and neck cancer: past, present and future. Expert Rev Anticanc. 2006; 6:1111–8.
5. Price KA, Cohen EE. Current treatment options for metastatic head and neck cancer. Curr Treat Options Oncol. 2012; 13:35–46. [PubMed: 22252884]

6. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, et al. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res. 1988; 48:3282–7. [PubMed: 3365707]
7. Furniss CS, McClean MD, Smith JF, Bryan J, Nelson HH, Peters ES, et al. Human papillomavirus 16 and head and neck squamous cell carcinoma. Int J Cancer. 2007; 120:2386–92. [PubMed: 17315185]
8. McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang SC, Byrnes G, et al. A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. Plos Genet. 2011:7.
9. Lesseur C, Diergaarde B, Olshan AF, Wunsch-Filho V, Ness AR, Liu G, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nat Genet. 2016; 48:1544–50. [PubMed: 27749845]
10. Michmerhuizen NL, Birkeland AC, Bradford CR, Brenner JC. Genetic determinants in head and neck squamous cell carcinoma and their influence on global personalized medicine. Genes Cancer. 2016; 7:182–200. [PubMed: 27551333]
11. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. Lancet. 2008; 371:1695–709. [PubMed: 18486742]
12. Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst. 2008; 100:407–20. [PubMed: 18334711]
13. Farshadpour F, Kranenborg H, Calkoen EV, Hordijk GJ, Koole R, Slootweg PJ, et al. Survival Analysis of Head and Neck Squamous Cell Carcinoma: Influence of Smoking and Drinking. Head Neck-J Sci Spec. 2011; 33:817–23.
14. Mayne ST, Cartmel B, Kirsh V, Goodwin WJ. Alcohol and Tobacco Use Prediagnosis and Postdiagnosis, and Survival in a Cohort of Patients with Early Stage Cancers of the Oral Cavity, Pharynx, and Larynx. Cancer Epidem Biomar. 2009; 18:3368–74.
15. Geisler SA, Olshan AF, Cai J, Weissler M, Smith J, Bell D. Glutathione S-transferase polymorphisms and survival from head and neck cancer. Head Neck. 2005; 27:232–42. [PubMed: 15668931]
16. Quintela-Fandino M, Hitt R, Medina PP, Gamarra S, Manso L, Cortes-Funes H, et al. DNA-repair gene polymorphisms predict favorable clinical outcome among patients with advanced squamous cell carcinoma of the head and neck treated with cisplatin-based induction chemotherapy. J Clin Oncol. 2006; 24:4333–9. [PubMed: 16896002]
17. Nogueira GAS, Lourenco GJ, Oliveira CBM, Marson FAL, Lopes-Aguiar L, Costa EFD, et al. Association between genetic polymorphisms in DNA mismatch repair-related genes with risk and prognosis of head and neck squamous cell carcinoma. International Journal of Cancer. 2015; 137:810–8. [PubMed: 25598504]
18. Marsit CJ, Black CC, Posner MR, Kelsey KT. A genotype-phenotype examination of cyclin D1 on risk and outcome of squamous cell carcinoma of the head and neck. Clin Cancer Res. 2008; 14:2371–7. [PubMed: 18413827]
19. da Costa Andrade VC, Parise O Jr, Hors CP, de Melo Martins PC, Silva AP, Garicochea B. The fibroblast growth factor receptor 4 (FGFR4) Arg388 allele correlates with survival in head and neck squamous cell carcinoma. Exp Mol Pathol. 2007; 82:53–7. [PubMed: 17084840]
20. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research (vol 10, pg 389, 2010). Nature Reviews Cancer. 2010; 10
21. Christensen BC, Avissar-Whiting M, Ouellet LG, Butler RA, Nelson HH, McClean MD, et al. Mature MicroRNA Sequence Polymorphism in MIR196A2 Is Associated with Risk and Prognosis of Head and Neck Cancer. Clinical Cancer Research. 2010; 16:3713–20. [PubMed: 20501619]
22. Christensen BC, Moyer BJ, Avissar M, Ouellet LG, Plaza SL, McClean MD, et al. A let-7 microRNA-binding site polymorphism in the KRAS 3′ UTR is associated with reduced survival in oral cancers. Carcinogenesis. 2009; 30:1003–7. [PubMed: 19380522]
23. Wilkins OM, Titus AJ, Gui J, Eliot M, Butler RA, Sturgis EM, et al. Genome-scale identification of microRNA-related SNPs associated with risk of head and neck squamous cell carcinoma. Carcinogenesis. 2017; 38:986–93. [PubMed: 28582492]

24. Guan X, Sturgis EM, Song X, Liu Z, El-Naggar AK, Wei Q, et al. Pre-microRNA variants predict HPV16-positive tumors and survival in patients with squamous cell carcinoma of the oropharynx. Cancer Lett. 2013; 330:233–40. [PubMed: 23219900]
25. Jamali Z, Aminabadi NA, Attaran R, Pournagiazar F, Oskouei SG, Ahmadpour F. MicroRNAs as prognostic molecular signatures in human head and neck squamous cell carcinoma: A systematic review and meta-analysis. Oral Oncol. 2015; 51:321–31. [PubMed: 25677760]
26. Avissar M, Christensen BC, Kelsey KT, Marsit CJ. MicroRNA Expression Ratio Is Predictive of Head and Neck Squamous Cell Carcinoma. Clinical Cancer Research. 2009; 15:2850–5. [PubMed: 19351747]
27. Zhu LJ, Sturgis EM, Lu ZM, Zhang H, Wei P, Wei QY, et al. Association between miRNA-binding site polymorphisms in double-strand break repair genes and risk of recurrence in patients with squamous cell carcinomas of the non-oropharynx. Carcinogenesis. 2017; 38:432–8. [PubMed: 28334093]
28. Zhu LJ, Sturgis EM, Zhang H, Lu ZM, Tao Y, Wei QY, et al. Genetic variants in microRNA-binding sites of DNA repair genes as predictors of recurrence in patients with squamous cell carcinoma of the oropharynx. Int J Cancer. 2017; 141:1355–64. [PubMed: 28646528]
29. Peters ES, McClean MD, Marsit CJ, Luckett B, Kelsey KT. Glutathione S-transferase polymorphisms and the synergy of alcohol and tobacco in oral, pharyngeal, and laryngeal carcinoma. Cancer Epidemiology Biomarkers & Prevention. 2006; 15:2196–202.
30. Li G, Sturgis EM, Wang LE, Chamberlain RM, Amos CI, Spitz MR, et al. Association of a p73 exon 2 G4C14-to-A4T14 polymorphism with risk of squamous cell carcinoma of the head and neck. Carcinogenesis. 2004; 25:1911–6. [PubMed: 15180941]
31. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3. 0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Research. 2014; 42:D86–D91. [PubMed: 24163105]
32. Schmeier S, Schaefer U, MacPherson CR, Bajic VB. dPORE-miRNA: polymorphic regulation of microRNA genes. PLoS One. 2011; 6:e16657. [PubMed: 21326606]
33. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res. 2010; 38:D640–51. [PubMed: 19906729]
34. Gong J, Liu C, Liu W, Wu Y, Ma Z, Chen H, et al. An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database (Oxford). 2015; 2015:bav029. [PubMed: 25877638]
35. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008; 36:D149–53. [PubMed: 18158296]
36. Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, et al. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics. 2012; 13:661. [PubMed: 23173617]
37. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015; 517:576–82. [PubMed: 25631445]
38. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016; 375:1109–12. [PubMed: 27653561]
39. Emmrich S, Streltsov A, Schmidt F, Thangapandi VR, Reinhardt D, Klusmann JH. LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. Mol Cancer. 2014:13. [PubMed: 24461128]
40. Shang C, Zhu W, Liu T, Wang W, Huang G, Huang J, et al. Characterization of long non-coding RNA expression profiles in lymph node metastasis of early-stage cervical cancer. Oncol Rep. 2016; 35:3185–97. [PubMed: 27035672]
41. Chen DH, Sun YT, Yuan Y, Han ZB, Zhang PJ, Zhang JS, et al. miR-100 Induces Epithelial-Mesenchymal Transition but Suppresses Tumorigenesis, Migration and Invasion. Plos Genet. 2014:10.
42. Lopez MF, Niu P, Wang L, Vogelsang M, Gaur M, Krastins B, et al. Opposing activities of oncogenic MIR17HG and tumor suppressive MIR100HG clusters and their gene targets regulate

replicative senescence in human adult stem cells. Npj Aging Mech Dis. 2017:3. [PubMed: 28649421]

43. Lu Y, Zhao X, Liu Q, Li C, Graves-Deal R, Cao Z, et al. lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/beta-catenin signaling. Nat Med. 2017
44. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 MicroRNA family. Cell. 2005; 120:635–47. [PubMed: 15766527]
45. Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev. 2007; 21:1025–30. [PubMed: 17437991]
46. Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res. 2007; 67:9762–70. [PubMed: 17942906]
47. Boyerinas B, Park SM, Hau A, Murmann AE, Peter ME. The role of let-7 in cell differentiation and cancer. Endocr Relat Cancer. 2010; 17:F19–36. [PubMed: 19779035]
48. Tosoni D, Puri C, Confalonieri S, Salcini AE, De Camilli P, Tacchetti C, et al. TTP specifically regulates the internalization of the transferrin receptor. Cell. 2005; 123:875–88. [PubMed: 16325581]
49. Kim YM, Stone M, Hwang TH, Kim YG, Dunlevy JR, Griffin TJ, et al. SH3BP4 Is a Negative Regulator of Amino Acid-Rag GTPase-mTORC1 Signaling. Mol Cell. 2012; 46:833–46. [PubMed: 22575674]
50. Zoncu R, Efeyan A, Sabatini DM. mTOR: from growth signal integration to cancer, diabetes and ageing. Nat Rev Mol Cell Bio. 2011; 12:21–35. [PubMed: 21157483]
51. Francavilla C, Rigbolt KTG, Emdal KB, Carraro G, Vernet E, Bekker-Jensen DB, et al. Functional Proteomics Defines the Molecular Switch Underlying FGF Receptor Trafficking and Cellular Outputs. Mol Cell. 2013; 51:707–22. [PubMed: 24011590]
52. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011; 146:353–8. [PubMed: 21802130]
53. Sawabe M, Ito H, Oze I, Hosono S, Kawakita D, Tanaka H, et al. Heterogeneous impact of alcohol consumption according to treatment method on survival in head and neck cancer: A prospective study. Cancer Sci. 2017; 108:91–100. [PubMed: 27801961]
54. Vokes EE, Agrawal N, Seiwert TY. HPV-Associated Head and Neck Cancer. J Natl Cancer Inst. 2015; 107:djv344. [PubMed: 26656751]
55. Smith EM, Rubenstein LM, Ritchie JM, Lee JH, Haugen TH, Hamsikova E, et al. Does pretreatment seropositivity to human papillomavirus have prognostic significance for head and neck cancers? Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008; 17:2087–96.
56. Holzinger D, Wichmann G, Baboci L, Michel A, Hofler D, Wiesenfarth M, et al. Sensitivity and specificity of antibodies against HPV16 E6 and other early proteins for the detection of HPV16-driven oropharyngeal squamous cell carcinoma. International journal of cancer. 2017; 140:2748–57. [PubMed: 28316084]

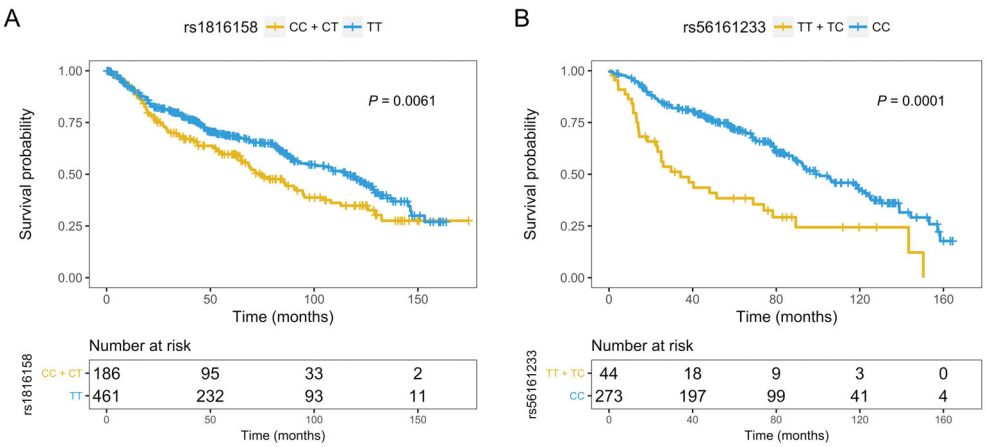

Figure 1. Kaplan-Meier analysis of rs1816158 (oral cavity cancer) and rs56161233 (laryngeal cancer)

Survival curves including all cases from both populations (discovery and validation) stratified by genotype are shown for (A) rs1816158 (oral cavity cancer), (B) rs56161233 (laryngeal cancer). Yellow lines represent survival times for heterozygous homozygous for the minor allele (effect allele), while blue lines depict survival times for subjects homozygous for the major allele (other allele). *P*-values for Log-rank tests are shown. Only subjects with tumors of the indicated sites are shown on the plots for each SNP.

Table 1

Clinical and pathological characteristics of study subjects

	Cases, *n* (%)		
	Massachusetts study ***n*** **= 847**	**M.D. Anderson study** ***n*** **= 1236**	**Total** ***n*** **= 2083**
Age at diagnosis			
≤50	181 (21.4)	336 (27.2)	517 (24.8)
>50 to ≤60	298 (35.2)	457 (37.0)	755 (36.2)
>60 to ≤70	236 (27.9)	303 (24.5)	539 (25.9)
>70	132 (15.6)	140 (11.3)	272 (13.1)
Sex			
Female	629 (74.3)	943 (76.3)	15742(75.5)
Male	218 (25.7)	293 (23.7)	511 (24.5)
Race			
Caucasian	796 (94)	1115 (90.2)	1911 (91.7)
Other	51 (6)	121 (9.8)	172 (8.3)
Smoking status			
Current	93 (11)	463 (37.5)	556 (26.7)
Former	501 (59.1)	393 (31.8)	894 (42.9)
Never	216 (25.5)	380 (30.7)	596 (28.6)
Missing	37 (4.4)	0 (0)	37 (1.8)
HPV16, 18, 33, 51 positivity[a]			
Yes	541 (63.9)		
No	306 (36.1)		
Tumor stage			
Stage I & II	243 (28.7)	297 (24)	540 (25.9)
Stage III & IV	604 (71.3)	939 (76)	1543 (74.1)
Tumor Site			
Oral cavity	298 (35.2)	362 (29.3)	662 (31.7)
Pharynx	413 (48.8)	685 (55.4)	1098 (52.7)
Larynx	136 (16.1)	189 (15.3)	325 (15.6)
Follow-up (events, time to follow-up)			
Censored: *n*, mean months	515, 81.3	749, 61.6	1264, 69.6
Death: *n*, mean months	332, 43.1	487, 48.6	819, 46.4

[a]Any HPV seropositivity for listed high-risk HPV serotypes (16, 11, 33 or 55), was only available for Massachusetts study subjects.

Table 2

Summary statistics for strongest miR-SNP associations with overall HNSCC survival in the joint analysis

					Discovery phase[a]		Validation phase[b]		Joint analysis[c]	
SNP	**Chr:pos[d]**	**Associated Gene[e]**	**Genomic Context[f]**	**Major/Minor allele**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**
Overall HNSCC SNPs										
rs41276823	9:116359270	*RGS3*	3′UTR	G/A	0.47 (0.32–0.71)	2.6E-04	0.86 (0.61–1.21)	3.8E-01	0.63 (0.48–0.82)	5.2E-04
rs3831960	1:45483218	*ZSWIM5*	3′UTR	C/DEL	1.58 (1.21–2.05)	6.5E-04	1.28 (1.03–1.61)	2.8E-02	1.34 (1.13–1.60)	7.0E-04
rs16988668	19:58850756	*ZSCAN22*	3′UTR	G/C	1.72 (1.26–2.34)	6.5E-04	1.37 (1.06–1.77)	1.7E-02	1.41 (1.15–1.72)	8.1E-04
Oral cavity SNPs										
rs3831960	1:45483218	*ZSWIM5*	3′UTR	C/DEL	2.54 (1.62–3.97)	4.4E-05	1.52 (1.03–2.26)	3.5E-02	1.73 (1.31–2.29)	1.3E-04
rs75820821	19:41829457	*CCDC97*	3′UTR	C/T	2.45 (1.52–3.96)	2.5E-04	1.27 (0.73–2.20)	4.0E-01	1.89 (1.32–2.71)	5.3E-04
rs1816158	11:122026460	*MIR100HG*	LncRNA	T/C	1.89 (1.31–2.71)	6.0E-04	1.22 (0.85–1.76)	2.7E-01	1.56 (1.21–2.00)	5.3E-04
Pharyngeal SNPs										
rs56312243	19:55899602	*RPL28*	3′UTR	C/T	2.22 (1.40–3.51)	7.0E-04	1.13 (0.76–1.67)	5.5E-01	1.50 (1.11–2.02)	7.8E-03
rs77506493	19:52795158	*ZNF766*	3′UTR	C/T	0.45 (0.28–0.71)	5.9E-04	0.81 (0.59–1.11)	2.0E-01	0.71 (0.55–0.91)	8.3E-03
rs45494801	19:55902084	*RPL28*	3′UTR	T/C	2.25 (1.43–3.53)	4.5E-04	1.11 (0.73–1.67)	6.3E-01	1.50 (1.10–2.02)	9.2E-03
Laryngeal SNPs										
rs56161233	2:235964133	*SH3BP4*	3′UTR	C/T	4.29 (2.13–8.67)	4.8E-05	1.82 (1.03–3.24)	4.0E-02	2.57 (1.71–3.86)	5.7E-06
rs3777710	6:17611089	*FAM8A1*	3′UTR	G/A	4.04 (1.98–8.22)	1.2E-04	1.74 (1.05–2.89)	3.2E-02	2.13 (1.43–3.16)	1.8E-04
rs12280753	11:116613660	*BUD13*	Downstream	C/T	5.65 (2.26–14.11)	2.1E-04	2.16 (1.32–3.53)	2.2E-03	2.14 (1.40–3.28)	4.1E-04

[a] Analysis adjusted for age at diagnosis, sex, race, HPV seropositivity (see Methods), tobacco use (lifetime pack-years smoked) and alcohol consumption (average drinks per week), and tumor stage. Overall HNSCC; *n* = 847, oral cavity; *n* = 298, pharynx; *n* = 413, larynx *n* = 136.

[b] Analysis adjusted for age at diagnosis, sex, race, smoking status and tumor stage. Overall HNSCC; *n* = 1236, oral cavity; *n* = 364, pharynx; *n* = 687, larynx *n* = 189.

[c] Analysis adjusted as described in b. Overall HNSCC; *n* = 2046, oral cavity; *n* = 648, pharynx; *n* = 1081, larynx *n* = 317. Note, 37 discovery phase subjects were missing data on smoking status (never, former, current) and were omitted from the joint population analysis.

[d] chr, chromosome; pos, position, according to NCBI Genome Build 37 (hg19).

[e] Gene that variant is located within or associated with (closest gene according to NCBI Genome Build 37 (hg19)).

[f] UTR, untranslated region; ncRNA, non-coding RNA variant.

Table 3

Association between HNSCC survival and expression of genes containing HNSCC-associated miR-SNPs

	Unadjusted		Adjusted[a]	
	HR (95% CI)	***P*-value**	**HR (95% CI)**	***P*-value**
***MIR100HG*-embedded genes (oral cavity cases only, *n* = 303)**				
miR-100[b]	1.18 (1.00–1.40)	0.048	1.25 (1.06–1.49)	0.009
miR-125b-1[b]	0.99 (0.78–1.25)	0.959	1.04 (0.82–1.31)	0.732
let7a-2[b]	0.90 (0.73–1.11)	0.326	0.92 (0.74–1.14)	0.446
Other (laryngeal cases only, *n* = 108)				
SH3BP4 expression	1.16 (0.68–2.00)	0.57	1.13 (0.65–1.97)	0.67

Expression data (miRNA-seq for miR-100, miR125b-1 and let7a-2; RNA-seq for *SH3BP4*) were tested for their association with survival using data from the TCGA-HNSC project

MIR100HG-derived genes and *SH3BP4* were tested for their association with survival times in oral cavity cancer cases and laryngeal cancer cases, respectively.

[a]Cox proportional hazards models were adjusted for age, sex, race (white vs non-white) and tumor stage (Low stage or high stage)

[b]Complete miRNA-seq data, averaged across all detected miRNA isoforms was used to test association between miRNA expression and survival times

Gene expression was modelled as a continuous variable in all analyses

Table 4

miRanda/mirSVR predicted miRNA-mRNA interactions overlapping SNPs associated with HNSCC survival

					miRNA expression percentile[e]	
SNP	**Target gene[a]**	**mRNA[b]**	**miRNA[c]**	**mirSVR score percentile[d]**	**Normal tissue**	**Tumor tissue**
rs16988668	*ZSCAN22*	NM_181846	miR-873-5p	0.17	0.51	0.67
	ZSCAN22	NM_181846	miR-339-5p	0.27	0.90	0.92
rs3831960	*ZSWIM5*	NM_020883	miR-96-3p	0.06	0.46	0.49
	ZSWIM5	NM_020883	miR-3171	0.17	0.16	0.52
rs56312243	*RPL28*	NM_001136134	miR-661	0.18		0.57
rs77506493	*ZNF766*	NM_001010851	miR-135b-3p	0.05	0.75	0.81
	ZNF766	AK024074	miR-135b-3p	0.09	0.75	0.81
rs56161233	*SH3BP4*	NM_014521	miR-2110	0.01	0.71	0.68
	SH3BP4	NM_014521	miR-548m	0.08		
	SH3BP4	NM_014521	miR-593-3p	0.13		
	SH3BP4	NM_014521	miR-3150a-3p	0.23	0.50	0.23
rs3777710	*FAM8A1*	NM_016255	miR-1292-5p	0.18	0.42	0.55

[a] miRNA target gene containing SNP, according to NCBI Genome Build 37 (hg19).

[b] RefSeq, NCBI or GenBank Transcript/sequence identifiers.

[c] miRBase IDs of miRs known or predicted to target associated gene.

[d] Calculated based on distribution of all available mirSVR predictions. More negative mirSVR scores, and therefore lower percentile scores, indicate more likely gene regulatory potential.

[e] TCGA-HNSC normal tissue miRNA expression, miRNA-mapped reads per million percentile.

Blank spaces indicate where relevant data was not available