Genome-scale identification of microRNA-related SNPs associated with risk of head and neck squamous cell carcinoma

Owen M. Wilkins, Alexander J. Titus, Jiang Gui, Melissa Eliot, Rondi A. Butler, Erich M. Sturgis, Guojun Li, Karl T. Kelsey, and Brock C. Christensen

Abstract
Polymorphisms in microRNAs and their target sites can disrupt microRNA-dependent gene regulation, and have been associated with cancer susceptibility. However, genome-scale analyses of microRNA-related genetic variation in cancer are lacking. We tested the associations of ~40,000 common [minor allele frequency (MAF) ≥5%], microRNA-related single nucleotide polymorphisms (miR-SNPs), with risk of head and neck squamous cell carcinoma (HNSCC) in a discovery population, and validated selected loci in an independent population among a total of 2198 cases and 2180 controls. Joint analyses across the discovery and validation populations revealed six novel miR–SNP associations with risk of HNSCC. An upstream variant of MIR548H4 (rs7834169), replicated its association with overall HNSCC risk as well as risk of oral cavity cancer. Four other variants were specifically associated with oral cavity cancer risk (rs16914640, rs1134367, rs7306991 and rs1373756). 3’ UTR variant of HADH, rs221347 and rs4975616, located within known cancer risk locus 5p15.33, were specific to risk of laryngeal cancer. High confidence predicted microRNA binding sites were identified for CLEC2D, LOC37443, KDM8 and HADH overlapping rs16914640, rs7306991, rs1134367 and rs221347, respectively. Furthermore, we identified several microRNA interactions with KDM8 and HADH predicted to be disrupted by genetic variation at rs1134367 and rs221347. These results suggest microRNA-related genetic variation may contribute to the genetic susceptibility of HNSCC, and that more powerful evaluation of this class of genetic variation and their relationship with cancer risk is warranted.

Introduction
Head and neck squamous cell carcinoma (HNSCC) is expected to be responsible for ~61,000 new cases and >13,000 deaths throughout 2016 in the USA alone. Considerable morbidity is associated with HNSCC as surgery and/or definitive radiation can result in disfigurement, difficulty swallowing, breathing and speaking. Established risk factors include alcohol consumption, tobacco use and human papillomavirus (HPV) infection and diet, oral hygiene and marijuana use have also been implicated. Patients diagnosed with late stage disease have >50% chance of recurrence or development of distant metastases therefore improvement of effective risk stratification and early detection strategies are critical toward reducing patient morbidity and mortality.

Genetic variation in protein-coding regions directing alcohol metabolism, DNA repair, apoptosis, cell cycle control, mitochondrial function and at HPV-related pathways has been associated...
with HNSCC risk (9–12). A recent Genome-wide association study (GWAS) of oral cavity cancer and oropharyngeal cancer identified seven novel loci associated with these cancer sites (13). While variation in protein-coding regions has received much attention in such studies, most disease associated variants lie within non-coding regions (14). Polymorphisms within microRNA (miRNA) targets sites, miRNA seed sites and miRNA processing genes can modify miRNA–mRNA binding, as well as create or destroy target sites entirely (15). However, such miRNA-related SNPs (miR-SNPs) are generally poorly represented on—and/or in low linkage disequilibrium (LD) with—SNPs on or imputed using Genome-wide association study (GWAS) arrays (16), and although candidate gene studies have identified miR-SNPs associated with HNSCC susceptibility (17–23), comprehensive evaluation of the relation between miR-SNPs and HNSCC susceptibility has yet to be investigated. To address these issues, we conducted a genome-scale evaluation of the association between miR-SNPs and HNSCC susceptibility, utilizing a novel genotyping array containing validated and predicted miR-SNPs at all levels of the miRNA pathway.

Materials and methods

Study participants

To identify miR-SNPs associated with HNSCC susceptibility, we obtained and analyzed data from two independent population-based case–control studies of HNSCC using a discovery (Massachusetts study; n, cases = 904, n, controls = 1051) and validation phase (M.D. Anderson study; n, cases = 1338, n, controls = 1356) approach. All subjects provided written informed consent as approved by the Institutional Review Board of the participating institutions. For the Massachusetts study, which has been described previously (10), incident cases of HNSCC were identified at nine hospitals in the Boston, MA, metropolitan area between 1999–2003 (Phase I) and 2006–2011 (Phase II). The state cancer registry confirmed >95% of incident cases in the study area were captured. Histology was reported by participating hospitals’ pathologists and confirmed by an independent study pathologist. Control subjects were matched to cases by age (±3 years), sex and town of residence. Demographic and exposure data were collected using self-administered questionnaires that were reviewed by trained study interviewers, and clinical information was obtained through medical chart reviews.

For the validation phase, histopathologically confirmed incident cases of HNSCC were recruited at The University of Texas M.D. Anderson Cancer Center between December 1996 and 2008. Details of patient recruitment to the study have been described previously (24). Hospital visitors at the M.D. Anderson Cancer Center over the same time were recruited as cancer-free controls and completed a questionnaire so that demographic and exposure data could be collected. Cases and controls were matched based on age and sex. Although categorical indicators of tobacco and alcohol use were available, continuous measures of these exposures (lifetime pack-years smoked and lifetime average number of drinks per week) were not available for most genotyped subjects. International Classification of Disease, Ninth Revision (ICD-9) codings and relevant pathological analyses were used to assign HNSCC cases to one of three groups; oral cavity, pharyngeal or laryngeal cancer. HNSCC cases included ICD-9 diagnosis codes 141, 143–146, 148, 149 and 161. For site-specific analyses with diagnosis codes were classified with ICD-9 codes as follows: oral cavity; 141.1–141.5, 141.8, 141.9, 143–145.2, 145.5–145.9, pharyngeal; 141.0, 141.6, 145.3, 145.4, 146, 148, 149.0, 149.1, laryngeal; 161.

Genotyping

DNA from Massachusetts study subjects was extracted from whole blood or buccal cells using the QIAamp DNA mini kit (Qiagen) and was genotyped using the Axiom miRNA Target Site Genotyping Array (Affymetrix). The standard array (no marker customization) was used. To reduce the burden of multiple testing, we limited our analyses to the 40 034 markers present at a MAF >5% (after quality control). In total, the array interrogates ~238 000 SNPs and SNPs in miRNAs, miRNA regulatory regions, miRNA target sites and miRNA processing proteins. Five online databases were used to select markers for the array: PolymiRTS (25), dPARE (26), Patrocles (27), miRNASNP (28) and miRcroRNA.org (29). The array also includes ~18 100 haplotype tagging miR-SNPs predicted to affect miRNA binding in mRNA 3’UTRs, identified by Thomas et al. (30). Quality control was performed according to the Axiom Best Practice Genotyping Analysis Workflow. Two samples below the Dish QC (DDC) metric threshold of 0.82, and 24 samples below the call rate threshold of 97% were excluded from analysis. About 16 597 markers fell below the suggested thresholds for SNP QC and were omitted. Of the remaining markers, 40 286 were present in the discovery population at a MAF >5%. 252 SNPs deviated from Hardy–Weinberg equilibrium (HWE) in Caucasian control study subjects (P < 1 × 10−6) and were removed, leaving the 40 034 markers for the discovery phase. Validation genotyping of DNA from M.D. Anderson study subjects was performed using the MassARRAY iPLEX gold assay (Sequenom). In the validation population 288 subjects with a call rate of >95%, four variants deviating from HWE in Caucasian controls, and monomorphic variants were removed.

Statistical analysis

Unconditional multivariable logistic regression for the association between miR-SNP genotype and HNSCC risk was performed using R Statistical Software version 3.3.0 to calculate odds ratios (ORs) and 95% confidence intervals (95% CI), assuming a dominant model of inheritance. HWE was calculated with exact tests using the HWExact function (pvaluetype = ‘dost’) from R package HardyWeinberg. Markers deviating from HWE at P < 1 × 10−6 were omitted. Statistical models in the discovery phase were adjusted for age (≥50, >50 to <60, ≥60 to <70, ≥70), sex (male, female), race (Caucasian, other), HPV serology (positive/negative for any of HPV types; HPV16, 18, 33, 51), alcohol consumption (lifetime average number of drinks per week) and tobacco use (lifetime pack-years). Subjects were ordered into discrete groups based on the population quartile distribution of alcohol consumption (≥1, >1 to ≤6, >6 to ≤31, >31) and tobacco use (≤2, >2 to ≤6, >6 to ≤14, >14). Genomic inflation factor λc was calculated using the R package GenABEL (31). To select markers for the validation phase, the strongest associations from each analysis, based on adjusted P values, were pruned using pairwise LD patterns determined from all available study participants selecting the two SNPs in lowest pairwise LD per gene calculated using the R package genetics. 32, 27, 28 and 27 markers were selected for validation genotyping in the overall, oral cavity cancer, pharyngeal cancer and laryngeal cancer-susceptible analyses, respectively (validation marker lists Supplementary Tables 1–4, available at Carcinogenesis Online), and tested with a joint analysis approach (32). Models for the validation population and joint analyses were adjusted for age, sex, race and smoking status (current, former, never). Extensive data for continuous measures of tobacco and alcohol consumption were not available in validation phase subjects. We considered validated variants as those with P < 1 × 10−3 in the joint analysis, a lower P in the joint analyses than the discovery phase analyses, and a consistent predicted direction of effect across study populations. Validated variants were selected for additional follow up analyses. The LD-pruning function in PLINK was used to determine the number of effectively independent variants in each analysis, using a 100-variant window size, a five-variant window shift and an r2 of 0.1. Population-specific genome-wide significance thresholds were based on Bonferroni correction (P = 3.5 × 10−6 to 1 d.p.).

Identification of HNSCC-associated loci predicted to disrupt miRNA target sites

Genomic coordinates for all available miRNA target site predictions were used to model the miranda algorithm and scored for meaningful down-regulation of target genes using the mirSVR algorithm (downloaded from the microRNA.org database [http://www.microrna.org/microrna/].)
To investigate the association of miR-SNPs across the genome with HNSCC risk, we genotyped ~238,000 SNPs in miRNA target sites, miRNA genes, and genes in the miRNA biogenesis pathway in a population-based case control study based in Massachusetts with the Affymetrix Axiom miRNA Target Site Array. For discovery, we tested the associations of 40,034 genetic variants genotyped using the Affymetrix miRTS array MAF ≥5%, with overall and tumor site-specific (oral cavity, pharyngeal, laryngeal) HNSCC risk, in 1051 controls and 904 cases from the Massachusetts study (Table 1). We applied unconditional logistic regression adjusted for age at diagnosis, sex, race, HPV seropositivity, tobacco consumption and alcohol consumption. Across the overall and site-specific analyses, the strongest associations were observed for rs9791083 (overall HNSCC, CN51KA1, OR; 0.73, 95% CI; 0.53–0.79), rs2018329 (oral cavity cancer, RUNX1, OR; 0.43, 95% CI; 0.19–0.64), rs59858117 (pharyngeal cancer, ATG3, OR; 2.47, 95% CI; 1.62–3.76) and rs221347 (laryngeal cancer, HADH, OR; 2.74, 95% CI; 1.79–4.19). Generally, the strongest associations with HNSCC risk were observed for overall HNSCC and oral cavity cancer, with 41 and 50 variants at an adjusted P < 1 × 10^{-4}, respectively, compared to 35 and 33 variants with adjusted P values below this threshold in the pharyngeal and laryngeal cancer analyses. Genomic inflation factor λ was calculated to measure P value inflation for the adjusted analyses and was close to 1 for all adjusted analyses, indicating minimal contribution of population stratification (Supplementary Figure 1, available at Carcinogenesis Online). Compared with the analysis of overall risk, effect size estimates in the site-specific analyses were generally larger (Supplementary Figure 2, available at Carcinogenesis Online), particularly for laryngeal cancer associated variants. Summary statistics for all associations tested in the discovery phase are available in Supplementary Datasets 1–4, available at Carcinogenesis Online.

Validation phase associations of miR-SNPs with HNSCC risk

An independent study population of incident HNSCC conducted at The University of Texas M.D. Anderson Cancer Center was used to validate findings from the discovery phase. MAFs of the 119 variants selected for validation were highly concordant across study populations (Supplementary Figure 3, available at Carcinogenesis Online). Eight associations from the discovery phase were validated in a joint analysis adjusting for age at diagnosis, sex, race and smoking status (Table 2, Supplementary Tables 1–4, available at Carcinogenesis Online). Supplementary Figure 4, available at Carcinogenesis Online, shows the relationship between effect size (ORs), MAF and statistical significance for the validated associations. For overall risk of HNSCC, rs7834169 upstream of MIR548H4, was replicated (Table 2), and does not overlap with known enhancer or promoter marks (Supplementary Table 5, available at Carcinogenesis Online). In the discovery phase, modest associations were observed for variants in pairwise LD with rs7834169 (Supplementary Figure 5A, available at Carcinogenesis Online), and conditional analysis revealed that they were dependent upon rs7834169 (Supplementary Figure 6A, available at Carcinogenesis Online). MIR548H4 was not expressed in the analysis of normal tissue from TCGA HNSCC subjects.

In the analysis of oral cavity cancer risk, four miR-SNPs were validated (Table 2), rs16914640, located within exon 1 of natural killer cell receptor CLEC2D, was associated with increased oral cavity cancer risk (Table 2, Supplementary Figure 4, available at Carcinogenesis Online). Another regional variant (rs7306991) in a pseudogene of CLEC2D replicated its association with oral cavity cancer, though these SNPs were not in LD (Table 2; Supplementary Figure 5E, available at Carcinogenesis Online). rs1134367, located in the 3′UTR of histone demethylase KDM8, was associated with reduced risk of oral cavity cancer (Table 2, Supplementary Figure 4, available at Carcinogenesis Online). Ancestry informative marker rs1373756 (Table 2) was also associated with oral cavity cancer risk.

While no variants were validated in the joint analyses of pharyngeal cancer risk, we identified two variants associated with risk of laryngeal cancer; rs221347 (3′UTR, HADH) and rs4975616 (upstream, CLPTM1L) (Table 2). rs221347 was the strongest association observed in any of the discovery phase analyses. Stratified analyses suggested potential effect modification of the rs221347 association by smoking status, as moderate to strong levels of association with laryngeal cancer risk were observed.
in never and former smokers, while no association was seen for current smokers across any of the study phases (Supplementary Table 6, available at Carcinogenesis Online). rs4975616 is located at rs7306991 bases and all three miRNAs were among the most abundantly expressed in either normal head and neck tissue, or primary tumor tissue samples (Supplementary Table 7, available at Carcinogenesis Online).

Predicted miRNA–mRNA interactions at HNSCC-associated loci

To identify miRNA–mRNA interactions disrupted by genetic variation, we intersected predicted miRNA target sites made using the miRanda algorithm with HNSCC associated SNPs outlined in Table 2 (29,33). Additionally, we calculated miRNA expression percentiles using miRNA-seq data of normal and primary tumor head and neck tissue samples from TCGA. We identified multiple high confidence predicted miRNA target sites at HNSCC-associated loci, many of which are expressed at a high level in normal and/or tumor head and neck tissue samples (Table 3). Predicted target sites for miR-3166 were identified for two sequences aligned to CLEC2D at rs16914640. Predicted miRNA target sites miR-3148, miR-605, miR-619 and miR-875-3p were identified for sequences aligned to LOC374443 at rs7306991 (Table 3). Binding sites for miR-3136 were identified for two target sites miR-3148, miR-605, miR-619 and miR-875-3p were sequences aligned to multiple high confidence predicted miRNA target sites at HNSCC-head and neck tissue samples from TCGA. We identified multiple miRNAs to

### Table 1. Clinical and pathological characteristics of study subjects

<table>
<thead>
<tr>
<th></th>
<th>Controls, n (%)</th>
<th>Cases, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Massachusetts study</td>
<td>M.D. Anderson study</td>
</tr>
<tr>
<td></td>
<td>n = 1051</td>
<td>n = 1356</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤50</td>
<td>169 (16.1)</td>
<td>441 (32.5)</td>
</tr>
<tr>
<td>&gt;50 to ≤60</td>
<td>345 (32.8)</td>
<td>468 (34.5)</td>
</tr>
<tr>
<td>&gt;60 to ≤70</td>
<td>331 (31.5)</td>
<td>316 (23.3)</td>
</tr>
<tr>
<td>&gt;70</td>
<td>206 (19.6)</td>
<td>131 (9.7)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>276 (26.3)</td>
<td>360 (26.5)</td>
</tr>
<tr>
<td>Male</td>
<td>775 (73.7)</td>
<td>996 (73.5)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>943 (89.7)</td>
<td>1160 (85.5)</td>
</tr>
<tr>
<td>Other</td>
<td>108 (10.3)</td>
<td>196 (14.5)</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>129 (12.3)</td>
<td>165 (12.2)</td>
</tr>
<tr>
<td>Former</td>
<td>489 (46.5)</td>
<td>405 (29.9)</td>
</tr>
<tr>
<td>Never</td>
<td>419 (39.9)</td>
<td>573 (42.3)</td>
</tr>
<tr>
<td>Missing</td>
<td>14 (1.3)</td>
<td>213 (15.7)</td>
</tr>
<tr>
<td>HPV16, 18, 33, 51 positivity*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>360 (34.3)</td>
<td>564 (62.4)</td>
</tr>
<tr>
<td>Negative</td>
<td>691 (65.7)</td>
<td>340 (37.6)</td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages I and II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages III and IV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral cavity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharynx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>243 (26.9)</td>
<td>331 (24.7)</td>
</tr>
<tr>
<td></td>
<td>57 (6.3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>328 (36.3)</td>
<td>384 (28.7)</td>
</tr>
<tr>
<td></td>
<td>145 (16)</td>
<td>220 (16.4)</td>
</tr>
</tbody>
</table>

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

To determine if genetic variation at the HNSCC-associated loci may alter miRNA–mRNA interactions, we used three databases on predicted effects of genetic variation at miRNA-related loci (PolymiRTSv3.0, miRNASNP and MirSNP). For the miRNAs whose binding was predicted to be disrupted at risk-associated loci miRNA expression percentiles were calculated as above. Genetic variation at rs1134367 was predicted to disrupt binding to 25 miRNAs to KDM8 (Table 4). Notably, context+ score differences predicting the effects of miR-149-3p, miR-4728-5p and miR-6883-5p binding to KDM8 were all in the 2nd percentile of all context+ scores obtained from the PolymiRTS database, indicating with high likelihood that the miRNA target site is disrupted by genetic variation at the HNSCC-associated loci identified in this study.

Predicted disruption of miRNA–mRNA interactions by HNSCC-associated miR-SNPs

rs1134367, two of which were expressed at or above the 73rd percentile in normal head and neck tissues, and at or above the 82nd percentile in tumor tissues (Table 3). These data highlight multiple miRNA-mRNA interactions with potential to be disrupted by genetic variation at the HNSCC-associated loci identified in this study.

rs1134367 was predicted to disrupt binding of 25 miRNAs to KDM8 (Table 4). Notably, context+ score differences predicting the effects of miR-149-3p, miR-4728-5p and miR-6883-5p binding to KDM8 were all in the 2nd percentile of all context+ scores obtained from the PolymiRTS database, indicating with high likelihood that the miRNA target site is disrupted by genetic variation at the HNSCC-associated loci identified in this study.

rs1134367 was predicted to disrupt binding of 25 miRNAs to KDM8 (Table 4). Notably, context+ score differences predicting the effects of miR-149-3p, miR-4728-5p and miR-6883-5p binding to KDM8 were all in the 2nd percentile of all context+ scores obtained from the PolymiRTS database, indicating with high likelihood that the miRNA target site is disrupted by genetic variation at the HNSCC-associated loci identified in this study.

rs1134367 was predicted to disrupt binding of 25 miRNAs to KDM8 (Table 4). Notably, context+ score differences predicting the effects of miR-149-3p, miR-4728-5p and miR-6883-5p binding to KDM8 were all in the 2nd percentile of all context+ scores obtained from the PolymiRTS database, indicating with high likelihood that the miRNA target site is disrupted by genetic variation at the HNSCC-associated loci identified in this study.
Table 2. Summary statistics for miR-SNPs associated with HNSCC susceptibility

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr:pos</th>
<th>Associated gene</th>
<th>Genomic context</th>
<th>EAF (cases/controls)</th>
<th>Discovery phase 95% CI</th>
<th>Validation phase 95% CI</th>
<th>Joint analysis 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall HNSCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.07/0.11</td>
<td>0.80 (0.63–1.00)</td>
<td>0.70 (0.59–0.82)</td>
</tr>
<tr>
<td>rs784169</td>
<td>8:2691029</td>
<td>MIR548H4</td>
<td>Intergenic</td>
<td></td>
<td>0.60 (0.46–0.76)</td>
<td>1.3E–04</td>
<td>1.7E–05</td>
</tr>
<tr>
<td>Oral cavity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.8E–04</td>
<td>1.35 (1.02–1.77)</td>
<td>0.68 (0.48–0.96)</td>
</tr>
<tr>
<td>rs16914640</td>
<td>12:9822387</td>
<td>CLEC2D</td>
<td>Missense</td>
<td>0.19/0.14</td>
<td>1.70 (1.37–2.17)</td>
<td>3.6E–04</td>
<td>5.7E–03</td>
</tr>
<tr>
<td>rs784169</td>
<td>8:2691029</td>
<td>MIR548H4</td>
<td>Intergenic</td>
<td>0.07/0.11</td>
<td>0.49 (0.33–0.73)</td>
<td>3.5E–04</td>
<td>5.4E–04</td>
</tr>
<tr>
<td>rs1134367</td>
<td>16:2723383</td>
<td>MDM2</td>
<td>3’ UTR</td>
<td>0.46/0.51</td>
<td>0.58 (0.43–0.77)</td>
<td>2.0E–04</td>
<td>3.6E–04</td>
</tr>
<tr>
<td>rs137576</td>
<td>18:49786306</td>
<td>DCC</td>
<td>Intergenic</td>
<td>0.08/0.12</td>
<td>0.49 (0.32–0.73)</td>
<td>5.8E–04</td>
<td>0.64 (0.48–0.96)</td>
</tr>
<tr>
<td>rs7306991</td>
<td>12:9810451</td>
<td>LOC374443</td>
<td>ncRNA</td>
<td>0.18/0.13</td>
<td>1.63 (1.22–2.18)</td>
<td>8.7E–04</td>
<td>1.35 (1.02–1.77)</td>
</tr>
<tr>
<td>Laryngeal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.8E–04</td>
<td>1.2E–01</td>
<td>1.45 (1.20–1.76)</td>
</tr>
<tr>
<td>rs221347</td>
<td>4:10895622</td>
<td>HADH</td>
<td>3’ UTR</td>
<td>0.14/0.10</td>
<td>2.74 (1.79–4.19)</td>
<td>3.2E–06</td>
<td>1.96 (1.48–2.61)</td>
</tr>
<tr>
<td>rs4975616</td>
<td>5:1315660</td>
<td>CLPTM1</td>
<td>Intergenic</td>
<td>0.37/0.45</td>
<td>0.52 (0.35–0.76)</td>
<td>8.1E–04</td>
<td>0.76 (0.54–1.07)</td>
</tr>
</tbody>
</table>

Table 3. miRanda/mirSVR predicted miRNA–mRNA interactions for SNPs associated with HNSCC risk

<table>
<thead>
<tr>
<th>SNP</th>
<th>Accession</th>
<th>Gene</th>
<th>Allele</th>
<th>miRNA</th>
<th>mirSVR score</th>
<th>Normal tissue</th>
<th>Tumor tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16914640</td>
<td>AK310554</td>
<td>CLEC2D</td>
<td>C</td>
<td>miR-3166</td>
<td>−0.142</td>
<td>0.42</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>BCO63128</td>
<td>CLEC2D</td>
<td>C</td>
<td>miR-3166</td>
<td>−0.134</td>
<td>0.42</td>
<td>0.37</td>
</tr>
<tr>
<td>rs1134367</td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>T</td>
<td>miR-3136</td>
<td>−0.246</td>
<td>0.51</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>AK308499</td>
<td>KDM8</td>
<td>T</td>
<td>miR-3136</td>
<td>−0.168</td>
<td>0.51</td>
<td>0.66</td>
</tr>
<tr>
<td>rs7306991</td>
<td>AYO084051</td>
<td>LOC374443</td>
<td>A</td>
<td>miR-3148</td>
<td>−0.296</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>AYO084051</td>
<td>LOC374443</td>
<td>A</td>
<td>miR-605</td>
<td>−0.232</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>NR_002814</td>
<td>LOC374443</td>
<td>A</td>
<td>miR-3148</td>
<td>−0.214</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>NR_002814</td>
<td>LOC374443</td>
<td>A</td>
<td>miR-605</td>
<td>−0.165</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td>rs221347</td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-548g</td>
<td>−1.131</td>
<td>0.43</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-3065-5p</td>
<td>−0.910</td>
<td>0.78</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-105</td>
<td>−0.134</td>
<td>0.73</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-548a-3p</td>
<td>−0.107</td>
<td>&lt;0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-548e</td>
<td>−0.106</td>
<td>0.59</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-548f</td>
<td>−0.106</td>
<td>0.06</td>
<td>0.78</td>
</tr>
</tbody>
</table>

HNSCC tumor tissue samples, or both. Six significant eQTLs have been documented for rs1134367 with KDM8 over a range of tissues (Supplementary Table 8, available at Carcinogenesis Online), consistent with the hypothesis that rs1134367 alters miRNA-dependent KDM8 regulation. In addition, four miRNA target sites in HADH were predicted to be disrupted by rs221347. Presence of the T allele for rs221347 was predicted to create a binding site for miR-548g-3p. The context+ score difference for the effects of rs221347 on miR-548g-3p binding to HADH were in the top 7% of scores. A very high confidence binding site for miR-548g-3p in HADH was also predicted by miRanda/mirSVR (Table 3). These data indicate miRNA-related genetic variation at HNSCC risk-associated rs1134367 and rs221347 alters miRNA-dependent gene regulation of KDM8 and HADH in normal head and neck tissue, as well as primary HNSCC tumor tissue.

Discussion

Previous studies testing the relation of miR-SNPs with HNSCC susceptibility and other cancers have focused on relatively few markers. In addition, miR-SNPs have generally not been well covered by GWAS arrays, making it difficult to accurately...
Table 4. Predicted effects of SNPs associated with HNSCC risk upon miRNA–mRNA interactions

<table>
<thead>
<tr>
<th>SNP</th>
<th>Accessiona</th>
<th>Geneb</th>
<th>Allelc</th>
<th>mirNad</th>
<th>Context + score difference (percentile)a</th>
<th>Predicted SNP effect (miRNASNP)</th>
<th>Predicted SNP effect (MirSNP)</th>
<th>miRNA expression (percentile)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1134367</td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-149-3p</td>
<td>0.02</td>
<td>Gain</td>
<td>Create</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>NM_024773</td>
<td>KDM8</td>
<td>C</td>
<td>miR-149-3p</td>
<td>0.02</td>
<td>Gain</td>
<td>Create</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-30b-3p</td>
<td>0.39</td>
<td>Gain</td>
<td>Create</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>NM_024773</td>
<td>KDM8</td>
<td>C</td>
<td>miR-30b-3p</td>
<td>0.39</td>
<td>Gain</td>
<td>Create</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-3689c</td>
<td>0.42</td>
<td>Gain</td>
<td>Create</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>NM_024773</td>
<td>KDM8</td>
<td>C</td>
<td>miR-3689c</td>
<td>0.42</td>
<td>Gain</td>
<td>Create</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>T</td>
<td>miR-4779</td>
<td>0.07</td>
<td>Gain</td>
<td>Create</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>NM_024773</td>
<td>KDM8</td>
<td>T</td>
<td>miR-4779</td>
<td>0.07</td>
<td>Gain</td>
<td>Create</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>T</td>
<td>miR-4536-5p</td>
<td>0.28</td>
<td>Gain</td>
<td>Create</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>NM_024773</td>
<td>KDM8</td>
<td>C</td>
<td>miR-4536-5p</td>
<td>0.28</td>
<td>Loss</td>
<td>Break</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-4728-5p</td>
<td>0.02</td>
<td>Gain</td>
<td>Create</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>NM_024773</td>
<td>KDM8</td>
<td>C</td>
<td>miR-4728-5p</td>
<td>0.02</td>
<td>Gain</td>
<td>Create</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>T</td>
<td>miR-4695-5p</td>
<td>0.42</td>
<td>Gain</td>
<td>Create</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-541-3p</td>
<td>0.33</td>
<td>Gain</td>
<td>Create</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-654-5p</td>
<td>0.34</td>
<td>Gain</td>
<td>Create</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6769a-5p</td>
<td>0.28</td>
<td>Gain</td>
<td>Create</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6769b-5p</td>
<td>0.27</td>
<td>Gain</td>
<td>Create</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6778-5p</td>
<td>0.22</td>
<td>Gain</td>
<td>Create</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6779-5p</td>
<td>0.39</td>
<td>Gain</td>
<td>Create</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6780a-5p</td>
<td>0.39</td>
<td>Gain</td>
<td>Create</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6785-5p</td>
<td>0.03</td>
<td>Gain</td>
<td>Create</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6799-5p</td>
<td>0.37</td>
<td>Gain</td>
<td>Create</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-6883-5p</td>
<td>0.02</td>
<td>Gain</td>
<td>Create</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-7106-5p</td>
<td>0.36</td>
<td>Gain</td>
<td>Create</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>NM_001145348</td>
<td>KDM8</td>
<td>C</td>
<td>miR-92a-2-5p</td>
<td>0.40</td>
<td>Gain</td>
<td>Create</td>
<td>0.60</td>
</tr>
<tr>
<td>rs221347</td>
<td>NM_001184705</td>
<td>HADH</td>
<td>T</td>
<td>miR-4729</td>
<td>Decrease</td>
<td>Decrease</td>
<td>Decrease</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>NM_001184705</td>
<td>HADH</td>
<td>T</td>
<td>miR-548-3p</td>
<td>0.07</td>
<td>Decrease</td>
<td>Create</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-548-3p</td>
<td>0.07</td>
<td>Decrease</td>
<td>Create</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>NM_005327</td>
<td>HADH</td>
<td>T</td>
<td>miR-4729</td>
<td>Decrease</td>
<td>Decrease</td>
<td>Decrease</td>
<td>0.43</td>
</tr>
</tbody>
</table>

RefSeq, NCBI or GenBank Transcript/sequence identifiers.
Gene location of variant according to NCBI Genome Build 37 (hg19).
Allele of SNP associated with miRNA–mRNA interaction.
mirBase IDs (www.mirbase.org) of miRs known or predicted to target-associated gene.
Context+ score differences were curated from the PolymiRTS database. Percentiles were calculated using the entire distribution of context+ score differences available from PolymiRTS.
TCGA-HNSCC miRNA expression from normal (n = 44) and tumor (n = 523) tissue samples. Expression percentiles for log2-transformed miRNA-mapped reads per million were calculated based on the distribution of all available sequenced miRNAs.

impute miR-SNP genotypes from existing data (16). Here, we employed a genome-scale approach using a novel genotyping platform to determine the contribution of common miR-SNPs to HNSCC susceptibility. We validated eight associations, six of which are in miRNA-related genetic variants. While only a single variant replicated in the analysis of overall HNSCC risk, four replicated in the analysis of oral cavity cancer risk, suggesting effects of miR-SNPs on cancer susceptibility are tissue-specific. All eight variants are located at loci not previously associated with HNSCC risk in the recent and previous GWAS studies of HNSCC (12,13). Genetic variation at validated loci was predicted to disrupt miRNA–mRNA interactions and expression of targeting miRNAs in head and neck tissues was confirmed. Together, these findings suggest miR-SNPs are associated with susceptibility to HNSCC and result in functional changes that contribute to HNSCC tumorigenesis.

The miRNA target site SNPs associated with HNSCC are predicted to alter gene regulation of genes whose role in cancer is supported. CLEC2D—a member of the natural killer cell C-type lectin receptors—is an established modulator of the natural killer cell cytotoxic response whose upregulation facilitates immune evasion in glioblastoma (36,37). Furthermore, the oral cavity cancer risk-associated variant rs7306991 lies within a pseudogene of CLEC2D. Pseudogene expression has been suggested to modulate parental gene expression, as competitive endogenous RNAs (ceRNAs), by competing for miRNA binding (38). BRAF pseudogene overexpression has been shown to induce lymphoma in mice (39), and genetic variation in pseudogenes may disrupt such regulation. Histone demethylase KDM8 has been implicated in cancer cell proliferation and breast cancer metastasis (40–42). Recent work suggests KDM8 is required for late steps of homologous recombination and genome integrity (43). Altered KDM8 expression due to variation at rs1134367 may alter histone methylation and contribute to HNSCC (44), and several significant eQTL hits for rs1134367 with KDM8 expression support this theory. HADH is a mitochondrial protein that plays an essential role in long-chain fatty acid β-oxidation (FAO) (45). HADH deficiency results in a rare disorder of mitochondrial FAO (46). Although located at different genomic loci than HADH, expression of the alpha subunit of HADH (HADHA) is deregulated.
in clear cell renal cell carcinoma (ccRCC) and has been suggested as a potential prognostic biomarker for ccRCC and lung cancer (47,48). Furthermore, fatty acid metabolism is now regarded as a key contributor to cancer cell proliferation (49), therefore it is plausible that deregulation of miRNA-mediated expression of HADH by rs221347 may contribute to carcinogenesis.

Given that previous studies investigating the roles of miR-SNPs in HNSCC risk have been limited to candidate-based approaches, the large scale of our study represents a major strength. Furthermore, use of two independent study populations allowed us to validate our findings and reduce the likelihood of false positive associations. Despite these strengths, our study has some underlying limitations. The tools for miRNA target site prediction are limited by our incomplete knowledge of the principles for functional miRNA targeting, which remains controversial (50). Consequently, our approach is insensitive to yet unidentified genetic variants that may occur in miRNA targets and alter risk of HNSCC. Additionally, relative to most GWA studies, our sample size is modest and replication in larger, more ethnically diverse populations is warranted. Differing demographic and clinical characteristic structures between the populations used in this study, as well as the presence of potential gene-gene interactions that may contribute to HNSCC susceptibility, which were not considered in this study, may have also reduced our ability to identify true associations. Furthermore, given the etiologic importance of HPV infection for oropharyngeal cancers, lack of adjustment for HPV positivity in the validation phase and joint population analyses of pharyngeal cancer risk may have contributed to non-replication observed in this subgroup. Future studies with more complete HPV data are needed to confirm our findings. Finally, where it did not already exist, functional studies of the variant genotypes identified here are needed to further understand the biology of observed associations and discount the potential contribution of other functional variants in LD. However, miR-SNPs are empirically in low LD with variants included on many GWAS panels, strengthening our findings.

In summary, we identified and validated miR-SNPs associated with risk of HNSCC. As our understanding of miRNA biology grows and novel miRNA target site prediction algorithms are developed, additional miR-SNPs may be identified. Here we demonstrated the utility of current knowledge to functionally annotate miR-SNPs associated with HNSCC.

Supplementary material
Supplementary data are available at Carcinogenesis online.

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

Acknowledgements
We would like to thank Kevin C. Johnson for discussions that improved this manuscript.

Conflict of Interest Statement: None declared.

References
35. Liu, C. et al. (2012) MiRSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics, 13, 661.