

ORIGINAL ARTICLE

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

Owen M. Wilkins¹, Alexander J. Titus¹, Jiang Gui², Melissa Eliot³, Rondi A. Butler³, Erich M. Sturgis^{4,5}, Guojun Li^{4,5}, Karl T. Kelsey^{3,6} and Brock C. Christensen^{1,7,8,*}

¹Department of Epidemiology, ²Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA, ³Department of Epidemiology, Brown University, Providence, RI 02912, USA, ⁴Department of Head and Neck Surgery, ⁵Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, TX 77030, USA, ⁶Department of Pathology and Laboratory Medicine, Brown University, Providence, RI 02912, USA, ⁷Department of Molecular and Systems Biology and ⁸Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA

*To whom correspondence should be addressed. Tel: +1 603 650 1827; Fax: +1 603 650 1129; Email: brock.c.christensen@dartmouth.edu

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 (1). Considerable morbidity is associated with HNSCC as surgery and/or definitive radiation can result in disfigurement, difficulty swallowing, breathing and speaking (2). Established risk factors include alcohol consumption, tobacco use and human papillomavirus (HPV) infection and diet, oral hygiene and marijuana use have also been

implicated (3–7). 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 (8).

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

Received: March 20, 2017; Revised: May 12, 2017; Accepted: June 1, 2017

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact journals.permissions@oup.com

Abbreviations

HWE	Hardy–Weinberg equilibrium
HNSCC	head and neck squamous cell carcinoma
HPV	human papillomavirus
miRNA	microRNA
MAFs	minor allele frequencies

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 Boards 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 cases 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 $\geq 5\%$ (after quality control). In total, the array interrogates ~238 000 SNPs and indels 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), dPORE (26), Patrocles (27), miRNASNP (28) and microRNA.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 (DQC) 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 $\geq 5\%$. 252 SNPs deviated from Hardy–Weinberg equilibrium (HWE) in Caucasian control study subjects ($P < 1 \times 10^{-3}$) 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 (pvalue.type = 'dost') from R package *HardyWeinberg*. Markers deviating from HWE at $P < 1 \times 10^{-3}$ 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 λ 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-specific 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 $P < 1 \times 10^{-3}$ 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 r^2 of 0.1. Population-specific genome-wide significance thresholds were based on Bonferroni correction ($P = 3.5 \times 10^{-6}$ to 1 d.p.).

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

Genomic coordinates for all available miRNA target site predictions made using 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/>)]

home.do)] were intersected with hg19 coordinates for HNSCC-associated SNPs (29,33). Low confidence interactions (mirSVR score > -0.1) were omitted (34). Three online databases; PolymiRTsv3.0 (25), miRNASNP (28) and MirSNP (35) were mined to identify miRNA target sites that may be disrupted by HNSCC-associated loci. PolymiRTsv3.0 utilizes TargetScan context+ scores to rank predicted miRNA target sites. Context+ scores predict the binding of a miRNA over a 3'UTR. Differences between context+ scores calculated using the reference and derived alleles for SNPs are provided by PolymiRTsv3.0. More negative context+ score differences suggest an increased likelihood that a variant disrupts miRNA targeting. Percentile ranks for context+ score differences were calculated based on all available scores. Lower percentile ranks indicate interactions more likely to be disrupted by genetic variation. The miRNASNP database (<http://www.bioguo.org/miRNASNP/index.php>) invokes the TargetScan and miRanda algorithms to make predictions of SNP effects on miRNA binding (28). MirSNP (<http://bioinfo.bjmu.edu.cn/mirsnp/search/>) utilizes the miRanda algorithm for miRNA target site prediction to determine the likely consequences of a given SNP (decrease/break or enhance/create miRNA target sites) (35). All available predictions were downloaded from each database and queried for the SNPs associated with HNSCC in this study.

TCGA miRNA expression analyses

miRNA isoform expression quantification from miRNA-seq analyses of The Cancer Genome Atlas (TCGA) HNSCC normal ($n = 44$) and primary tumor tissue ($n = 523$) samples was downloaded from the Genomic Data Commons Data Portal. To obtain strand specific miRNA expression for genes producing two miRNAs from the same precursor, miRBase accessions were used to determine expression of the relevant miRNA isoform. Average miRNA expression (miRNA mapped reads per million) across all samples was determined and restricted to miRNAs predicted to target genomic loci overlapping SNPs associated with HNSCC risk in this study. Percentile scores were calculated for the \log_2 transformed average gene expression values. miRNA-mRNA interactions identified from in silico analyses for which miRNA expression across all normal and tumor tissue was not available in isoform expression quantification files were omitted.

Code availability

R Code used to perform the analyses presented in this study will be made available upon request to the authors.

Data availability

Submission of genotypes and phenotypes to dbGAP is in process and an accession number was not available at the time of going to press.

Results

Discovery phase associations of miR-SNPs with HNSCC risk

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 $\geq 5\%$, 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, CSNK1A1, 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 $P < 1 \times 10^{-3}$, 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 5F, 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

Table 1. Clinical and pathological characteristics of study subjects

	Controls, n (%)			Cases, n (%)		
	Massachusetts study	M.D. Anderson study	Total	Massachusetts study	M.D. Anderson study	Total
	n = 1051	n = 1356	n = 2407	n = 904	n = 1338	n = 2242
Age at diagnosis						
≤50	169 (16.1)	441 (32.5)	610 (25.3)	189 (20.9)	366 (27.4)	555 (24.8)
>50 to ≤60	345 (32.8)	468 (34.5)	813 (33.8)	320 (35.4)	486 (36.3)	806 (36)
>60 to ≤70	331 (31.5)	316 (23.3)	647 (26.9)	251 (27.8)	334 (25)	585 (26.1)
>70	206 (19.6)	131 (9.7)	337 (14)	144 (15.9)	152 (11.4)	296 (13.2)
Sex						
Female	276 (26.3)	360 (26.5)	636 (26.4)	233 (25.8)	315 (23.5)	548 (24.4)
Male	775 (73.7)	996 (73.5)	1771 (73.6)	671 (74.2)	1023 (76.5)	1694 (75.6)
Race						
Caucasian	943 (89.7)	1160 (85.5)	2103 (87.4)	847 (93.7)	1212 (90.6)	2059 (91.8)
Other	108 (10.3)	196 (14.5)	304 (12.6)	57 (6.3)	126 (9.4)	183 (8.2)
Smoking status						
Current	129 (12.3)	165 (12.2)	294 (12.2)	102 (11.3)	493 (36.8)	595 (26.5)
Former	489 (46.5)	405 (29.9)	894 (37.1)	539 (59.6)	437 (32.7)	976 (43.5)
Never	419 (39.9)	573 (42.3)	992 (41.2)	223 (24.7)	404 (30.2)	627 (28)
Missing	14 (1.3)	213 (15.7)	227 (9.4)	40 (4.4)	4 (0.3)	44 (2)
HPV16, 18, 33, 51 positivity*						
Positive	360 (34.3)			564 (62.4)		
Negative	691 (65.7)			340 (37.6)		
Tumor stage						
Stages I and II				243 (26.9)	331 (24.7)	574 (25.6)
Stages III and IV				604 (66.8)	1007 (75.3)	1611 (71.9)
Missing				57 (6.3)	0 (0)	57 (2.5)
Tumor site						
Oral cavity				328 (36.3)	384 (28.7)	712 (31.8)
Pharynx				431 (47.7)	734 (54.9)	1165 (52)
Larynx				145 (16)	220 (16.4)	365 (16.3)

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

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 in known cancer risk locus 5p15.33 and was included on the genotyping array as part of a set of known disease-associated loci curated from the National Human Genome Research Institute (NHGRI) GWAS catalog. Similar results were observed in sensitivity analyses of overall and site-specific HNSCC risk restricted to Caucasian subjects (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 in *KDM8*

at rs1134367, and six miRNAs were predicted to target *HADH* at rs221347, 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.

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

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 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 rs1134367 (Table 4). The effects of rs1134367 on miR-149-3p, miR-4728-5p and miR-6883-5p binding to *KDM8* were concordant with predicted effects from the miRNASNP and MirSNP databases and all three miRNAs were among the most abundantly expressed in either normal head and neck tissue, or primary

Table 2. Summary statistics for miR-SNPs associated with HNSCC susceptibility

SNP	Chr:pos ^c	Associated gene ^d	Genomic context ^e	EAF (cases/controls) ^f	Discovery phase ^a		Validation phase ^b		Joint analysis ^b	
					OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Overall HNSCC										
rs7834169	8:26910291	MIR548H4	Intergenic	0.07/0.11	0.60 (0.46–0.78)	1.3E–04	0.80 (0.63–1.00)	4.8E–02	0.70 (0.59–0.82)	1.7E–05
Oral cavity										
rs16914640	12:9822387	CLEC2D	Missense	0.19/0.14	1.70 (1.27–2.27)	3.6E–04	1.46 (1.12–1.91)	5.7E–03	1.60 (1.32–1.94)	1.5E–06
rs7834169	8:26910291	MIR548H4	Intergenic	0.07/0.11	0.49 (0.33–0.73)	3.5E–04	0.68 (0.48–0.96)	2.8E–02	0.57 (0.44–0.73)	1.5E–05
rs1134367	16:27232883	KDM8	3'UTR	0.46/0.51	0.58 (0.43–0.77)	2.0E–04	0.74 (0.56–0.97)	2.9E–02	0.68 (0.56–0.82)	9.5E–05
rs1373756	18:49786306	DCC	Intergenic	0.08/0.12	0.49 (0.32–0.73)	5.9E–04	0.68 (0.48–0.96)	2.7E–02	0.61 (0.47–0.79)	1.6E–04
rs7306991	12:9810451	LOC374443	ncRNA	0.18/0.13	1.63 (1.22–2.18)	8.7E–04	1.35 (1.02–1.77)	3.5E–02	1.45 (1.20–1.76)	1.6E–04
Laryngeal										
rs221347	4:108955622	HADH	3'UTR	0.14/0.10	2.74 (1.79–4.19)	3.2E–06	1.46 (0.97–2.19)	7.0E–02	1.96 (1.48–2.61)	2.9E–06
rs4975616	5:1315660	CLPTM1L	Intergenic	0.37/0.45	0.52 (0.35–0.76)	8.1E–04	0.76 (0.54–1.07)	1.2E–01	0.64 (0.50–0.82)	3.8E–04

^aAnalysis adjusted for age at diagnosis, sex, race, HPV seropositivity (16, 11, 33 or 55), tobacco use (lifetime pack-years smoked) and alcohol consumption (average drinks per week).

^bAnalysis adjusted for age at diagnosis, sex, race and smoking status (never, former and current).

^cchr, chromosome; pos, position, according to NCBI Genome Build 37 (hg19).

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

^eUTR, untranslated region; ncRNA, non-coding RNA variant.

^fEAF, Effect allele frequencies calculated using all study subjects.

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

SNP	Accession ^a	Gene ^b	Allele ^c	miRNA ^d	mirSVR score	miRNA expression percentile ^e	
						Normal tissue	Tumor tissue
rs16914640	AK310554	CLEC2D	C	miR-3166	–0.142	0.42	0.37
	BC063128	CLEC2D	C	miR-3166	–0.134	0.42	0.37
rs1134367	NM_001145348	KDM8	T	miR-3136	–0.246	0.51	0.66
	AK308499	KDM8	T	miR-3136	–0.168	0.51	0.66
rs7306991	AY084051	LOC374443	A	miR-3148	–0.296	0.01	0.19
	AY084051	LOC374443	A	miR-605	–0.232	0.62	0.56
	NR_002814	LOC374443	A	miR-3148	–0.214	0.01	0.19
	NR_002814	LOC374443	A	miR-605	–0.165	0.62	0.56
	AY084051	LOC374443	A	miR-619	–0.127		0.12
	AY084051	LOC374443	A	miR-875-3p	–0.114		0.67
rs221347	NM_005327	HADH	T	miR-548g	–1.131		0.43
	NM_005327	HADH	T	miR-3065-5p	–0.910	0.78	0.82
	NM_005327	HADH	T	miR-105	–0.134	0.73	0.92
	NM_005327	HADH	T	miR-548a-3p	–0.107		<0.01
	NM_005327	HADH	T	miR-548e	–0.106	0.59	0.62
	NM_005327	HADH	T	miR-548f	–0.106	0.06	0.78

^aRefSeq, NCBI or GenBank Transcript/sequence identifiers.

^bGene location of variant according to NCBI Genome Build 37 (hg19).

^cAllele of SNP associated with miRNA–mRNA interaction.

^dmiRBase IDs (www.mirbase.org) of miRs known or predicted to target-associated gene.

^eTCGA-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.

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

SNP	Accession ^a	Gene ^b	Allele ^c	miRNA ^d	Context + score difference (percentile) ^e	Predicted SNP effect (miRNASNP)	Predicted SNP effect (MirSNP)	miRNA expression (percentile) ^f	
								Normal tissue	Tumor tissue
rs1134367	NM_001145348	KDM8	C	miR-149-3p	0.02	Gain	Create	0.78	0.78
	NM_024773	KDM8	C	miR-149-3p	0.02	Gain	Create	0.78	0.78
	NM_001145348	KDM8	C	miR-30b-3p	0.39	Gain	Create	0.79	0.79
	NM_024773	KDM8	C	miR-30b-3p	0.39	Gain	Create	0.79	0.79
	NM_001145348	KDM8	C	miR-3689c	0.42	Gain	Create		0.31
	NM_024773	KDM8	C	miR-3689c	0.42	Gain	Create		0.31
	NM_001145348	KDM8	T	miR-4779		Gain		0.07	0.27
	NM_024773	KDM8	T	miR-4779		Gain	Create	0.07	0.27
	NM_001145348	KDM8	T	miR-4536-5p	0.28	Gain	Create	0.59	0.12
	NM_024773	KDM8	C	miR-4536-5p	0.28	Loss	Break	0.59	0.12
	NM_001145348	KDM8	C	miR-4728-5p	0.02	Gain	Create		0.49
	NM_024773	KDM8	C	miR-4728-5p	0.02	Gain	Create		0.49
	NM_001145348	KDM8	T	miR-4695-5p	0.42			0.59	0.44
	NM_001145348	KDM8	C	miR-541-3p	0.33			0.77	0.77
	NM_001145348	KDM8	C	miR-654-5p	0.34			0.74	0.72
	NM_001145348	KDM8	C	miR-6769a-5p	0.28				0.17
	NM_001145348	KDM8	C	miR-6769b-5p	0.27			0.36	0.32
	NM_001145348	KDM8	C	miR-6778-5p	0.22			0.08	0.59
	NM_001145348	KDM8	C	miR-6779-5p	0.39			0.18	0.13
	NM_001145348	KDM8	C	miR-6780a-5p	0.39			0.05	0.19
	NM_001145348	KDM8	C	miR-6785-5p	0.03			0.03	0.09
	NM_001145348	KDM8	C	miR-6799-5p	0.37				0.13
	NM_001145348	KDM8	C	miR-6883-5p	0.02				0.74
NM_001145348	KDM8	C	miR-7106-5p	0.36				0.26	
NM_001145348	KDM8	C	miR-92a-2-5p	0.40				0.60	
rs221347	NM_001184705	HADH	T	miR-4729			Decrease		0.43
	NM_001184705	HADH	T	miR-548g-3p	0.07		Create		<0.01
	NM_005327	HADH	T	miR-548g-3p	0.07		Create		<0.01
	NM_005327	HADH	T	miR-4729			Decrease		0.43

^aRefSeq, NCBI or GenBank Transcript/sequence identifiers.

^bGene location of variant according to NCBI Genome Build 37 (hg19).

^cAllele of SNP associated with miRNA–mRNA interaction.

^dmiRBase IDs (www.mirbase.org) of miRNAs known or predicted to target-associated gene.

^eContext+ score differences were curated from the PolymiRTS database. Percentiles were calculated using the entire distribution of context+ score differences available from PolymiRTS.

^fTCGA-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 GWA 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 supports 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

- Siegel, R.L. et al. (2016) Cancer statistics, 2016. *CA. Cancer J. Clin.*, 66, 7–30.
- Sanderson, R.J. et al. (2002) Squamous cell carcinomas of the head and neck. *BMJ*, 325, 822–827.
- Furniss, C.S. et al. (2007) Human papillomavirus 16 and head and neck squamous cell carcinoma. *Int. J. Cancer*, 120, 2386–2392.
- Blot, W.J. et al. (1988) Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res.*, 48, 3282–3287.
- Edefonti, V. et al. (2012) Nutrient-based dietary patterns and the risk of head and neck cancer: a pooled analysis in the International Head and Neck Cancer Epidemiology consortium. *Ann. Oncol.*, 23, 1869–1880.
- Peters, E.S. et al. (2008) Dairy products, leanness, and head and neck squamous cell carcinoma. *Head Neck*, 30, 1193–1205.
- Galeone, C. et al. (2010) Coffee and tea intake and risk of head and neck cancer: pooled analysis in the international head and neck cancer epidemiology consortium. *Cancer Epidemiol. Biomarkers Prev.*, 19, 1723–1736.
- Worsham, M.J. (2011) Identifying the risk factors for late-stage head and neck cancer. *Expert Rev. Anticancer Ther.*, 11, 1321–1325.
- Lacko, M. et al. (2014) Genetic susceptibility to head and neck squamous cell carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.*, 89, 38–48.
- Peters, E.S. et al. (2006) Glutathione S-transferase polymorphisms and the synergy of alcohol and tobacco in oral, pharyngeal, and laryngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, 15, 2196–2202.
- Marsit, C.J. et al. (2008) A genotype-phenotype examination of cyclin D1 on risk and outcome of squamous cell carcinoma of the head and neck. *Clin. Cancer Res.*, 14, 2371–2377.
- McKay, J.D. et al. (2011) A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet.*, 7, e1001333.
- Lesseur, C. et al. (2016) Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat. Genet.*, 48, 1544–1550.
- Khurana, E. et al. (2016) Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.*, 17, 93–108.
- Ryan, B.M. et al. (2010) Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer*, 10, 389–402.
- Richardson, K. et al. (2011) A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. *BMC Genomics*, 12, 504.
- Niu, Y.M. et al. (2015) Significant association between functional microRNA polymorphisms and head and neck cancer susceptibility: a comprehensive meta-analysis. *Sci. Rep.*, 5, 12972.
- Song, X. et al. (2013) MicroRNA variants increase the risk of HPV-associated squamous cell carcinoma of the oropharynx in never smokers. *PLoS One*, 8, e56622.
- Ma, X.P. et al. (2013) Association between microRNA polymorphisms and cancer risk based on the findings of 66 case-control studies. *PLoS One*, 8, e79584.
- Guan, X. et al. (2013) A functional variant at the miR-885-5p binding site of CASP3 confers risk of both index and second primary malignancies in patients with head and neck cancer. *FASEB J.*, 27, 1404–1412.
- Liu, Z. et al. (2011) A functional variant at the miR-184 binding site in TNFAIP2 and risk of squamous cell carcinoma of the head and neck. *Carcinogenesis*, 32, 1668–1674.
- Liu, Z. et al. (2010) Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck. *Cancer*, 116, 4753–4760.
- Christensen, B.C. et al. (2010) Mature microRNA sequence polymorphism in MIR196A2 is associated with risk and prognosis of head and neck cancer. *Clin. Cancer Res.*, 16, 3713–3720.
- Li, G. et al. (2004) Association of a p73 exon 2 G4C14-to-A4T14 polymorphism with risk of squamous cell carcinoma of the head and neck. *Carcinogenesis*, 25, 1911–1916.
- Bhattacharya, A. et al. (2014) PolymiRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.*, 42, D86–D91.
- Schmeier, S. et al. (2011) dPORE-miRNA: polymorphic regulation of microRNA genes. *PLoS One*, 6, e16657.

27. Hiard, S. et al. (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res.*, 38, D640–D651.
28. Gong, J. et al. (2015) An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Database (Oxford)*, 2015, bav029.
29. Betel, D. et al. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, 36, D149–D153.
30. Thomas, L.F. et al. (2011) Inferring causative variants in microRNA target sites. *Nucleic Acids Res.*, 39, e109.
31. Aulchenko, Y.S. et al. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23, 1294–1296.
32. Skol, A.D. et al. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, 38, 209–213.
33. Betel, D. et al. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, 11, R90.
34. Cancer Genome Atlas, N. (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 517, 576–82.
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.
36. Rosen, D.B. et al. (2008) Functional consequences of interactions between human NKR-P1A and its ligand LLT1 expressed on activated dendritic cells and B cells. *J. Immunol.*, 180, 6508–6517.
37. Roth, P. et al. (2007) Malignant glioma cells counteract antitumor immune responses through expression of lectin-like transcript-1. *Cancer Res.*, 67, 3540–3544.
38. Karreth, F.A. et al. (2013) ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.*, 3, 1113–1121.
39. Karreth, F.A. et al. (2015) The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell*, 161, 319–332.
40. Ishimura, A. et al. (2012) Jmjd5, an H3K36me2 histone demethylase, modulates embryonic cell proliferation through the regulation of Cdkn1a expression. *Development*, 139, 749–759.
41. Hsia, D.A. et al. (2010) KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc. Natl. Acad. Sci. USA*, 107, 9671–9676.
42. Zhao, Z. et al. (2015) Overexpression of histone demethylase JMJD5 promotes metastasis and indicates a poor prognosis in breast cancer. *Int. J. Clin. Exp. Pathol.*, 8, 10325–10334.
43. Amendola, P.G. et al. (2017) JMJD-5/KDM8 regulates H3K36me2 and is required for late steps of homologous recombination and genome integrity. *PLoS Genet.*, 13, e1006632.
44. Mancuso, M. et al. (2009) H3K4 histone methylation in oral squamous cell carcinoma. *Acta Biochim. Pol.*, 56, 405–410.
45. Eaton, S. et al. (2000) The mitochondrial trifunctional protein: centre of a beta-oxidation metabolon? *Biochem. Soc. Trans.*, 28, 177–182.
46. Martins, E. et al. (2011) Short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency: the clinical relevance of an early diagnosis and report of four new cases. *J. Inherit. Metab. Dis.*, 34, 835–842.
47. Zhao, Z. et al. (2016) Prognostic significance of two lipid metabolism enzymes, HADHA and ACAT2, in clear cell renal cell carcinoma. *Tumour Biol.*, 37, 8121–8130.
48. Kageyama, T. et al. (2011) HADHA is a potential predictor of response to platinum-based chemotherapy for lung cancer. *Asian Pac. J. Cancer Prev.*, 12, 3457–3463.
49. Currie, E. et al. (2013) Cellular fatty acid metabolism and cancer. *Cell Metab.*, 18, 153–161.
50. Kim, D. et al. (2016) General rules for functional microRNA targeting. *Nat. Genet.*, 48, 1517–1526.