### Title

Smoking status regulates a novel panel of PIWI-interacting RNAs in head and neck squamous cell carcinoma.




### Authors

Krishnan, Aswini R
Korrapati, Avinaash
Zou, Angela E
et al.



### Publication Date

2017-02-01

### DOI

10.1016/j.oraloncology.2016.12.022

Peer reviewed

HHS Public Access

Author manuscript

*Oral Oncol*. Author manuscript; available in PMC 2018 February 01.

Published in final edited form as:

*Oral Oncol*. 2017 February ; 65: 68–75. doi:10.1016/j.oraloncology.2016.12.022.

# Smoking status regulates a novel panel of PIWI-interacting RNAs in head and neck squamous cell carcinoma

**Aswini R. Krishnan**[1], **Avinaash Korrapati**[1], **Angela E. Zou**[1], **Yuanhao Qu**[1], **Xiao Qi Wang**[2], **Joseph A. Califano**[1], **Jessica Wang-Rodriguez**[3], **Scott M. Lippman**[4], **Melbourne F. Hovell**[5], and **Weg M. Ongkeko**[1,†]

[1]Department of Otolaryngology-Head and Neck Surgery, University of California, San Diego, La Jolla, California, United States of America

[2]Department of Surgery, The University of Hong Kong, Pokfulam, Hong Kong, China

[3]Veterans Administration Medical Center and Department of Pathology, University of California San Diego, La Jolla, California, United States of America

[4]Department of Medicine and Moores Cancer Center, University of California, San Diego, La Jolla, California, United States of America

[5]Graduate School of Public Health, San Diego State University, San Diego, California, United States of America

## Abstract

**Objective—**Smoking remains a primary etiological factor in head and neck squamous cell carcinoma (HNSCC). Given that non-coding RNAs (ncRNAs), including PIWI-interacting RNAs (piRNAs), have emerged as mediators of initiation and progression in head and neck malignancies, we undertook a global study of piRNA expression patterns in smoking-associated HNSCC.

**Materials and Methods—**Using RNA-sequencing data from 256 current smoker and lifelong nonsmoker samples in The Cancer Genome Atlas (TCGA), we analyzed the differential expression patterns of 27,127 piRNAs across patient cohorts stratified by tobacco use, with HPV16 status and tumor status taken into account. We correlated their expression to clinical characteristics and to smoking-induced alteration of PIWI proteins, the functional counterparts of piRNAs. Finally, we correlated our identified piRNAs and PIWI proteins to known chromosomal aberrations in HNSCC to understand their wider-ranging genomic effects.

**Results and Conclusion—**Our analyses implicated a 13-member piRNA panel in smoking-related HNSCC, among which *NONHSAT123636* and *NONHSAT113708* are associated with tumor stage, *NONHSAT067200* with patient survival, and *NONHSAT081250* with smoking-

†Postal Address of Corresponding Author: Weg Ongkeko, VA Medical Center San Diego, 3350 La Jolla Village Dr., West Wing Rm 6029, San Diego, CA 92161.



**Conflict of Interest Statement**

The authors have no conflict of interest.

altered *PIWIL1* protein expression. 6 piRNAs as well as *PIWIL1* correlated with genomic alterations common to HNSCC, including *TP53* mutation, *TP53-3p* co-occurrence, and *3q26, 8q24*, and *11q13* amplification. Collectively, our findings provide novel insights into the etiology-specific piRNA landscape of smoking-induced HNSCC.

### Keywords

Smoking; head and neck neoplasms; RNA; small interfering

## Introduction

Head and neck squamous cell carcinomas (HNSCCs) represent the sixth deadliest cancer worldwide [1, 2]. Despite advances in diagnosis and treatment, HNSCC patient mortality has improved minimally over the past three decades [3]. HNSCC risk factors include tobacco smoking, alcohol consumption, and more recently, HPV infection [4, 5]; however, tobacco remains the most potent and longstanding etiological agent, increasing disease risk by up to 25-fold [6, 7].

Despite the prevalence of smoking-induced HNSCC, current understanding of its pathogenesis and progression remains limited. One reason is that HNSCCs exhibit a wide range of molecular profiles based on anatomic site, tumor stage, and etiology, making it difficult to comprehensively study molecular alterations in this disease [8, 9]. Previously, studies have shown that tobacco smoke induces carcinogenesis through genotoxic components, such as nitrosamines and polycyclic aromatic hydrocarbons, that form DNA adducts and activate pathways including AKT and PKA [2, 10, 11]. Tobacco also induces loss-of-function mutations in *TP53, CDKN2A, FAT1*, and *PIK3CA*, along with copy number aberrations in *3q26/28* and *11q13/22* [2, 9, 12]. Meanwhile, studies exploring the role of epigenetic interactions in HNSCC have identified distinct methylation patterns in tumor clusters based on patient smoking status [13–15]. Taken together, these findings suggest that it is crucial to understand the interplay between genetic, epigenetic, and transcriptional alterations in order to garner a comprehensive view of the mechanisms of smoking-induced HNSCC pathogenesis.

Non-coding RNAs (ncRNAs) have been gaining recognition for their involvement in genetic and epigenetic regulation [16, 17], with recent findings implicating ncRNAs in a number of cancers [18–21]. In HNSCC, studies have documented the dysregulation of long ncRNAs (*HOTAIR, MALAT1, MEG3, NEAT1*), microRNAs (*miR-21/miR-181b/miR-345, miR-221/miR-375*, and *miR-200a/miR-125a*), small nucleolar RNAs (*SNORD116-20, SNORA14B, SNORD35B*), and PIWI-interacting RNAs (*piR-34736, piR-49145, piR-57125, piR-35469/piR-35468/piR-35467/piR-35466*) in tumor vs. normal tissues [22–25]. However, the functional relevance and etiological association of these dysregulations remain largely unexplored, including knowledge of which ncRNAs specifically are altered by smoking. This is critical as the heterogeneity of HNSCC makes it important to understand ncRNA dysregulation by HNSCC subtype.

As the largest class of small ncRNA molecules, PIWI-interacting RNAs (piRNAs) contribute to various biological processes, from germline and somatic development to chromosome

rearrangement and gene deregulation [26–28]. piRNAs are thought to function via the PIWI subfamily of Argonaute proteins (PIWL1-4), forming piRNA/PIWI complexes that promote transposon silencing and transcriptional regulation [29, 30]. Our previous study of general piRNA dysregulation in HNSCC found that select piRNAs (namely *piR-34736*) correlate with tobacco use, suggesting that a specific panel of piRNAs may be linked to smoking-induced HNSCC [25]. The extent of these findings, however, remains significantly limited, as there is currently no knowledge of piRNA expression and function in a purely tobacco-specific context. We therefore aimed to comprehensively characterize piRNA dysregulation in smoking-induced HNSCCs. Using RNA-sequencing data from 256 current smoker and lifelong nonsmoking datasets from The Cancer Genome Atlas, we identified 13 piRNAs differentially expressed in smoking-associated HNSCCs and analyzed their clinical relevance. We then identified similarly dysregulated PIWI proteins and correlated their expression to individual piRNAs. Finally, we evaluated the associations of our piRNAs and PIWI proteins to mutations and copy number variations (CNVs) in order to develop a more thorough molecular understanding of smoking-induced HNSCC.

## Results

### Identification of dysregulated piRNAs in smoking-related HNSCC

In order to identify piRNAs specifically dysregulated in smoking-induced HNSCCs, we examined 256 current smoker and lifelong nonsmoker head and neck RNA-seq datasets from TCGA with available clinical data (dataset IDs in Supplementary Table 1). Datasets were divided into 6 cohorts based on smoking vs. nonsmoking status, with HNSCC vs. adjacent normal tissue classification and HPV16(+) vs. HPV(−) status accounted for : (1) HPV(−) HNSCC Smoker, (2) HPV(−) HNSCC Nonsmoker, (3) HPV16(+) HNSCC Smoker, (4) HPV16(+) HNSCC Nonsmoker, (5) HPV(−) Normal Nonsmoker (Figure 1a).

To identify smoking-dysregulated piRNAs, we used negative binomial-based differential expression testing to compare piRNA expression between current smoking and lifelong nonsmoking cohorts, with HPV16(+) and HPV(−) datasets analyzed separately to minimize the confounding effects of HPV status on smoking-related HNSCC pathogenesis. We found 58 piRNAs dysregulated in the HPV(−) HNSCC Smoker vs. HPV(−) Normal Nonsmoker comparison, 8 piRNAs dysregulated in the HPV(−) HNSCC Smoker vs. HPV(−) HNSCC Nonsmoker comparison, and 6 piRNAs dysregulated in the HPV16(+) HNSCC Smoker vs. HPV16(+) HNSCC Nonsmoker comparison (*FDR* < 0.05; Supplementary Table 2; Figure 1a).

We focused on characterizing the 58 piRNAs dysregulated between HPV(−) HNSCC Smoker and HPV(−) Normal Nonsmoker patients, as these were most likely to be specifically involved in smoking-induced HNSCC pathogenesis. We compared the piRNAs found to be generally dysregulated in HNSCCs to these 58 piRNAs and found both overlapping and unique piRNAs between the comparisons, suggesting that HNSCCs preserve a heterogeneity that must be separately evaluated by etiology (Figure 1b) [25]. We performed additional analyses on these transcripts to more thoroughly investigate piRNA involvement in smoking-induced HNSCC pathogenesis and progression.

### Determination of clinically significant smoking-dysregulated piRNAs

First, we assessed all 58 piRNAs dysregulated between HPV(−) HNSCC Smoker and HPV(−) Normal Nonsmoker patients for clinical and functional potential in smoking-induced HNSCC by evaluating their expression relative to a range of HNSCC clinical features, including tumor anatomic site, tumor stage, and metastatic indicators. Analyses were performed using piRNA expression values of all HPV(−) HNSCC Smokers in order to minimize confounding clinical variables in the patient cohort and directly evaluate the clinical relevance of the candidate piRNAs in smoking-related HNSCC.

Using the Kruskal-Wallis test ($p < 0.05$), we found that *NONHSAT123636*, transcript piRNA downregulated in smoking-related HNSCCs, exhibited lowered expression with increasing clinical stage and with lymphovascular invasion, while upregulated piRNAs *NONHSAT113708* and *NONHSAT015828* demonstrated elevated expression in patients with higher clinical stages. *NONHSAT081250*, downregulated in smoking-induced HNSCCs, displayed reduced expression with greater nodal extracapsular spread and variable expression in HPV(−) Nonsmoking HNSCCs of different anatomical sites (Figure 2a – 2d).

We then analyzed all 58 candidate piRNAs to assess the relationship between piRNA expression level and overall survival among all HPV(−) Smoking HNSCC patients with available data (n=140). Modeling piRNA expression as a binary variable (high/low), we found that low expression of *NONHSAT067200*, a piRNA downregulated in smoking-induced HNSCCs, was significantly associated with poor survival in patients ($p$=0.0361, hazard ratio [HR] = 1.897, Figure 2e). We also performed multivariate Cox regression on *NONHSAT067200* and found that association with patient outcome was independent of prognostic factors in HNSCC such as age, gender, clinical stage, and tumor grade (Table 1).

We present our five piRNAs displaying clinical correlations (*NONHSAT123636, NONHSAT113708, NONHSAT015828, NONHSAT081250*, and *NONHSAT067200*) in Table 2a.

### Determination of commonly dysregulated piRNA transcripts in smoking-induced HNSCC

Next, we further analyzed the 58 dysregulated piRNAs (Figure 1c) by identifying transcripts displaying common dysregulation in multiple Smoker vs. Nonsmoker comparisons. Our analyses revealed a panel of 8 commonly dysregulated transcripts, presented in Table 2b, which are likely to exhibit general involvement in smoking-induced HNSCC pathogenesis.

### PIWI Protein dysregulation in HPV-related HNSCC and Smoking-related HNSCC

We then evaluated the dysregulation of *PIWIL1, PIWIL2, PIWIL3*, and *PIWIL4* mRNA in smoking-induced HNSCC, utilizing all available TCGA head and neck mRNA datasets for current smoking and lifelong nonsmoking patients (dataset IDs in Supplementary Table 3). We performed negative binomial-based differential expression testing between the same patient cohorts as initially used to evaluate piRNA dysregulation. We filtered these differential expression data to only retain PIWIL genes significantly dysregulated in the HPV(−) Cancer Smoker vs. HPV(−) Normal Nonsmoker comparison and verified that they exhibited consistent differential expression, if applicable, in additional smoker vs.

nonsmoker comparisons ($p < 0.05$, *FDR* < 0.05, Supplementary Table 4). Our analyses revealed *PIWIL1* to be upregulated by 9.55 fold ($p = 0.0035$, *FDR* = 0.014) in the HPV(−) Cancer Smoker vs. HPV(−) Normal Nonsmoker comparison and 5.30 fold ($p = 0.010$, *FDR* = 0.015) in the HPV(+) Cancer Smoker vs. HPV(+) Cancer Nonsmoker comparison.

### Correlation of individual piRNAs to PIWIL1 upregulation

To investigate the potential molecular roles of individual piRNAs in relation to *PIWIL1*, we correlated the expression of each of 58 piRNAs to the expression of *PIWIL1* in all HPV(−) HNSCC Smoking patients. We categorized *PIWIL1* expression (in cpm) into intervals by powers of 10, and using Kruskal-Wallis tests ($p < 0.05$) found *NONHSAT081250*, a downregulated piRNA in smoking-induced HNSCC, to exhibit significantly reduced expression in patients with increased expression of *PIWIL1* (Figure 3a).

### Association of smoking-dysregulated piRNAs and *PIWIL1* with known genomic alterations in HNSCC

Existing studies propose that piRNA/PIWI complexes mediate epigenetic regulation of the genome, altering gene expression and the behavior of transposable elements [26–28]. In order to more thoroughly understand the role of piRNAs and PIWI proteins in these processes, we analyzed the association of all 13 key smoking-dysregulated piRNAs (Table 2) and *PIWIL1* with frequent somatic mutations and copy number variations documented in HNSCC (Supplementary Table 5a/5b). Analyses were performed using data from all HPV(−) HNSCC Smoking patients, with mutation and copy number calls obtained from The Broad Institute Firehose database.

We employed the Wilcoxon rank-sum test to correlate piRNA and PIWI protein expression values (cpm) to mutation and copy number status. Our analyses revealed that *NONHSAT108298* and *NONHSAT123636*, piRNAs downregulated in smoking-induced HNSCCs, exhibit significantly lower expression in the presence of *TP53* mutations compared to the wild type, while upregulated piRNAs *NONHSAT015828, NONHSAT077364*, and *NONHSAT077463* displayed elevated expression in the presence of mutated oncogenes ($p < 0.05$, Figure 4a). Amplification of *3q26, 8q24, 11q13*, and *5p15.33* and deletion of *10q23.31* and *17q25.4*, commonly observed aberrations in HNSCC, were significantly associated with reduced expression of *NONHSAT081250, NONHSAT108298, NONHSAT123636*, piRNAs underexpressed in smoking-induced HNSCCs ($p < 0.05$, Figures 4b–4c). In addition, HNSCC-associated CNVs of *10p15.3, 11p23.1*, and *13q22.1* correlated not only with lowered expression of piRNAs *NONHSAT108298* and *NONHSAT081250*, but also with elevated expression of *PIWIL1* ($p < 0.05$, Figure 4d, Figure 3b).

Recent studies indicate that *TP53* mutation is frequently accompanied by *3p* chromosomal loss in HNSCC, and that co-occurrence of these alterations associates with a decrease in HNSCC survival rates [31]. We therefore attempted to identify whether our 13 smoking-dysregulated piRNAs and *PIWIL1* correlated with combined *TP53-3p* alteration. We found that *NONHSAT081250* and *NONHSAT108298*, transcripts downregulated in smoking-induced HNSCCs, displayed significantly lowered expression in the presence of *TP53-3p*

co-occurrence, suggesting an interaction between these piRNAs and alterations ($p < 0.05$, Figure 4e).

The remaining correlations of piRNA and *PIWIL1* to mutations and CNVs are presented in Table 3.

## Discussion

Smoking is the most prevalent cause of HNSCC, which remains a dangerous and intractable malignancy due to its vast clinical, etiological, and genetic heterogeneity [8, 9]. However, recent findings suggest that epigenetic factors, particularly ncRNA alterations, may reveal specific signatures of HNSCC subtypes and provide insights into the pathogenesis of distinct HNSCCs, including smoking-induced HNSCCs [23].

To our knowledge, we are the first to comprehensively investigate piRNA alterations in smoking-induced HNSCC. Past studies of smoking-dysregulated ncRNAs have focused primarily on miRNAs and lncRNAs, with virtually no investigation of smoking-associated piRNA alterations, possibly due to the difficulty of elucidating the mechanisms of piRNA-mediated genomic regulation and transformation [32].

Through RNA-seq analysis of 265 current smoker vs. lifelong nonsmoker datasets from The Cancer Genome Atlas, we identified 58 piRNAs significantly dysregulated between HPV(−) HNSCC Smoker and Normal Nonsmoker tissues ($p < 0.05$, $FDR < 0.05$). Taking into account the heterogeneity of HNSCC, we performed a multi-tiered analysis on all 58 transcripts to more thoroughly filter for those exhibiting greatest implication in this disease. First, we identified all piRNAs displaying clinical correlation in smoking-induced HNSCCs and found four transcripts (*NONHSAT113708, NONHSAT081250, NONHSAT123636*, and *NONHSAT015828*) associated with tumor stage, metastasis, and/or anatomic site, as well as one transcript (*NONHSAT067200*) associated with patient survival (Kruskal-Wallis and Cox regression analyses, all $p < 0.05$). These piRNAs may contribute to various clinical phenotypes and potentially serve as prognostic biomarkers for smoking-induced HNSCCs. Second, we identified eight piRNAs displaying directly consistent dysregulation among multiple smoking vs. nonsmoking patient cohorts. Despite not exhibiting the same clinical significance as the five transcripts identified previously, these piRNAs may be more generally involved in smoking-associated HNSCC pathogenesis, perhaps by promoting initial induction of the disease, and may potentially function as diagnostic biomarkers.

Our past investigations of ncRNA alterations in generic HNSCCs have indicated that four out of our five clinically significant piRNAs (*NONHSAT015828, NONHSAT081250, NONHSAT113708*, and *NONHSAT123636*) exhibit directionally consistent dysregulation between generic and smoking-induced HNSCCs [25]. On the other hand, among the eight piRNAs commonly dysregulated between multiple smoker vs. nonsmoker cohorts, only three transcripts (*NONHSAT077463, NONHSAT069719*, and *NONHSAT108298*) were mutually altered. Moreover, *NONHSAT108298* and *NONHSAT069719* displayed directionally inconsistent dysregulations in smoking-induced HNSCCs compared to HNSCCs in general. Cumulatively, these suggest that while some transcripts may be

commonly implicated between smoking-related and generic HNSCCs, smoking-induced HNSCCs still preserve a distinct piRNA landscape that must be analyzed separately from HNSCCs of different origin.

In addition, *NONHSAT105870* was previously identified to be implicated in breast cancer, suggesting the ability of the same piRNAs to modulate other malignancies [26]. To the best of our knowledge, four of our transcripts are completely novel, and all 13 remain uncharacterized in any malignancy beyond previous reports of their dysregulation. To further investigate the functional potential of these piRNAs, we assessed these piRNAs in the context of PIWI proteins. Using mRNA expression data for 253 patients from The Cancer Genome Atlas, we identified *PIWIL1* to be significantly upregulated in current smoker vs. lifelong nonsmoker samples, in accordance with previous studies identifying overexpression of this gene in other carcinomas ($p < 0.05$, $FDR < 0.05$) [33–36]. Correlations of our 13-member piRNA panel to *PIWIL1* indicated that *NONHSAT081250*, a piRNA downregulated in smoking-associated HNSCCs, exhibited significantly reduced expression with upregulation of *PIWIL1*, suggesting a potential mechanistic interaction between these elements ($p < 0.05$). As piRNA/PIWI complexes are known to mediate epigenetic and transcriptional alterations to the genome, we also associated our 13 piRNA candidates and *PIWIL1* with frequent chromosomal aberrations in HNSCC. Notably, we found CNVs in *11p23.1* and *13q22.1* to correlate not only with downregulation of NONHSAT081250 but also with upregulation of *PIWIL1*, further substantiating the presence of genomic interplay between the two transcripts.

In addition, *NONHSAT081250* and *NONHSAT123636* were found to associate with other frequent CNVs in HNSCC, including amplification of *3q26* and *8q24*, loci of the *PIK3CA* oncogene and the *c-MYC* proto-oncogene respectively [37]. Current theory suggests that these alterations occur early in cancer development, suggesting the potential of these piRNAs to regulate the preliminary stages of malignant transformation and serve as diagnostic or prognostic biomarkers [38–40]. Among a range of piRNA correlations with HNSCC-related somatic mutations, our analyses revealed *NONHSAT108298* and *NONHSAT123636* to associate with *TP53* mutations as well as with *TP53* mutation-3*p* deletion co-occurrence, a phenomenon associated with drastically reduced HNSCC patient survival [31]. Taken together, these findings reveal the potential of our piRNA candidates to assume a vast array of roles in smoking-related HNSCC induction and to modulate the effects of well-established HNSCC-associated genomic aberrations.

Despite our comprehensive evaluation of piRNA dysregulation and their PIWI protein and genomic interactions, some phenomena still remain unexplored. For example, *CDKN2A*, a gene near-universally mutated in smoking-induced HNSCCs, did not display correlations with either of our 13 piRNA candidates or *PIWIL1* [9]. *PIWL1* itself, shown to induce epigenetic silencing of *PTEN* in endometrial carcinomas, did not exhibit the same correlation in smoking-associated head and neck malignancies [41]. Finally, *piR-34376*, the transcript identified previously to correlate with smoking status in generic HNSCCs, was not significantly differentially expressed in our present analyses [25]. Cumulatively, these suggest that even etiology-specific HNSCCs remain diverse with multiple synergistic and

disparate molding factors, substantiating the heterogeneous nature of the disease and the necessity to further investigate its genetic and epigenetic mediators.

## Materials and Methods

### RNA-seq datasets and clinical data

MapSplice-aligned TCGA BAM files were obtained from the UCSC Cancer Genomics Hub (https://cghub.ucsc.edu/) on 11 March 2015. To investigate piRNA expression, we downloaded RNA-seq datasets for all 256 current smoker and lifelong nonsmoker head and neck tissue samples. The TCGA barcodes for all patients whose datasets were used in this study are provided in Supplementary Table 1.

Patient clinical data, including smoking status, were downloaded from the TCGA Data Portal on 20 March 2015 (https://tcga-data.nci.nih.gov/tcga/findArchives.htm).

### piRNA expression analysis

piRNA read counts were generated from sequencing datasets via BEDtools coverageBed (https://github.com/arq5x/bedtools2) using piRNA annotation files. The piRNA BED file containing 27,127 piRNA transcripts was obtained from NONCODEv4 (http://www.bioinfo.org/NONCODEv4/), a database integrating ncRNA data from RefSeq, Ensembl, and published literature. Read count tables were imported into edgeR v3.0 (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html), and lowly expressed piRNAs were filtered from the analysis. Following TMM normalization, pairwise comparisons were applied to identify significantly differentially expressed piRNAs between the Current Smoking and Lifelong Nonsmoking patient cohorts, as visualized in Figure 1a.

### Association of piRNA expression with clinical covariates and patient survival

All 58 piRNAs dysregulated between HPV(−) Smoking HNSCC and HPV(−) Nonsmoking Normal cohorts were evaluated for clinical significance. Employing the Kruskal-Wallis test, we investigated piRNA association with anatomic neoplasm subdivision, tumor stage and grade, lymphovascular and perineural invasion, and extracapsular nodal spread, using clinical data and piRNA expression values (cpm) from HPV(−) smoking HNSCC patients. Patients with no available information for a given characteristic were filtered from analyses involving that variable.

Survival analyses were performed on the 58 piRNAs using Cox proportional hazards models, with piRNA expression in tumors (cpm) modeled as a binary variable based on expression above or below the median. Because HPV status has been shown to profoundly influence molecular signatures and clinical outcomes in HNSCC, we limited our cohort to HPV(−) HNSCC Smokers to minimize confounding variables. We first performed univariate Kaplan-Meier analysis and univariate Cox regression analysis to identify candidates significantly associated with patient outcome ($p < 0.05$), and then performed multivariate Cox analysis to evaluate whether correlations were independent of clinical variables such as age (grouped into 10-year intervals), gender, and tumor grade and stage.

### Assessment of PIWI protein dysregulation

Level 3 mRNA expression data containing pre-generated read counts for all 253 available current smoker and lifelong nonsmoker head and neck tissue samples were obtained from the UCSC Cancer Genomics Hub (https://cghub.ucsc.edu/) on 16 June 2016 (dataset IDs in Supplementary Table 3). Read counts for *PIWIL1-4* were imported into edgeR v3.0 (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html). Following filtering of lowly expressed genes and TMM normalization, pairwise comparisons were applied to identify significantly differentially expressed PIWI proteins between the same Current Smoking and Lifelong Nonsmoking patient cohorts as used earlier to evaluate piRNA differential expression (Figure 1a).

We filtered this data to select all PIWI protein candidates dysregulated in the HPV(−) Smoking HNSCC vs. HPV(−) Nonsmoking Normal comparison that, if differentially expressed in other smoker vs. nonsmoker comparisons, exhibited directionally consistent dysregulation.

### Correlation of individual piRNA dysregulation to PIWI protein dysregulation

We assessed all 13 piRNAs showing significant clinical associations and common dysregulation between smoker and nonsmoker cohorts for correlation with dysregulated PIWI proteins. We grouped PIWI protein expression values (cpm) for each patient into exponential intervals increasing by powers of 10. Using the Kruskal-Wallis test, we evaluated the correlation between piRNA expression levels (cpm) and PIWI protein expression intervals.

### Association of ncRNA expression with tumor mutations and copy number aberrations

TCGA tumor mutation calls were obtained from mutation annotation files (maf) generated by the Broad Institute GDAC Firehose on 20 August 2016. We focused our analysis on 26 most frequently mutated genes in HNSCCs, as determined by whole exome sequencing of an independent tumor cohort by Stransky et al. [42] (Supplementary Table 5a). Wilcoxon rank sum tests were employed to test for significant associations between piRNAs and *PIWIL1* expression level (cpm) and mutational status.

Copy number variations for the TCGA tumors were obtained from the GISTIC2 pipeline in Firehose on 20 August 2016. Similarly, 73 significant (99% confidence) focal amplifications and deletions, along with all amplifications on *3q26, 8q24*, and *11q13* (the most frequent CNVs in HNSCC), were analyzed for correlation to piRNAs and *PIWIL1* expression level using Wilcoxon rank sum tests [43] (Supplementary Table 5b).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Funding**

This work was supported by funding from the National Institutes of Health, grant number DE023242 to W.M.O. and by Academic Senate grants from the University of California to W.M.O. and J.W.R.

**Role of the Funding Source**

Federally sponsored research to promote the understanding of tobacco-related head and neck cancer.

## Abbreviations

| | |
|---|---|
| **HNSCC** | Head and neck squamous cell carcinoma |
| **ncRNA** | Noncoding RNA |
| **piRNA** | PIWI-interacting RNA |

## References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010; 127:2893–2917. [PubMed: 21351269]
2. Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. Annu Rev Pathol. 2009; 4:49–70. [PubMed: 18729723]
3. Carvalho AL, Nishimoto IN, Califano JA, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. Int J Cancer. 2005; 114:806–816. [PubMed: 15609302]
4. Galbiatti AL, Padovani-Junior JA, Maniglia JV, Rodrigues CD, Pavarino EC, Goloni-Bertollo EM. Head and neck cancer: causes, prevention and treatment. Braz J Otorhinolaryngol. 2013; 79:239–247. [PubMed: 23670332]
5. Ragin CC, Modugno F, Gollin SM. The epidemiology and risk factors of head and neck cancer: a focus on human papillomavirus. J Dent Res. 2007; 86:104–114. [PubMed: 17251508]
6. Kim L, King T, Agulnik M. Head and neck cancer: changing epidemiology and public health implications. Oncology (Williston Park). 2010; 24:915–919. 24. [PubMed: 21138172]
7. Andre K, Schraub S, Mercier M, Bontemps P. Role of alcohol and tobacco in the aetiology of head and neck cancer: a case-control study in the Doubs region of France. Eur J Cancer B Oral Oncol. 1995; 31B:301–309. [PubMed: 8704646]
8. Suh Y, Amelio I, Guerrero Urbano T, Tavassoli M. Clinical update on cancer: molecular oncology of head and neck cancer. Cell Death Dis. 2014; 5:e1018. [PubMed: 24457962]
9. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015; 517:576–582. [PubMed: 25631445]
10. Phillips DH. DNA adducts in human tissues: biomarkers of exposure to carcinogens in tobacco smoke. Environ Health Perspect. 1996; 104(Suppl 3):453–458. [PubMed: 8781363]
11. Shih RH, Cheng SE, Hsiao LD, Kou YR, Yang CM. Cigarette smoke extract upregulates heme oxygenase-1 via PKC/NADPH oxidase/ROS/PDGFR/PI3K/Akt pathway in mouse brain endothelial cells. J Neuroinflammation. 2011; 8:104. [PubMed: 21861928]
12. Li H, Wawrose JS, Gooding WE, Garraway LA, Lui VW, Peyser ND, et al. Genomic analysis of head and neck squamous cell carcinoma cell lines and human tumors: a rational approach to preclinical model selection. Mol Cancer Res. 2014; 12:571–582. [PubMed: 24425785]
13. Hasegawa M, Nelson HH, Peters E, Ringstrom E, Posner M, Kelsey KT. Patterns of gene promoter methylation in squamous cell cancer of the head and neck. Oncogene. 2002; 21:4231–4236. [PubMed: 12082610]
14. Fertig EJ, Markovic A, Danilova LV, Gaykalova DA, Cope L, Chung CH, et al. Preferential activation of the hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. PLoS One. 2013; 8:e78127. [PubMed: 24223768]
15. Choudhury JH, Ghosh SK. Promoter Hypermethylation Profiling Identifies Subtypes of Head and Neck Cancer with Distinct Viral, Environmental, Genetic and Survival Characteristics. PLoS One. 2015; 10:e0129808. [PubMed: 26098903]

16. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006; 15(Spec No 1):R17–R29. [PubMed: 16651366]
17. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014; 157:77–94. [PubMed: 24679528]
18. Calore F, Lovat F, Garofalo M. Non-coding RNAs and cancer. Int J Mol Sci. 2013; 14:17085–17110. [PubMed: 23965974]
19. Huang T, Alvarez A, Hu B, Cheng SY. Noncoding RNAs in cancer and cancer stem cells. Chin J Cancer. 2013; 32:582–593. [PubMed: 24206916]
20. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015; 21:1253–1261. [PubMed: 26540387]
21. Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. Annu Rev Pathol. 2014; 9:287–314. [PubMed: 24079833]
22. Yang QQ, Deng YF. Long non-coding RNAs as novel biomarkers and therapeutic targets in head and neck cancers. Int J Clin Exp Pathol. 2014; 7:1286–1292. [PubMed: 24817925]
23. Denaro N, Merlano MC, Russi EG, Lo Nigro C. Non coding RNAs in head and neck squamous cell carcinoma (HNSCC): a clinical perspective. Anticancer Res. 2014; 34:6887–6896. [PubMed: 25503114]
24. Shiiba M, Uzawa K, Tanzawa H. MicroRNAs in Head and Neck Squamous Cell Carcinoma (HNSCC) and Oral Squamous Cell Carcinoma (OSCC). Cancers (Basel). 2010; 2:653–669. [PubMed: 24281087]
25. Zou AE, Zheng H, Saad MA, Rahimy M, Ku J, Kuo SZ, et al. The non-coding landscape of head and neck squamous cell carcinoma. Oncotarget. 2016
26. Ng KW, Anderson C, Marshall EA, Minatel BC, Enfield KS, Saprunoff HL, et al. Piwi-interacting RNAs in cancer: emerging functions and clinical utility. Mol Cancer. 2016; 15:5. [PubMed: 26768585]
27. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem. 2015; 84:405–433. [PubMed: 25747396]
28. Ku HY, Lin H. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. Natl Sci Rev. 2014; 1:205–218. [PubMed: 25512877]
29. Sasaki T, Shiohama A, Minoshima S, Shimizu N. Identification of eight members of the Argonaute family in the human genome. Genomics. 2003; 82:323–330. [PubMed: 12906857]
30. Meister G. Argonaute proteins: functional insights and emerging roles. Nat Rev Genet. 2013; 14:447–459. [PubMed: 23732335]
31. Gross AM, Orosco RK, Shen JP, Egloff AM, Carter H, Hofree M, et al. Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss. Nat Genet. 2014; 46:939–943. [PubMed: 25086664]
32. Maccani MA, Knopik VS. Cigarette smoke exposure-associated alterations to non-coding RNA. Front Genet. 2012; 3:53. [PubMed: 22509180]
33. Wang DW, Wang ZH, Wang LL, Song Y, Zhang GZ. Overexpression of hiwi promotes growth of human breast cancer cells. Asian Pac J Cancer Prev. 2014; 15:7553–7558. [PubMed: 25292027]
34. Wang Y, Liu J, Wu G, Yang F. Manipulations in HIWI level exerts influence on the proliferation of human non-small cell lung cancer cells. Exp Ther Med. 2016; 11:1971–1976. [PubMed: 27168836]
35. Yang L, Bi L, Liu Q, Zhao M, Cao B, Li D, et al. Hiwi Promotes the Proliferation of Colorectal Cancer Cells via Upregulating Global DNA Methylation. Dis Markers. 2015; 2015:383056. [PubMed: 26355242]
36. Grochola LF, Greither T, Taubert H, Moller P, Knippschild U, Udelnow A, et al. The stem cell-associated Hiwi gene in human adenocarcinoma of the pancreas: expression and risk of tumour-related death. Br J Cancer. 2008; 99:1083–1088. [PubMed: 18781170]
37. Brown JR, Hanna M, Tesar B, Werner L, Pochet N, Asara JM, et al. Integrative genomic analysis implicates gain of PIK3CA at 3q26 and MYC at 8q24 in chronic lymphocytic leukemia. Clin Cancer Res. 2012; 18:3791–3802. [PubMed: 22623730]

38. Estilo CL, P OC, Ngai I, Patel SG, Reddy PG, Dao S, et al. The role of novel oncogenes squamous cell carcinoma-related oncogene and phosphatidylinositol 3-kinase p110alpha in squamous cell carcinoma of the oral tongue. Clin Cancer Res. 2003; 9:2300–2306. [PubMed: 12796399]

39. Woenckhaus J, Steger K, Werner E, Fenic I, Gamerdinger U, Dreyer T, et al. Genomic gain of PIK3CA and increased expression of p110alpha are associated with progression of dysplasia into invasive squamous cell carcinoma. J Pathol. 2002; 198:335–342. [PubMed: 12375266]

40. Beebe-Dimmer JL, Levin AM, Ray AM, Zuhlke KA, Machiela MJ, Halstead-Nussloch BA, et al. Chromosome 8q24 markers: risk of early-onset and familial prostate cancer. Int J Cancer. 2008; 122:2876–2879. [PubMed: 18360876]

41. Chen Z, Che Q, Jiang FZ, Wang HH, Wang FY, Liao Y, et al. Piwil1 causes epigenetic alteration of PTEN gene via upregulation of DNA methyltransferase in type I endometrial cancer. Biochem Biophys Res Commun. 2015; 463:876–880. [PubMed: 26056945]

42. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011; 333:1157–1160. [PubMed: 21798893]

43. Lin M, Smith LT, Smiraglia DJ, Kazhiyur-Mannar R, Lang JC, Schuller DE, et al. DNA copy number gains in head and neck squamous cell carcinoma. Oncogene. 2006; 25:1424–1433. [PubMed: 16247453]

## Highlights

- RNA-seq analysis implicated a 13-member piRNA panel in smoking-related HNSCC
- 5 piRNAs correlated with clinical variables, such as tumor stage and survival
- mRNA expression analysis revealed dysregulation of PIWIL1 in smoking-related HNSCC
- 6 piRNAs as well as PIWIL1 associated with common genomic alterations in HNSCC

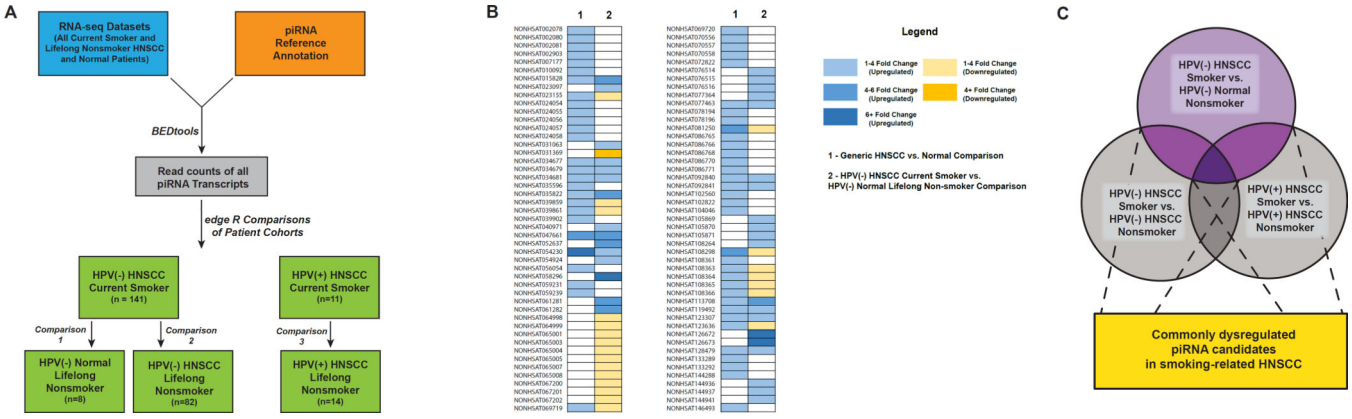


**Figure 1.**

(A) Schematic detailing the RNA-seq analysis pipeline used to identify smoking-dysregulated piRNA candidates ($p < 0.05$, $FDR < 0.05$). (B) Comparison of piRNA dysregulation between generic HNSCC vs. normal tissues and piRNA dysregulation in smoking-related HNSCC vs. nonsmoking normal tissues with piRNAs color-coded by fold change, highlighting the etiological heterogeneity of HNSCCs. Boxes with no color indicate that the piRNA was not found dysregulated in that comparison. (C) Schematic of method used to identify piRNA candidates commonly dysregulated in smoking-induced HNSCCs.

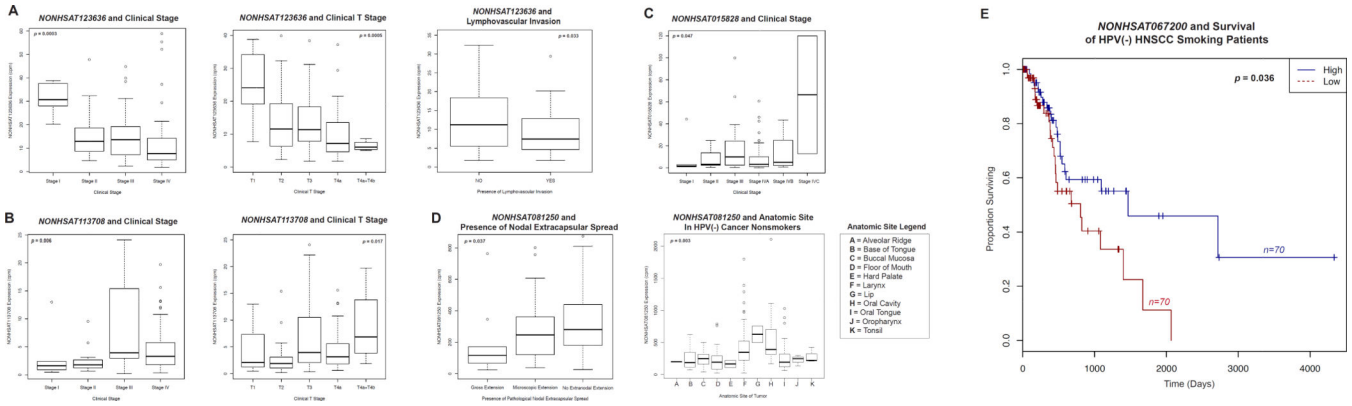


**Figure 2.**

(A–D) Boxplots indicating downregulation of piRNAs *NONHSAT123636* and *NONHSAT081250* and upregulation of piRNAs *NONHSAT113708* and *NONHSAT015828* with increased clinical stage and tumor metastasis, as well as significant variation in *NONHSAT081250* expression based on tumor site (Kruskal-Wallis, $p < 0.05$). (E) Kaplan-Meier curve depicting survival outcomes based on relative high and low expression of *NONHSAT067200*.

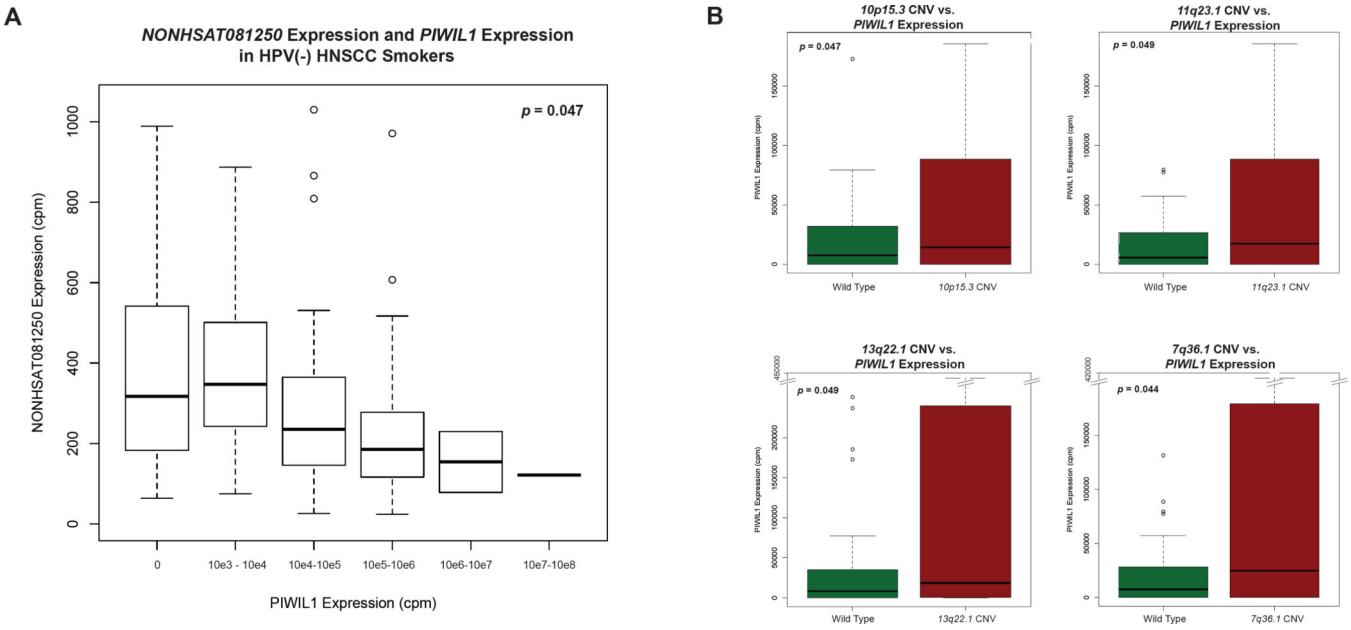


**Figure 3.**

(A) Boxplot showing consistent decrease in *NONHSAT081250* expression with increased expression of PIWI protein *PIWIL1* (Kruskal-Wallis $p < 0.05$). (B) Boxplots depicting correlations between *PIWIL1* expression and specific CNVs commonly found in HNSCC (Wilcoxon rank sum, $p < 0.05$).

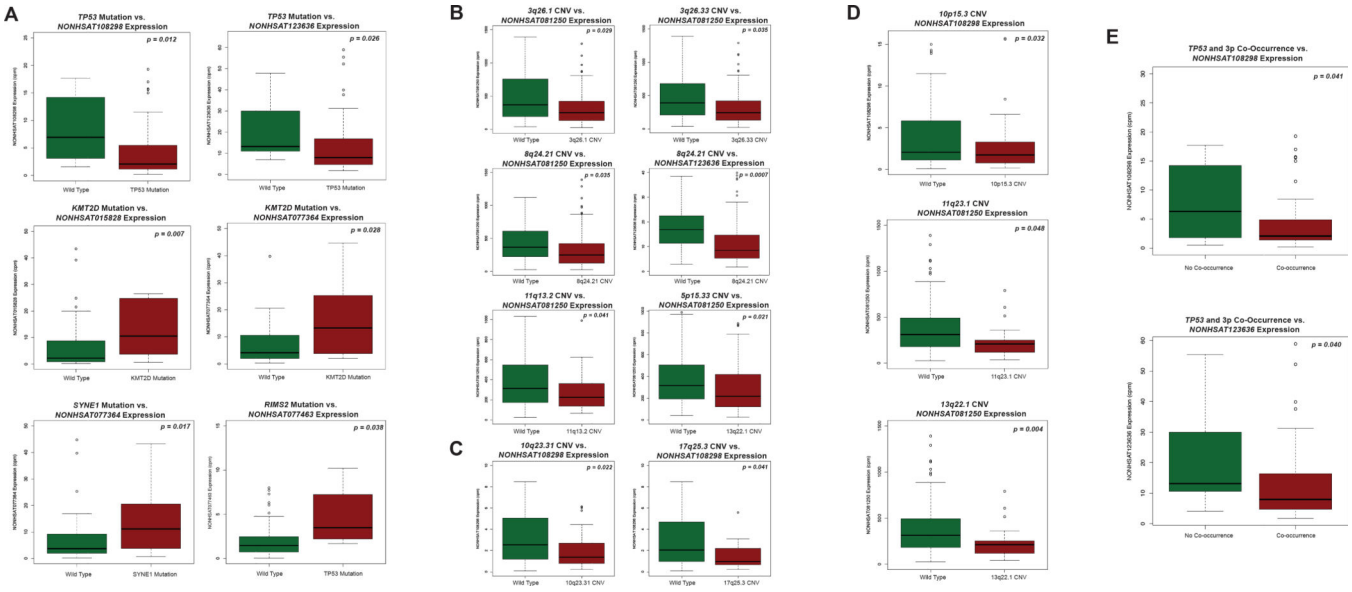


**Figure 4.**

Boxplots correlating expression of smoking-dysregulated piRNAs to (A) specific somatic mutations, (B) specific copy number amplifications, (C) specific copy number deletions, and (D) to CNVs previously associated with elevated *PIWIL1* expression in HNSCC (Wilcoxon rank sum, $p < 0.05$). (E) Association of smoking-dysregulated piRNA expression with concurrent loss of *TP53* and *3p* in HNSCCs (Wilcoxon rank sum, $p < 0.05$).

## Table 1

Multivariate Cox regression analysis on *NONHSAT067200* reveals significant association between low transcript expression and poor prognosis.

| | HR | Lower 95% | Upper 95% | P-value |
|---|---|---|---|---|
| NONHSAT067200 (low vs. high) | 2.523706 | 1.2066 | 5.279 | 0.0139 |
| Age | 0.996979 | 0.9661 | 1.029 | 0.8505 |
| gender (Male vs. Female) | 0.443991 | 0.1969 | 1.001 | 0.0504 |
| Clinical Stage (relative to Stage I) | | | | |
| StageII | 1.787716 | 0.2394 | 13.35 | 0.5712 |
| StageIII | 0.890394 | 0.1535 | 5.166 | 0.897 |
| StageIV | 1.366213 | 0.2598 | 7.185 | 0.7126 |
| Neoplasm Histologic Grade (relative to G1) | | | | |
| G2 | 2.89568 | 0.6743 | 12.435 | 0.1527 |
| G3 | 2.95133 | 0.5963 | 14.606 | 0.1847 |

## Table 2

a – piRNAs dysregulated between HPV(−) HNSCC Current Smokers and Normal Lifelong Nonsmokers with significant clinical correlations. b – piRNAs commonly dysregulated between multiple current smoker vs. lifelong nonsmoker dataset cohorts.

**a: Clinically Significant piRNAs**

| | **HPV(−) Cancer Smoker vs. HPV(−) Normal Nonsmoker Comparison** | | |
|---|---|---|---|
| | **Fold Change** | **p-value** | **FDR** |
| NONHSAT123636 | 0.339323958 | 1.18E-005 | 1.85E-004 |
| NONHSAT081250 | 0.35231788 | 2.50E-005 | 3.57E-004 |
| NONHSAT113708 | 4.544719015 | 5.67E-005 | 7.25E-004 |
| NONHSAT015828 | 4.245166385 | 9.99E-005 | 1.18E-003 |
| NONHSAT067200 | 0.486658784 | 5.48E-003 | 2.81E-002 |

**b: Commonly Dysregulated piRNAs**

| | HPV(+) Cancer Smoker vs. HPV(+) Cancer Nonsmoker Comparison | | | HPV(−) Cancer Smoker vs. HPV(−) Normal Nonsmoker Comparison | | |
|---|---|---|---|---|---|---|
| | Fold Change | p-value | FDR | Fold Change | p-value | FDR |
| NONHSAT069719 | 0.293154275 | 0.000377948 | 0.01593488 | 0.256065962 | 8.93E-009 | 3.36E-007 |
| NONHSAT105869 | 3.331103084 | 0.000235996 | 0.01082467 | 3.565024027 | 4.62E-004 | 4.03E-003 |
| NONHSAT105870 | 3.331103084 | 0.000235996 | 0.01082467 | 3.564975594 | 4.62E-004 | 4.03E-003 |
| NONHSAT105871 | 3.331121555 | 0.00023598 | 0.01082467 | 3.564885402 | 4.62E-004 | 4.03E-003 |
| | HPV(−) Cancer Smoker vs. HPV(−) Cancer Nonsmoker Comparison | | | HPV(−) Cancer Smoker vs. HPV(−) Normal Nonsmoker Comparison | | |
| | Fold Change | p-value | FDR | Fold Change | p-value | FDR |
| NONHSAT052637 | 1.441735587 | 1.45E-003 | 1.99E-002 | 4.983904762 | 2.84E-005 | 3.97E-004 |
| NONHSAT077364 | 1.437821319 | 1.42E-003 | 1.98E-002 | 3.381257526 | 7.46E-004 | 5.92E-003 |

| **b: Commonly Dysregulated piRNAs** | | | | | | |
|---|---|---|---|---|---|---|
| NONHSAT077463 | 1.448806978 | 1.23E-003 | 1.77E-002 | 3.414273903 | 8.19E-004 | 6.35E-003 |
| NONHSAT108298 | 0.557702406 | 8.67E-008 | 5.67E-006 | 0.303693787 | 9.82E-007 | 2.07E-005 |

**Table 3**

Additional correlations between piRNA expression and common genomic alterations in HNSCC.

| piRNA | Mutated Gene | p-value |
|---|---|---|
| NONHSAT015828 | PTEN | 0.04500568 |
| NONHSAT077463 | SYNE1 | 0.046761032 |
| piRNA | CNV | p-value |
| NONHSAT108298 | 20p12.2 | 0.010639047 |
| NONHSAT108298 | 2q21.2 | 0.016429416 |
| NONHSAT123636 | 8q11.21 | 0.041808507 |
| NONHSAT123636 | 20p12.2 | 0.035671195 |
| NONHSAT123636 | 20q11.22 | 0.020166353 |
| NONHSAT123636 | 2q21.2 | 0.036380564 |
| NONHSAT081250 | 7p11.2 | 0.048063426 |
| NONHSAT081250 | 8q11.21 | 0.040061748 |
| NONHSAT081250 | 18q23 | 0.00114509 |