

UC San Diego

UC San Diego Previously Published Works

Title

Transcriptome sequencing uncovers novel long noncoding and small nucleolar RNAs dysregulated in head and neck squamous cell carcinoma

Permalink

<https://escholarship.org/uc/item/0vs4t96j>

Journal

RNA, 21(6)

ISSN

1355-8382

Authors

Zou, Angela E
Ku, Jonjei
Honda, Thomas K
[et al.](#)

Publication Date

2015-06-01

DOI

10.1261/rna.049262.114

Peer reviewed

Transcriptome sequencing uncovers novel long non-coding and small nucleolar RNAs dysregulated in head and neck squamous cell carcinoma

Angela E. Zou¹, Jonjei Ku¹, Thomas K. Honda¹, Vicky Yu¹, Selena Z. Kuo¹, Hao Zheng¹, Yinan Xuan¹, Andrew Hinton², Kevin T. Brumund¹, Jonathan H. Lin³, Jessica Wang-Rodriguez³, Weg M. Ongkeko^{1*}

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

²Department of Pediatrics, University of California, San Diego, La Jolla, California, United States of America

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

* Corresponding author

Shortened Title:

Novel non-coding RNAs in head and neck cancer

Keywords:

HNSCC; Long non-coding RNAs; RNA-sequencing; Small nucleolar RNAs

Abstract

Head and neck squamous cell carcinoma persists as one of the most common and deadly malignancies, with early detection and effective treatment still posing formidable challenges. To expand our currently sparse knowledge of the non-coding alterations involved in the disease and identify potential biomarkers and therapeutic targets, we globally profiled the dysregulation of small nucleolar and long non-coding RNAs in head and neck tumors. Using next-generation RNA-sequencing data from 40 pairs of tumor and matched normal tissues, we found 2,808 long non-coding RNA (lncRNA) transcripts significantly differentially expressed by a fold change magnitude ≥ 2 . Meanwhile, RNA-sequencing analysis of 31 tumor-normal pairs yielded 33 significantly dysregulated small nucleolar RNAs (snoRNA). In particular, we identified 2 dramatically downregulated lncRNAs and 1 downregulated snoRNA whose expression levels correlated significantly with overall patient survival, suggesting their functional significance and clinical relevance in head and neck cancer pathogenesis. We confirmed the dysregulation of these non-coding RNAs in head and neck cancer cell lines derived from different anatomic sites, and determined that ectopic expression of the 2 lncRNAs inhibited key EMT and stem cell genes and reduced cellular proliferation and migration. As a whole, non-coding RNAs are pervasively dysregulated in head and squamous cell carcinoma. The precise molecular roles of the three transcripts identified warrants further characterization, but our data suggest that they are likely to play substantial roles in head and neck cancer pathogenesis and are significantly associated with patient survival.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth leading cancer by incidence worldwide (Ferlay et al. 2010). Despite advances in treatment and increased knowledge about underlying genetic and environmental risk factors behind the disease, the five-year survival rate for HNSCC has persisted at 50% for more than three decades, with most cases remaining undiagnosed until metastasis (Nemunaitis and O'Brien 2002). A deeper understanding of the drivers and mechanisms behind HNSCC pathogenesis, as well as continued identification of candidate biomarkers and prognostic factors, is necessary to inform better diagnostic and therapeutic strategies.

While a number of gene mutations and their impacts on signaling pathways have been implicated in HNSCC onset and progression, there still remains much to be elucidated about the role that perturbations in the non-coding genome may play in these processes. Non-coding RNAs (ncRNAs) have been increasingly viewed as key players in human disease, with recent studies establishing their involvement in gene regulation, cell differentiation, self-renewal, and cell proliferation, among other pivotal functions, as well as their dysregulation in many cancers (Esteller 2011). Existing studies of their roles in HNSCC, however, have primarily focused on profiling microRNAs (Avisar et al. 2009; Childs et al. 2009; Ramdas et al. 2009; Hui et al. 2010), a class of small (~22 nt) and relatively well-documented ncRNAs that comprise but a fraction of the non-coding transcriptome (Esteller 2011).

In contrast, long non-coding RNAs (lncRNAs), transcripts exceeding 200 nt in length, are encoded by a large proportion of the so-called “dark matter” of the human genome (Mercer et al.

2009; Kapranov et al. 2010). While a vast majority of lncRNAs remain functionally uncharacterized, the few that are have been linked to a range of biological processes, including chromatin modification, regulation of transcription factors, mRNA processing and degradation, and even cell-cell signaling (Geisler and Collier 2013). Furthermore, studies suggest that their expression patterns are even more tissue-specific than those of protein-coding genes (Cabili et al. 2011) and that they are involved in pivotal tumor suppressive and oncogenic pathways. Increased expression of the lncRNA *HOTAIR* in primary breast tumors, for instance, was shown to strongly correlate with metastasis and death (Gupta et al. 2010), with similar upregulation and prognostic significance observed in a number of other cancers, including non-small cell lung cancer and pancreatic cancer (Zhang et al. 2014). Similarly, overexpression of the lncRNA *MALAT-1* has been linked to hepatocellular carcinoma and metastasis in small-cell lung cancer (Lin et al. 2007). Several studies of these well-characterized lncRNAs in HNSCC have revealed promising associations as well, with *HOTAIR*, *MEG-3*, *MALAT-1*, and *NEAT-1* each exhibiting dysregulation and/or prognostic significance in several subtypes of HNSCC (Fang et al. 2014, Feng et al. 2012, Li et al. 2013, Tang et al. 2013, Yang and Deng 2014). Additionally, *GAS5* (growth arrest specific transcript 5) has been linked to low expression and poor prognosis in HNSCC (Gee et al. 2011), and most recently, Shen et al. reported the dysregulation of two relatively uncharacterized transcripts, *AC026166.2* and *RP11-169D4.1*, in LSCCs using microarray and qRT-PCR expression profiling (Shen et al. 2014). However, aside from these investigations involving mostly known, cancer-associated lncRNAs, the dysregulation and functional significance of lncRNAs in HNSCC has remained largely uncharacterized.

Similarly, small nucleolar RNAs (snoRNAs), which are components of small nucleolar ribonucleoproteins (snoRNPs) and intermediate in length (60-300 nt) (Esteller 2011; Williams

and Farzaneh 2012), are also emerging as potential players in the development of cancer. As facilitators of the methylation and pseudouridylation of rRNA and other post-transcriptional modifications involved in ribosome production (Esteller 2011), snoRNAs were long considered to be merely housekeeping RNAs. Studies, however, have increasingly implicated snoRNA dysregulation as a potential determinant of cell fate and a driver of oncogenesis. Drastic downregulation of human *S5* snoRNA genes was observed between meningiomas and normal brain tissue (Chang et al. 2002), while a panel of six snoRNAs was found to be overexpressed in non-small cell lung cancer (Liao et al. 2010). More recent studies have revealed that a germline homozygous deletion of 2 base pairs in the snoRNA *U50* gene was significantly linked to prostate cancer development, and that a heterozygous genotype for the same deletion, both somatically and in germline, was frequently observed in breast cancers (Dong et al. 2009). The identification of dysregulated snoRNAs in multiple cancer types thus far suggests that promising associations between snoRNAs and HNSCC pathogenesis may exist as well.

In our study, we aimed to profile the global expression patterns and dysregulation of lncRNAs and snoRNAs in HNSCC by utilizing RNA-sequencing (RNA-seq), a next-generation deep sequencing technology that captures the transcriptional landscape of the human genome through reverse transcription of isolated RNA and high-throughput sequencing of the resulting cDNA. By comparing the RNA-seq profiles of lncRNA and snoRNA expression levels in 40 primary HNSCC tumors with those of their matched normal counterparts, we generated a panel of differentially expressed ncRNAs and further characterized selected candidates to gauge the extent of their dysregulation and their functional relevance in HNSCC.

Results

Analysis of primary tumor and matched normal RNA-seq datasets uncovers widespread dysregulation of lncRNAs and snoRNAs in HNSCC

To identify HNSCC-associated lncRNAs, we searched for alterations in their expression using RNA-seq data from 40 patients in The Cancer Genome Atlas (TCGA) repository (Table 1a, Additional File 8a). For each of these patients, raw sequence data of both the primary HNSCC tumor and matched adjacent normal tissue were available. To profile lncRNA expression levels, we mapped each RNA-seq library to a reference annotation of 32,108 human lncRNAs. Using the R/Bioconductor package edgeR, we filtered and normalized the read count data and compared lncRNA expression, in the form of counts-per-million (cpm), between HNSCC and normal samples. From the 12,407 consistently detected lncRNA transcripts and isoforms, a total of 7,120 differentially expressed lncRNAs (FDR < 0.05) were identified between the HNSCC and adjacent normal tissues (Figure 1). Among these, 2,808 transcripts were differentially expressed by at least ± 2 -fold (FDR < 0.05), ranging from *lnc-LCE5A-1*, with predicted downregulation of approximately 112-fold in HNSCC, to *lnc-BCHE* with predicted upregulation of approximately 58-fold in HNSCC. Further filtering yielded 222 lncRNA transcripts differentially expressed by at least ± 8 fold (FDR < 0.0001) (Figure 2, Additional File 1).

To explore the dysregulation of snoRNAs in HNSCC, we obtained the expression profiles of 325 snoRNAs in 31 HNSCC samples and their matched adjacent normal tissues from The Cancer Genome Atlas (TCGA) (Table 1b, Additional File 8b). We normalized the snoRNA read count data in edgeR and compared snoRNA expression, in the form of cpm, between the HNSCC and normal samples. In total, 33 differentially expressed snoRNAs (FDR < 0.05) were identified

between the HNSCC and matched normal tissues, ranging from *SNORD116-20* with predicted downregulation of more than 4-fold in HNSCC to *SNORD60* with predicted upregulation of about 3-fold in HNSCC (Figure 3, Additional File 2).

Known, cancer-associated lncRNAs in HNSCC and stratification of lncRNA dysregulation by HNSCC anatomic site

GAS5 (growth arrest specific transcript 5), a lncRNA downregulated in breast cancers and shown to induce apoptosis (Mourtada-Maarabouni et al. 2009), was recently implicated in HNSCCs as well (Gee et al. 2011). While we found *GAS5* downregulation in HNSCC to be statistically significant, with 29 splice isoforms of the lncRNA appearing in our original panel of 7,120 differentially expressed lncRNA transcripts, the difference in *GAS5* expression levels between HNSCC and normal tissues was determined to be less than 2-fold with $p > 0.0001$, so we did not identify *GAS5* as a candidate for further study (Additional File 3). *MEG-3* (maternally expressed gene 3), a tumor suppressor lncRNA implicated in p53 activation (Zhang et al. 2010), was also downregulated in HNSCCs by less than 2-fold, $p > 0.0001$ (Additional File 3). We observed no significant dysregulation of many other lncRNAs previously linked to cancers, including *HOTAIR*, *MALAT-1*, *ANRIL*, *NEAT-1*, and *UCA-1*, in our overall cohort.

To investigate possible heterogeneity in lncRNA expression among different HNSCC tumor sites, we stratified our patients by anatomic subdivision (Table 1a) and performed tumor vs. paired normal differential expression analysis on each subcohort using edgeR. Among our 15 patients presenting oral squamous cell carcinoma, 777 lncRNA transcripts were differentially expressed by at least ± 4 -fold ($p < 0.0001$). Notably, we found that *HOTAIR* was significantly upregulated and numerous splice isoforms of *MEG-3* and *GAS5* were downregulated in primary

tumors (Additional File 4a). Meanwhile, analysis of the 13 tongue squamous cell carcinoma pairs yielded 1,020 lncRNA transcripts dysregulated by ≥ 4 -fold in magnitude ($p < 0.0001$), with downregulation observed in *GAS5* (Additional File 4b). Finally, in our 11 tumor-normal laryngeal squamous cell carcinoma (LSCC) paired samples, we found 657 lncRNA transcripts deregulated by at least ± 4 -fold. While none of the aforementioned well-documented lncRNAs were determined to be dysregulated in LSCC, we did observe slight downregulation of RP11-169D4.1 (Additional File 4c), a lncRNA recently associated with survival in LSCC patients (Shen et al. 2014).

Univariate and multivariate analyses reveal 2 lncRNAs and 1 snoRNA significantly associated with HNSCC patient outcome

We next examined the relationship between the expression levels of selected ncRNAs in HNSCC patients and overall survival in order to determine which lncRNAs and snoRNAs could potentially be of functional significance in HNSCC pathogenesis. We measured the expression levels of the 30 most significantly dysregulated lncRNAs in an additional 45 randomly selected HNSCC RNA-seq datasets from TCGA, doubling the size of our cohort ($n = 85$), and modeled lncRNA expression level as a binary variable, with the “low” group corresponding to the 50% of HNSCC patients with relative low expression of the lncRNA, and “high” corresponding to the 50% of HNSCC patients exhibiting relative high expression of the lncRNA. Low expression of 2 lncRNAs that RNA-seq analysis identified as dramatically downregulated in HNSCC, *lnc-LCE5A-1* and *lnc-KCTD6-3* (Table 2), correlated with poor survival under univariate Kaplan-Meier analysis, with a Cox regression model also yielding significant results (*lnc-LCE5A-1*: $p = 0.048$, Hazard Ratio (HR) = 1.65; *lnc-KCTD6-3*: $p = 0.032$, HR = 1.73) (Figure 4a-b). We obtained snoRNA expression profiles from an additional 99 randomly selected HNSCC tumors

from TCGA, resulting in a cohort of size of $n = 130$. Univariate analysis revealed that lower expression of *SNORD35B* (*U35B*), a snoRNA downregulated in HNSCC, served as an adverse prognostic factor for survival, with Cox regression analysis also finding significant correlation ($p = 0.009$, HR = 1.94) (Figure 4c).

We then performed multivariate analysis for *lnc-LCE5A-1*, *lnc-KCTD6-3*, and *SNORD35B* and found that expression of these 3 non-coding RNAs operated as prognostic factors independent of established clinical risk factors, including pathologic stage and tumor grade, among other patient characteristics (Table 3). To eliminate age as a confounding variable in patient survival and ncRNA expression level, we conducted a secondary analysis excluding patients > 85 years of age. Additionally, we modeled patient age as both a continuous and binary covariate (< or >= 75 years) in order to capture nonlinear associations between patient age and survival. Under this analysis, we found that high vs. low expression of all three ncRNAs remained a significant independent predictor of patient outcomes (Additional File 5).

Transcript characterization and coverage mapping of survival-associated ncRNAs

Visualization of both lncRNAs in the Ensembl genome browser revealed their antisense overlap with protein-coding regions, while *SNORD35B* was found to be encoded in an intron of *RPS11* (Figure 5). To confirm that the mapped reads spanned the entirety of our identified lncRNAs, we plotted the average per-base coverage of each exon in the lncRNAs (Additional File 6a-b). We also compared the sequences and locations of *SNORD35B* and its paralog, *SNORD35A* (Additional File 6c).

***In-vitro* verification of survival-associated ncRNAs**

To verify the dysregulation of *lnc-LCE5A-1*, *lnc-KCTD6-3*, and *SNORD35B*, we examined their relative expression in two normal oral epithelial cell lines (OKF4 and OKF6) compared to five established HNSCC cell lines (UMSCC-10B, UMSCC-22B, HN-1, HN-12, and HN-30) using real-time qRT-PCR (quantitative reverse transcriptase PCR). We found that the three ncRNAs were nearly exclusively downregulated in the HNSCC cell lines, although *lnc-LCE5A-1* and *lnc-KCTD6-3* were both downregulated to a lesser extent than predicted by RNA-seq analysis of the patient sequence data (Figure 6).

Overexpression of *lnc-LCE5A-1* and *lnc-KCTD6-3* in HNSCC cell lines decreases stem cell and EMT gene expression and reduces migration and proliferation

To explore the functional significance of our 2 HNSCC-dysregulated lncRNAs, we first examined their relationship with established epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) genes. We transfected *lnc-LCE5A-1* and *lnc-KCTD6-3* expression plasmids into HNSCC cell lines and measured their overexpression using qRT-PCR (Additional File 7). Ectopic expression of *lnc-LCE5A-1* increased the expression of *CDH-1* (E-cadherin) in HNSCC cells, while decreasing the expression of *OCT-4*, *NANOG*, and *VIM* (Vimentin) (Figure 7a). Meanwhile, overexpression of *lnc-KCTD6-3* reduced the expression of *NANOG*, *BMI-1*, *TWIST*, and *VIM* in HNSCC cell lines (Figure 7b).

Because the expression of EMT-related genes was consistently altered in lncRNA-transfected cells, we examined whether these lncRNAs could inhibit cell migration, a key component of the EMT phenotype, using wound healing migration assays. Over a 24-hour period, we observed markedly reduced migration of *lnc-LCE5A-1* and *lnc-KCTD6-3*-transfected HNSCC cells as compared to vector-transfected controls (Figure 8).

Additionally, to investigate the effects of both lncRNAs on cell proliferation, we performed an MTS assay on *lnc-LCE5A-1* and *lnc-KCTD6-3*-transfected HNSCC cell lines. Both lncRNAs led to a reduction in proliferation of at least 25% over 3 days (Figure 9).

Discussion

The molecular basis of HNSCC has largely been studied in the context of environmental risk factors and predispositions in the protein-coding genome. Despite extensive research and advances in HNSCC treatment, there exists a paucity of biomarkers for early detection and few targetable gene mutations, and survival rates for the disease have shown little improvement (Agrawal et al. 2011). Meanwhile, much remains to be characterized regarding the epigenetic and transcriptomic landscapes of HNSCC; their alterations and functional roles may help shape a new generation of HNSCC diagnostic and therapeutic strategies.

To our knowledge, this is the first study utilizing next-generation sequencing to globally profile the dysregulation of lncRNAs and snoRNAs in HNSCC. We first compiled and compared the expression patterns of 32,108 lncRNA transcripts in HNSCC and adjacent normal tissues. Among the 12,407 lncRNA transcripts consistently detected in these samples, nearly one-fourth (2,808) were significantly differentially expressed in HNSCC by a fold change of ≥ 2 , with 222 transcripts differentially expressed by 8-fold or more. We also compared the expression profiles of 325 snoRNAs in 31 HNSCC and matched normal tissues, and of the 135 snoRNAs expressed, about one-fifth (33) were significantly dysregulated in HNSCC. Such a sizeable volume and magnitude of dysregulation in both lncRNAs and snoRNAs, coupled with existing findings regarding the differential expression of various microRNAs (Avissar et al. 2009; Childs et al.

2009; Ramdas et al. 2009; Hui et al. 2010) and well-documented lncRNAs (Yang and Deng 2014), suggests that aberrations associated with the non-coding transcriptome are pervasive in HNSCC and if functionally characterized, may substantially expand our knowledge about the epigenetic mechanisms and regulatory alterations involved in HNSCC pathogenesis.

Previous studies on lncRNAs in HNSCC have primarily taken place in the context of known cancer-associated transcripts. Multiple isoforms of *GAS5* and *MEG-3*, lncRNAs previously implicated in HNSCCs (Gee et al. 2011, Tang et al. 2013), were found to be differentially expressed in our cohort, but with small fold change magnitudes. Several other known cancer-associated lncRNAs have also been implicated in HNSCCs, but most are only limited to tumors arising in certain anatomic sites. *HOTAIR*, for instance, showed elevated expression in nasopharyngeal, laryngeal, and oral squamous cell carcinomas, but did not exhibit significant upregulation in tongue squamous cell carcinomas; similarly, *NEAT-1*, *MALAT-1*, and *UCA-1* were inconsistently dysregulated among laryngeal, oral, and tongue squamous cell carcinomas (Yang and Deng 2014). Our primary study profiled ncRNA dysregulation independently of HNSCC anatomic subdivision, with samples originating from multiple tumor sites. We failed to identify *HOTAIR*, *NEAT-1*, *UCA-1* or *MALAT-1*, among several other cancer-associated lncRNAs, as differentially expressed across our cohort, suggesting a degree of consistency with previous findings.

A number of studies have identified molecular distinctions among HNSCCs of varying anatomic origin (Belbin et al. 2008, Kokko et al. 2011, Lleras et al. 2013, Rodrigo et al. 2001). CD44 expression, for instance, was found to be predictive of five-year survival in primary pharyngeal and laryngeal tumors, but not in tumors arising in the oral cavity (Kokko et al. 2011).

Meanwhile, genome-wide methylation profiling revealed unique DNA methylation signatures associated with different HNSCC anatomic sites (Lleras et al. 2013). Given the tissue specificity of lncRNAs and their prolific involvement in many aspects of genomic and transcriptional regulation (Cabili et al. 2011), changes in lncRNA expression may contribute to the emergence of site-specific characteristics in HNSCC.

To further study anatomic site-specific lncRNA expression and dysregulation, we stratified our patients by primary tumor site (oral, tongue, and laryngeal) and performed differential expression analysis on each subcohort. We found significant dysregulation of *HOTAIR* and *MEG-3* in oral squamous cell carcinomas, consistent with a prior study by Tang et al. (Tang et al. 2013); however, in contrast with their report, we observed no significant dysregulation of *NEAT-1* or *UCA-1*. We also found no significant overexpression of *UCA-1* in our 13 patients with tongue squamous cell carcinoma (TSCC), although its dysregulation was previously reported by Fang et al. (Fang et al. 2014). In accordance with Fang et al., however, we failed to implicate *HOTAIR*, *NEAT-1*, *MALAT-1*, *MEG-3*, or *HULC* as differentially expressed in TSCC. Finally, in our 11 laryngeal squamous cell carcinomas (LSCC), we observed moderate downregulation of *RP11-169D4.1*, a transcript recently associated with LSCC patient survival (Shen et al. 2014), yet found no significant dysregulation of *HOTAIR*, *MALAT-1*, or *AC026166.2*, although all three lncRNAs were previously implicated in LSCCs (Yang and Deng 2014, Shen et al. 2014). These discrepancies regarding the dysregulation of lncRNAs in various HNSCC anatomic sites merit further investigation; larger patient cohorts or multi-center studies, as well as more uniform, next-generation expression profiling approaches, may play important roles in resolving such differences. The 2 lncRNAs we identify here, however, consistently appear among the most downregulated transcripts in all three anatomic sites (Additional File 4). Taken together, these

preliminary analyses suggest that there is a high likelihood of heterogeneity among the non-coding pathways altered in HNSCCs arising from different anatomic sites, but that the dysregulation of certain lncRNAs may still persist as unifying features in HNSCC development and progression.

We also conducted our study independently of patients' HPV status and tobacco and/or alcohol consumption, even though these factors have been shown to produce distinctions in HNSCC gene and microRNA signatures and prompt different modes of clinical management (Hui et al. 2010; Agrawal et al. 2011; Stransky et al. 2011). The fact that most existing and well-characterized genes, microRNAs, and lncRNAs are of limited applicability in explaining the pathogenesis of HNSCC as a whole further underscores the importance of identifying alternative dysregulated ncRNAs as biomarkers or prognostic indicators in HNSCC, and determining, in turn, their molecular roles in HNSCC pathogenesis.

Using RNA-seq, we were able to quantify the expression levels of not only known, cancer-associated lncRNAs, but also those of previously uncharacterized lncRNA transcripts. Despite evidence of key variations at the molecular level distinguishing certain HNSCC subtypes, we were able to identify universal alterations in the transcriptomic landscape of HNSCC that occurred regardless of the roles of environmental factors or anatomic sites in HNSCC etiology. To identify lncRNAs and snoRNAs with the greatest potential functional significance, we examined the correlation between the expression levels of our candidate transcripts in HNSCC patients and these patients' overall survival using univariate and multivariate Cox regression analyses, identifying 2 lncRNAs, *lnc-LCE5A-1* and *lnc-KCTD6-3*, and 1 snoRNA, *SNORD35B*, whose low expression in HNSCC tumors was significantly associated with poor survival. We are

the first to associate these two lncRNAs with disease, as well as the first to implicate snoRNAs as both widely dysregulated and prognostically significant in HNSCC.

We verified the dysregulation of these three transcripts *in-vitro* by comparing their expression levels in normal oral epithelial cell lines with those in HNSCC cell lines. We did find that the fold changes of these ncRNAs between HNSCC and normal cell lines were inconsistent with those predicted by the RNA-seq patient data analysis, with *lnc-LCE5A-1* and *lnc-KCTD6-3* exhibiting a significantly more moderate degree of dysregulation than indicated by the clinical data. Such a disparity may be attributed to the fact that the cell lines were not matched samples; furthermore, dramatically reduced differences in endogenous lncRNA expression between normal and cancerous cell lines appears to be a feature in other lncRNA studies as well (Gupta et al. 2010).

We next investigated the functional significance of *lnc-LCE5A-1* and *lnc-KCTD6-3* in terms of their effects on EMT and cancer stem cell (CSC) characteristics. Enforced expression of the lncRNAs in HNSCC cell lines led to downregulation of CSC- and EMT-inducing genes, with *lnc-LCE5A-1* promoting the upregulation of key EMT-inhibitor *CDH-1* (E-cadherin) as well. *lnc-LCE5A-1* and *lnc-KCTD6-3*-transfected cells also exhibited notably reduced rates of cellular migration and proliferation, further indicating that these lncRNAs may play significant roles in inhibiting EMT and stem cell-like phenotypes.

The association of both lncRNAs with EMT and CSC signatures underscores the importance and potential clinical significance of the EMT-CSC model of tumorigenesis and disease progression in HNSCC. Because EMT induction leads to the acquisition of invasive and migratory

capabilities in cancer cells and promotes the formation of tumor-initiating, self-renewing CSCs (Reya et al. 2001, Singh and Settleman 2010), its potential regulation by *lnc-LCE5A-1* and *lnc-KCTD6-3* may account for the reduced survival observed among patients expressing low levels of these lncRNAs. Further characterization of the nature and mechanism behind *lnc-LCE5A-1* and *lnc-KCTD6-3* modulation of EMT and CSC characteristics may establish these lncRNAs as promising targets of HNSCC therapies intending to diminishing risks of recurrence and metastasis.

While we assessed the roles of these lncRNAs in terms of their impact on well-established tumorigenic pathways, the *cis*-antisense orientation of both lncRNAs with protein-coding genes is intriguing and may contribute to the functional scope of these lncRNAs as well. *CRNN* (also known as *C1orf10*) overlaps the intron of *lnc-LCE5A-1* (Figure 5a) and encodes the protein cornulin, which is expressed in the upper cell layers of differentiated squamous tissues and linked to epithelial differentiation (Contzler et al. 2005). Low expression of *CRNN* has been observed in esophageal and oral squamous cell carcinomas and implicated in cell proliferation (Imai et al. 2005; Luthra et al. 2007). Because antisense intersections of lncRNAs and mRNAs may influence expression of the latter (Guil and Esteller 2012), further studies on the association in expression between *lnc-LCE5A-1* and *CRNN*, as well as characterization of the possible molecular mechanisms behind such a relationship, may be valuable.

Similarly, *lnc-KCTD6-3* exhibits an antisense overlap with introns of *FAM107A* (Figure 5b), a gene found to exhibit low or complete loss of expression in a number of cancers, including renal cell, non-small cell lung, and prostate cancers, as well as astrocytomas, while studies of enforced expression of *FAM107A* show that it inhibits cell proliferation and promotes apoptosis (Yamato

et al. 1999). Additionally, *lnc-KCTD6-3* overlaps *FAM3D*, a gene observed to be upregulated in the bloodstream of colon cancer patients that has been proposed as a potential biomarker for early detection (Solmi et al. 2004). Investigation of the *cis*-regulatory potential of *lnc-KCTD6-3* with respect to *FAM107A* and *FAM3D* may prove fruitful as well.

Meanwhile, *SNORD35B*, a box C/D snoRNA thought to mediate the 2'-O-methylation of 28S ribosomal RNA (residue C4506)(Nicoloso et al. 1996, Lestrade and Weber 2006), is located in an intronic region and co-transcribed with host gene *RPS11* (Figure 5c). While the role of the *RPS11* in cancer is unknown, Nadano et al. reported aberrant expression of the ribosomal protein S11 in a panel of cancer cell lines. Intriguingly, these differences in S11 expression were not shown to affect ribosomal biogenesis or protein composition (Nadano et al. 2000), prompting speculation of a non-ribosomal function for S11, and in turn, *RPS11* and its intronic snoRNA. Many snoRNAs, including previously disease-associated snoRNAs, are intronically oriented; nine box C/D snoRNAs, for instance, are encoded in the introns of lncRNA *GAS5* and were implicated as possible factors in oncogenesis as well due to their low expression and correlation with survival in breast carcinomas (Gee et al. 2011). Additionally, four box C/D snoRNAs, *SNORD32A*, *SNORD33*, *SNORD34*, and *SNORD35A*, are located in the introns of *RPL13A*, a component of the 60s ribosomal subunit protein (Michel et al. 2011). Interestingly, a cell line with reduced expression of these snoRNAs did not result in differential methylation of ribosomal RNAs as predicted, but linked downregulation of the snoRNAs to increased resistance to lipotoxicity and regulation of metabolic stress pathways (Michel et al. 2011).

Like *SNORD35B*, *SNORD35A* was also predicted to facilitate methylation of 28S rRNA in terms of canonical snoRNA function, and we found it to be significantly downregulated in HNSCCs as

well (Additional File 2). These parallels suggest that *SNORD35B*, as a *trans*-duplicated paralog of *SNORD35A* with sequence and structural similarity (Additional File 6), may similarly assume a repertoire of non-nucleolar or non-ribosomal functions. Functional analysis of both paralogs in the context of cancer would therefore be valuable, and could further cement snoRNAs as a novel and potentially fruitful source of insight into HNSCC pathogenesis.

Conclusions

Our findings suggest that dysregulation of the non-coding transcriptome in HNSCC is both extensive and enormously complex. In particular, we have identified three transcripts whose expression levels correlate significantly to patient survival, including two novel lncRNAs that confer a tumor suppressive phenotype. Although further characterization of their molecular mechanisms remains necessary, these ncRNAs may play functionally vital roles in HNSCC pathogenesis. Taken as a whole, our findings demonstrate that RNA-seq transcriptome profiling on matched tumor and normal pairs generates novel insights into cancer biology, resulting in the implication of many previously uncharacterized elements in HNSCC pathogenesis and also yielding findings that may be applicable to other malignancies.

Methods

lncRNA Expression Profiling:

RNA-seq libraries were obtained from The Cancer Genome Atlas (TCGA) (<https://tcga-data.nci.nih.gov/tcga/>), which contained 40 patients with either RNASeq or RNASeqv2 data for HNSCCs (arising in the oral cavity, tongue, or larynx) and their adjacent normal tissue samples (TCGA IDs in Additional File 8). For each of these datasets, sequence reads had been aligned

using the BWA algorithm (<http://bio-bwa.sourceforge.net/>, default parameters) to a reference transcript database derived from hg19 (UCSC).

A BED annotation file containing 32,108 human lncRNA transcripts and splice isoforms was downloaded from LNCipedia (Volders et al. 2013), a database curating lncRNA structures and sequences from multiple sources, including the lncRNadb, Broad Institute, Ensembl, Gencode, and NONCODE. The BEDtools utility coverageBed (Quinlan and Hall 2010) was used to generate lncRNA expression values in the form of integer read counts for each dataset by computing the number of alignments from each library that overlapped each lncRNA feature given by the annotation file.

snoRNA Expression Profiling:

RNAseqv2-generated expression profiles of known snoRNAs were available in TCGA for 31 HNSCCs (arising in the oral cavity, tongue, or larynx) and their adjacent normal tissues (TCGA IDs in Additional File 8). snoRNA expression values in the form of integer read counts were obtained for each dataset.

lncRNA and snoRNA Differential Expression Analysis:

Using the R/Bioconductor software package edgeR (<http://www.bioconductor.org/packages/release/bioc/html/edgeR.html>), differential expression analysis was performed on the lncRNA and snoRNA read count data produced for the HNSCC and normal sample pairs. lncRNAs and snoRNAs with very low expression (counts per million < 1 for more than one-half of the samples) were excluded from the analysis. Recalculation and accounting for differences between sample library sizes was accomplished by using trimmed mean of M-values (TMM) to compute normalization factors.

To accommodate the paired nature of the experimental designs, the negative binomial generalized linear model (GLM) functionality in edgeR, along with Cox-Reid dispersion

estimates, was employed. These methods have demonstrated success in assessing differential expression in multi-factor experiments while appropriately accounting for biological and technical variation (McCarthy et al. 2012). lncRNAs and snoRNAs whose fold changes were found to vary between HNSCC and paired normal samples by a fold change magnitude greater than 1 and whose false discovery rates (FDR) fell below 0.05 were identified by edgeR to be differentially expressed. To form a more selective panel of lncRNAs for further analysis, filtering of the edgeR lncRNAs was performed based on the following criteria: $|\text{HNSCC/Normal}| \geq 8$ -fold change and $\text{FDR} < 0.0001$, yielding 222 lncRNAs of interest. All 33 snoRNAs determined by edgeR to be differentially expressed were retained as candidates.

lncRNA and snoRNA Survival Analysis:

lncRNA expression was profiled (according to the procedure above) for an additional 45 randomly selected HNSCC TCGA datasets, doubling the size of our cohort ($n = 85$), while 100 additional randomly selected RNASeq-generated snoRNA expression profiles were obtained from TCGA, resulting in a cohort size of 130. All data regarding patient survival, demographics, and clinicopathological characteristics was obtained from TCGA. For both univariate and multivariate analyses, we modeled lncRNA and snoRNA expression level as a binary variable, with the “low” group corresponding to the one-half of HNSCC patients exhibiting relative low expression of the ncRNA, and “high” corresponding to the other half of the HNSCC patients with relative high expression of the ncRNA. For multivariate Cox analysis, patient age was modeled as a continuous variable, and clinical stage and pathologic stage were treated as categorical variables (with Stage I set as the reference). Gender (male vs. female) and tumor grade (G3-G4 vs. G1-G2) were modeled as binary variables. Patients with incomplete or unavailable data in any of these categories were excluded from the multivariate analysis, resulting in cohort sizes of 75 for the lncRNA analyses and 109 for the snoRNA analysis (TCGA

IDs used listed in Additional File 9). The Statsoft software STATISTICA was used for survival models.

Cell Lines and Cell Culture:

The two non-cancerous oral epithelial cell lines used in the qRT-PCR experiments, OKF4 and OKF6, were gifts from the Rheinwald Lab at Harvard Medical School. Both were cultured in Keratinocyte-SFM(1X) with L-glutamine, supplemented with 0.2 ng/mL human recombinant epidermal growth factor (EGF) 1-53, 25 ug/mL bovine pituitary extract (BPE), 0.3 mM calcium chloride, and penicillin streptomycin. Once they attained 30% confluency, they were cultured in equal parts supplemented Keratinocyte-SFM medium and DFK medium. DFK was made with equal parts DMEM and F-12 and supplemented with 0.2 ng/mL EGF 1-53, 25 ug/mL BPE, 2mM L-glutamine, and penicillin streptomycin.

Five established HNSCC cell lines, UMSSC-10B (base of tongue), UMSSC-22B (larynx), HN-1 (pharynx), HN-12 (tongue), and HN-30 (pharynx), were used in the *in-vitro* portion of this study. UMSSC-10B and UMSSC-22B were gifts from Dr. Tom Carey, University of Michigan, and HN-1, HN-12, and HN-30 were gifts from Dr. J.S. Gutkind, National Institute for Dental and Craniofacial Research. Cell lines were cultured in DMEM supplemented with 10% fetal bovine serum (FBS), 2% streptomycin sulfate (Invitrogen), and 2% L-glutamine (Invitrogen), and incubated at 37°C in 5% CO₂ and 21% O₂.

Validation of Differential Expression by qRT-PCR:

Total RNA was isolated using the RNeasy mini kit (QIAGEN). After polyadenylation (Ambion) and annealing of 1.0 µg total RNA, cDNA was synthesized using M-MLV reverse transcriptase (Promega), according to the manufacturer's instructions. Real-time PCR reaction mixes were created using FastStart Universal SYBR Green Master Mix (Roche Diagnostics) and run on a StepOnePlus™ Real-Time PCR System (Applied Biosystems) using the following program:

95°C for 10 min, 95°C for 30 s, and 60°C for 1 min, for 50 cycles. Results were analyzed using the $\Delta\Delta C_t$ method, and experiments were performed in technical triplicates with *GAPDH* gene expression measured as endogenous control. Strand-specific primers were custom designed by the authors and created by Eurofins Genomics (Huntsville, AL, USA). The following sequences were used: *GAPDH* forward: 5'-CTTCGCTCTCTGCTCCTCC-3', reverse: 5'-CAATACGACCAAATCCGTTG-3', *lnc-LCE5A-1* forward: 5'-GGGCACCTCAAGAAAAGCAT-3', reverse: 5'-GAGCACAGCCACACACTAAA-3', *lnc-KCTD6-3* forward: 5'-AGCCACAGCCACCCTAAAAT-3', reverse: 5'-ACAGCCTCACTCACTGCCTA-3', snoRNA *SNORD35B* forward: 5'-GCAGATGATGTTTGTTCACG-3', reverse: 5'-CGGCATCAGTTTTACCAAGTG-3'.

lncRNA Expression Plasmid Transfection:

Expression plasmids for *lnc-LCE5A-1* and *lnc-KCTD6-3* were custom-designed (Life Technologies) by cloning the respective lncRNA sequences into the pUC19 vector. Plasmids were transiently transfected using Lipofectamine 2000 (Invitrogen), following the manufacturer's specifications. The pUC19 empty vector was used as a control and transfection efficiency for all three plasmids was monitored using GFP as a reporter.

Quantitative Real-Time PCR for Stem Cell and EMT Gene Expression:

Total RNA was collected 48 hours after transfection with the lncRNA expression plasmids using the RNeasy mini kit (QIAGEN). cDNA was synthesized using the Superscript III RT-PCR kit (Invitrogen, Carlsbad, CA) per the manufacturer's instructions. Real-time PCR reaction mixes were created using FastStart Universal SYBR Green Master Mix (Roche Diagnostics), and run on a StepOnePlus™ Real-Time PCR System (Applied Biosystems) using the following program: 95°C for 10 min, 95°C for 30 s, and 60°C for 1 min, for 40 cycles. Results were analyzed using the $\Delta\Delta C_t$ method, and experiments were performed in technical triplicates with

GAPDH gene expression measured as endogenous control. Primers were custom ordered (Eurofins MWG Operon) using the following sequences: *GAPDH* forward: 5'-CTTCGCTCTCTGCTCCTCC-3', reverse: 5'-CAATACGACCAAATCCGTTG -3'. *OCT-4* forward: 5'- GCAAAGCAGAAACCCTCGTGC-3', reverse: 5'-ACCACACTCGGACCACATCCT-3'. *NANOG* forward: 5'- GATTTGTGGGCCTGAAGAAA-3', reverse: 5'-TTGGGACTGGTGGGAAGAATC-3'. *BMI-1* forward: 5'-TCCACAAAGCACACACATCA-3', reverse: 5'- CTTTCATTGTCTTTTCCGCC-3'. *VIM* forward: 5'- GGAAATGGCTCGTCACCTTCGT-3', reverse: 5'-AGAAATCCTGCTCTCCTCGCCT-3'. *TWIST* forward: 5'-GGGCCGGAGACCTAGATGTCATTG-3', reverse: 5'-GAATGCAGAGGTGTGAGGATGGTG-3'. *CDH-1* forward: 5'-CTGATGTGAATGACAACGCC-3', reverse: 5'-TAGATTCTTGGGTTGGGTCG-3'.

MTS Cell Proliferation Assay

UMSCC-10B, UMSCC-22B, HN-1, and HN-30 cells were plated into a 96-well flat-bottom tissue culture plate (Falcon) at a density of 5,000 cells per well. After a 24-hour plating period, cells were transfected with the lncRNA expression plasmids. Following a 48- to 72-hour incubation period, cellular proliferation was analyzed using the CellTiter 96 AQueous non-radioactive proliferation assay (Promega) in accordance with the manufacturer's protocol. All assays were performed in triplicate wells and experiments were individually performed twice.

Wound Healing Migration Assay:

UMSCC-10B and UMSCC-22B cells were cultured in 6-well plates until confluent and were transiently transfected with the lncRNA expression plasmids. 48 hours after transfection, a line in the plate was scored using a P200 pipette tip. The cells were incubated in growth medium, and

at each 12-hour interval, the scratched wound was photographed using a light microscope (Nikon Eclipse TE2000-U).

List of abbreviations

HNSCC: Head and neck squamous cell carcinoma; ncRNA: non-coding RNA; lncRNA: long non-coding RNA; snoRNA: small nucleolar RNA; RNA-seq: RNA sequencing; TCGA: The Cancer Genome Atlas; cpm: counts per million

Competing Interests

The authors declare no competing interests.

Acknowledgements

This work was supported by funding from the National Institutes of Health, grant number DE023242 to WMO, and by funding from the Department of Veterans Affairs, Veterans Health Administration, BLR&D IO1BX002284 to JHL.

References:

- Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J et al. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**(6046): 1154-1157.
- Avissar M, Christensen BC, Kelsey KT, Marsit CJ. 2009. MicroRNA expression ratio is predictive of head and neck squamous cell carcinoma. *Clin Cancer Res* **15**(8): 2850-2855.
- Belbin TJ, Schlecht NF, Smith RV, Adrien LR, Kawachi N, Brandwein-Gensler, M, ... Prystowsky MB. 2008. Site-Specific Molecular Signatures Predict Aggressive Disease in HNSCC. *Head and Neck Pathology*, **2**(4), 243–256.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**(18): 1915-1927.
- Chang LS, Lin SY, Lieu AS, Wu TL. 2002. Differential expression of human 5S snoRNA genes. *Biochem Biophys Res Commun* **299**(2): 196-200.
- Childs G, Fazzari M, Kung G, Kawachi N, Brandwein-Gensler M, McLemore M, Chen Q, Burk RD, Smith RV, Prystowsky MB et al. 2009. Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma. *Am J Pathol* **174**(3): 736-745.
- Contzler R, Favre B, Huber M, Hohl D. 2005. Cornulin, a new member of the "fused gene" family, is expressed during epidermal differentiation. *J Invest Dermatol* **124**(5): 990-997.
- Dong XY, Guo P, Boyd J, Sun X, Li Q, Zhou W, Dong JT. 2009. Implication of snoRNA U50 in human breast cancer. *J Genet Genomics* **36**(8): 447-454.
- Esteller M. 2011. Non-coding RNAs in human disease. *Nat Rev Genet* **12**(12): 861-874.
- Fang Z, Wu L, Wang L, Yang Y, Meng Y, Yang H. 2014. Increased expression of the long non-coding RNA UCA1 in tongue squamous cell carcinomas: a possible correlation with cancer metastasis. *Oral Surg Oral Med Oral Pathol Oral Radiol* **117**(1): 89-95.
- Feng J, Tian L, Sun Y, Li D, Wu T, Wang Y, Liu M. 2012. Expression of long non-coding RNA MALAT-1 is correlated with progress and apoptosis of laryngeal squamous cell carcinoma. *Head Neck Oncol*. Sep 9;**4**(2):46.
- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**(12): 2893-2917.
- Gee HE, Buffa FM, Camps C, Ramachandran A, Leek R, Taylor M, Patil M, Sheldon H, Betts G, Homer J et al. 2011. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer* **104**(7): 1168-1177.
- Geisler S, Coller J. 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* **14**(11): 699-712.
- Guil S, Esteller M. 2012. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol* **19**(11): 1068-1075.
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL et al. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**(7291): 1071-1076.
- Hui AB, Lenarduzzi M, Krushel T, Waldron L, Pintilie M, Shi W, Perez-Ordóñez B, Jurisica I, O'Sullivan B, Waldron J et al. 2010. Comprehensive MicroRNA profiling for head and neck squamous cell carcinomas. *Clin Cancer Res* **16**(4): 1129-1139.
- Imai FL, Uzawa K, Nimura Y, Moriya T, Imai MA, Shiiba M, Bukawa H, Yokoe H, Tanzawa H. 2005. Chromosome 1 open reading frame 10 (C1orf10) gene is frequently down-regulated and inhibits cell proliferation in oral squamous cell carcinoma. *Int J Biochem Cell Biol* **37**(8): 1641-1655.

- Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PH, Reaman G, Milos P, Arceci RJ, Thompson JF et al. 2010. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol* **8**: 149.
- Kokko LL, Hurme S, Maula SM, Alanen K, Grénman R, et al. 2011. Significance of site-specific prognosis of cancer stem cell marker CD44 in head and neck squamous-cell carcinoma. *Oral Oncol* **47**: 510–516.
- Li DD, Feng JP, Wu TY, Wang YD, Sun YN, Ren JY and Liu M. 2013. Long intergenic noncoding RNA HOTAIR is overexpressed and regulates PTEN methylation in laryngeal squamous cell carcinoma. *Am J Pathol* **182**: 64-70.
- Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, Liu Z, Jiang F. 2010. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Mol Cancer* **9**: 198.
- Lin R, Maeda S, Liu C, Karin M, Edgington TS. 2007. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* **26**(6): 851-858.
- Lleras, R, Smith, RV, Adrien, LR, Schlecht, NF, Burk, RD, Harris, T et al. 2013. Unique DNA methylation loci distinguish anatomic site and HPV status in head and neck squamous cell carcinoma. *Clin Cancer Res* **19**(19):5444–5455.
- Luthra MG, Ajani JA, Izzo J, Ensor J, Wu TT, Rashid A, Zhang L, Phan A, Fukami N, Luthra R. 2007. Decreased expression of gene cluster at chromosome 1q21 defines molecular subgroups of chemoradiotherapy response in esophageal cancers. *Clin Cancer Res* **13**(3): 912-919.
- McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**(10): 4288-4297.
- Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**(3): 155-159.
- Michel CI, Holley CL, Scruggs BS, Sidhu R, Brookheart RT, Listenberger LL, Behlke MA, Ory DS, Schaffer JE. 2011. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metab* **14**(1): 33-44.
- Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. 2009. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* **28**(2): 195-208.
- Nadano, D., Ishihara, G., Aoki, C., Yoshinaka, T., Irie, S. & Sato, T. A. 2000. *Jpn. J. Cancer Res.* **91**, 802-810.
- Nemunaitis J, O'Brien J. 2002. Head and neck cancer: gene therapy approaches. Part 1: adenoviral vectors. *Expert Opin Biol Ther* **2**(2): 177-185.
- Nicoloso M, Qu LH, Michot B, Bachellerie JP. 1996. Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *J Mol Biol* **260**(2): 178-195.
- Prensner JR, Chinnaiyan AM. 2011. The emergence of lncRNAs in cancer biology. *Cancer Discov* **1**(5): 391-407.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.
- Ramdas L, Giri U, Ashorn CL, Coombes KR, El-Naggar A, Ang KK, Story MD. 2009. miRNA expression profiles in head and neck squamous cell carcinoma and adjacent normal tissue. *Head Neck* **31**(5): 642-654.
- Rodrigo JP, Suarez C, Gonzalez MV, Lazo PS, Ramos S, Coto E, Alvarez I, Garcia LA, Martinez JA. 2001. Variability of genetic alterations in different sites of head and neck cancer. *Laryngoscope* **111**:1297–1301

- Shen Z, Li Q, Deng H, Lu D, Song H, et al. 2014. Long Non-Coding RNA Profiling in Laryngeal Squamous Cell Carcinoma and Its Clinical Significance: Potential Biomarkers for LSCC. *PLoS ONE* **9**(9): e108237.
- Singh A, Settleman J. 2010. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29**:4741-4751.
- Solmi R, De Sanctis P, Zucchini C, Ugolini G, Rosati G, Del Governatore M, Coppola D, Yeatman TJ, Lenzi L, Caira A et al. 2004. Search for epithelial-specific mRNAs in peripheral blood of patients with colon cancer by RT-PCR. *Int J Oncol* **25**(4): 1049-1056.
- Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A et al. 2011. The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**(6046): 1157-1160.
- Tang H, Wu Z, Zhang J, Su B. 2013. Salivary lncRNA as a potential marker for oral squamous cell carcinoma diagnosis. *Mol Med Rep* **7**(3): 761-766.
- Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P. 2013. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* **41**(Database issue): D246-251.
- Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Research*; **34**(Database issue):D158-D162.
- Williams GT, Farzaneh F. 2012. Are snoRNAs and snoRNA host genes new players in cancer? *Nat Rev Cancer* **12**(2): 84-88.
- Yamato T, Orikasa K, Fukushige S, Orikasa S, Horii A. 1999. Isolation and characterization of the novel gene, TU3A, in a commonly deleted region on 3p14.3-->p14.2 in renal cell carcinoma. *Cytogenet Cell Genet* **87**(3-4): 291-295.
- Yang QQ, Deng YF. 2014. Long non-coding RNAs as novel biomarkers and therapeutic targets in head and neck cancers. *Int J Clin Exp Pathol* **7**(4): 1286-1292.
- Zhang J, Zhang P, Wang L, Piao HL, Ma L. 2014. Long non-coding RNA HOTAIR in carcinogenesis and metastasis. *Acta Biochim Biophys Sin (Shanghai)* **46**(1): 1-5.
- Zhang X, Rice K, Wang Y, Chen W, Zhong Y, Nakayama Y, ... Klibanski A. 2010. Maternally Expressed Gene 3 (MEG3) Noncoding Ribonucleic Acid: Isoform Structure, Expression, and Functions. *Endocrinology*, **151**(3), 939–947.

Figure Legends

Figure 1. Plot of lncRNA log-fold changes against log-counts per million. All 12,407 transcripts consistently expressed in patient samples and evaluated in edgeR are visualized here, plotted as a function of their $\log_2(\text{fold change})$ in HNSCC compared to normal tissue versus their $\log_2(\text{counts per million})$, where counts per million (cpm) is a measure of RNA abundance as captured by normalized read count data. The 7,120 lncRNA transcripts identified to be differentially expressed are highlighted in red. Transcripts not deemed to be significantly dysregulated are shown in black. Red dots lying outside of the 2 blue lines denote transcripts significantly dysregulated by a magnitude of 2-fold change or more.

Figure 2. Heatmaps of the most differentially expressed lncRNAs in HNSCC. (a) Heatmap depicting normalized lncRNA expression levels (in the form of cpm) across normal and HNSCC tissues in our 40-patient cohort. Only the 222 lncRNAs differentially expressed by a magnitude of at least 8-fold in HNSCC with FDRs < 0.0001 are shown (Additional File 1). (b) & (c) Higher resolution images of blue- and orange-labeled regions show, among other lncRNAs, marked downregulation of *lnc-LCE5A-1* and *lnc-KCTD6-3* in HNSCC tumors.

Figure 3. Heatmap depicting snoRNA expression and dysregulation in HNSCC. All 135 snoRNAs consistently detected in our cohort of 31 HNSCC and normal tissue pairs are represented. 33 of these snoRNAs were determined to be significantly dysregulated in HNSCC (Additional File 2).

Figure 4. Kaplan-Meier curves show correlation between expression of 3 ncRNAs in HNSCC tumors and overall survival. Patients were divided into high and low expression groups depending on whether their expression of the given ncRNA fell above or below the median. Low expression of (a) *lnc-LCE5A-1*, (b) *lnc-KCTD6-3*, and (c) snoRNA *SNORD35B* was significantly associated with poorer survival in univariate analyses.

Figure 5. Blat visualization of survival-associated ncRNAs via the Ensembl Genome Browser. (a) Blat search of *lnc-LCE5A-1* reveals its antisense overlap with gene *CRNN*. (b) Blat search of *lnc-KCTD6-3* reveals its antisense overlap with gene *FAM107A* and exonic antisense overlap with *FAM3D*. (c) Blat search of snoRNA *SNORD35B* shows its location in an intron of *RPS11* (40S ribosomal protein S11).

Figure 6. qRT-PCRs demonstrate that the 3 ncRNAs of interest are downregulated *in-vitro*.

Expression levels of *lnc-LCE5A-1*, *lnc-KCTD6-3*, and *SNORD35B* were compared between 2 normal oral epithelial cell lines (OKF4 and OKF6), shown in black and 5 HNSCC cell lines (UMSCC-10B, UMSCC-22B, HN-1, HN-12, and HN-30), shown in grey, with expression in OKF4 serving as the reference. (a) & (c) *lnc-LCE5A-1* and *SNORD35B* exhibited significant downregulation across all cancerous cell lines, while (b) *lnc-KCTD6-3* showed downregulation in four of the five cell lines.

Figure 7. Ectopic expression of *lnc-LCE5A-1* and *lnc-KCTD6-3* inhibits cancer stem cell and EMT-inducing genes.

(a) *lnc-LCE5A-1*-transfected HNSCC cell lines exhibited induction of *CDH-1* (E-cadherin) and reduction in *VIM*, *NANOG*, and *OCT-4* expression levels, while (b) *lnc-KCTD6-3*-transfected cell lines exhibited downregulation of *VIM*, *TWIST*, *BMI-1*, and *NANOG*.

Figure 8. Overexpression of *lnc-LCE5A-1* and *lnc-KCTD6-3* dramatically reduces HNSCC cell migration.

Wound healing migration assays in (a) UMSCC-10B and (b) UMSCC-22B HNSCC cell lines show decreased cell motility in *lnc-LCE5A-1* and *lnc-KCTD6-3*-transfected cell lines. Graphs plot average scratch wound width at each 12-hour timepoint.

Figure 9. Overexpression of *lnc-LCE5A-1* and *lnc-KCTD6-3* reduces HNSCC cell proliferation.

3-day MTS proliferation assays demonstrate at least 25% reduction in growth in (a) *lnc-LCE5A-1*-transfected cell lines and (b) *lnc-KCTD6-3*-transfected cell lines as compared to vector-transfected cells.

Description of additional data files

Additional data file 1 contains tables listing the most upregulated and downregulated lncRNAs in HNSCC (≥ 8 -fold change magnitude, FDR < 0.0001). Additional data file 2 is a table listing the 33 differentially expressed snoRNAs in HNSCC (FDR < 0.05). Additional data file 3 is a table listing the known, HNSCC-associated lncRNAs significantly dysregulated in our 40-patient HNSCC cohort. Additional data file 4 contains tables summarizing the most differentially expressed and known, cancer-associated lncRNAs among HNSCCs stratified by anatomic site. Additional data file 5 contains a focused analysis of ncRNA expression level and patient survival vs. patient age, as well as tables summarizing the demographics and clinicopathological characteristics of the expanded HNSCC patient cohort used in the survival analyses. Additional data file 6 contains annotated genome browser images for the survival-associated ncRNAs, including average read coverage of lncRNA exons and sequence alignment of the *SNORD35* paralogs. Additional data file 7 contains qRT-PCR data verifying the lncRNA transfections. Additional data file 8 contains lists of TCGA IDs corresponding to the HNSCC patients used for the matched pair lncRNA and snoRNA differential expression analyses. Additional data file 9

contains lists of TCGA IDs corresponding to the expanded cohort of HNSCC patients used for the survival analyses.

Table 1. Summary of patient demographics and clinicopathological characteristics. (a) Clinical data for patients used in the lncRNA expression profiling and analysis. **(b)** Clinical data for patients used in the snoRNA differential expression analysis. ^A Clinical data not available for one patient

(a) Characteristic	Patients (n = 39)^A	
	No.	%
Age		
Median	64	
Range	29-87	
Sex		
Male	27	69%
Female	12	31%
Vital Status		
Living	8	21%
Deceased	31	79%
Tumor Site		
Oral Cavity	15	38%
Tongue	13	33%
Larynx	11	28%
Clinical Stage		
Stage I	2	5%
Stage II	13	33%
Stage III	16	41%
Stage IV	8	21%
Pathologic Stage		
Stage I	2	5%
Stage II	15	38%
Stage III	7	18%
Stage IV	15	38%
Tumor Grade		
G1-G2	25	64%
G3-G4	10	26%
GX	4	10%

(b) Characteristic	Patients (n = 31)	
	No.	%
Age		
Median	64	
Range	29-87	
Sex		
Male	22	71%
Female	9	29%
Vital Status		
Living	7	23%
Deceased	24	77%
Tumor Site		
Oral Cavity	12	39%
Tongue	9	29%
Larynx	10	32%
Clinical Stage		
Stage I	2	6%
Stage II	10	32%
Stage III	12	39%
Stage IV	7	23%
Pathologic Stage		
Stage I	2	6%
Stage II	12	39%
Stage III	3	10%
Stage IV	14	45%
Tumor Grade		
G1-G2	20	65%
G3-G4	8	26%
GX	3	10%

Table 2. Genomic coordinates and aliases for survival-associated ncRNAs.

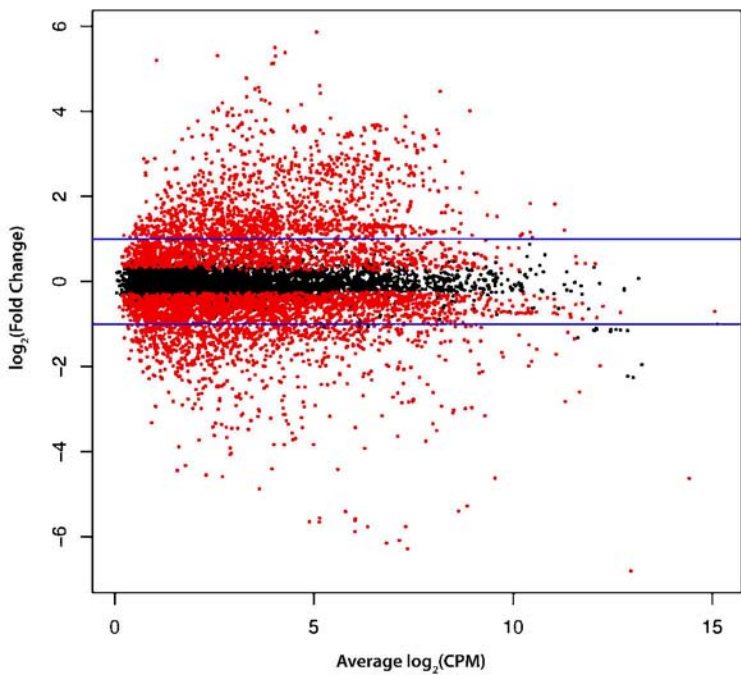
Gene name	Genomic Coordinates (hg19)	Aliases (Ensembl and NONCODEv4)
<i>lnc-LCE5A-1</i>	chr1: 152,346,430-152,417,932	RP1-91G5.3, ENSG00000227415 (ENST00000411804) NONHSAG002953 (NONHSAT006486)
<i>lnc-KCTD6-3</i>	chr3:58592807-58620167	RP11-475O23.3-001, ENSG00000244383 (ENST00000464125) NONHSAG035285 (NONHSAT090141)
<i>SNORD35B</i>	chr19: 50,000,977-50,001,063	SNORD35B, RNU35B, ENSG00000200530 NONHSAT067208

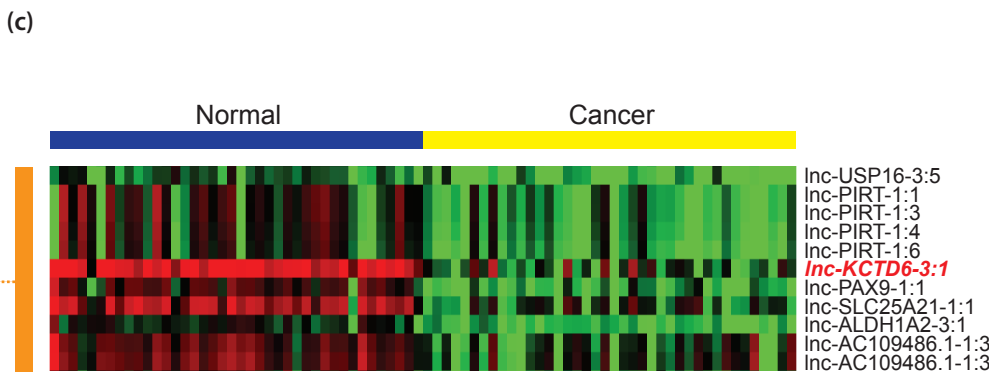
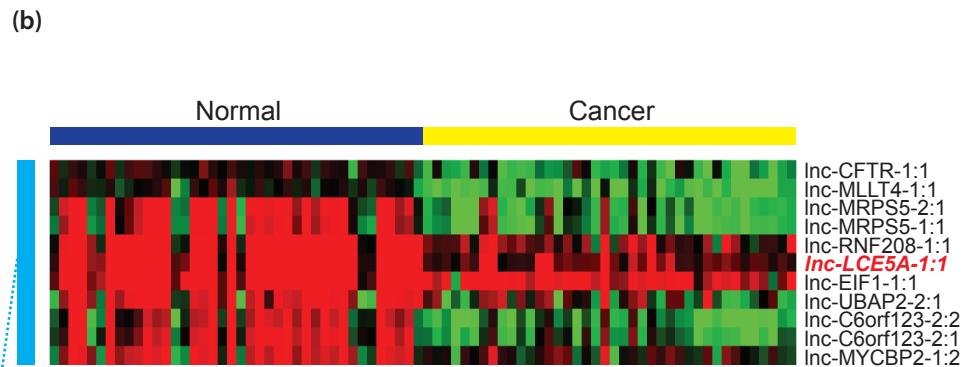
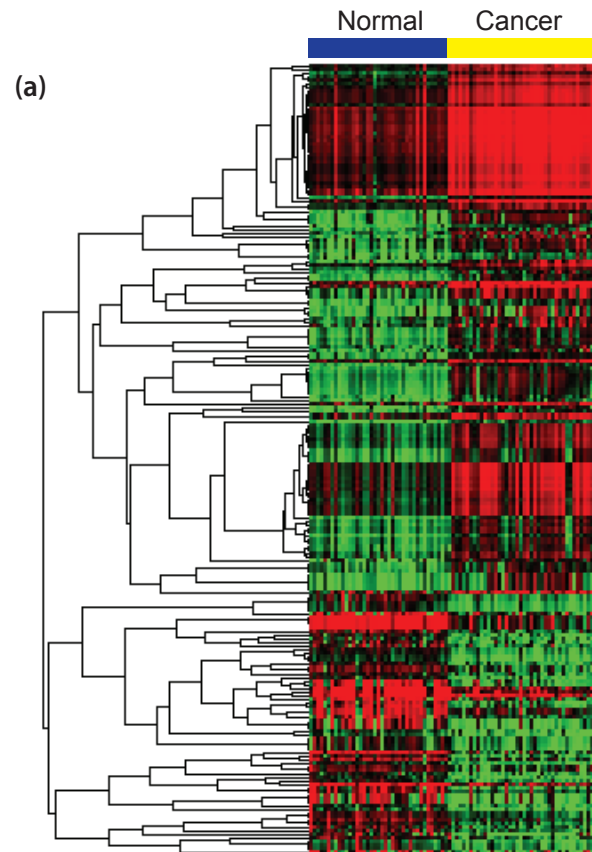
Table 3. Multivariate Cox proportional hazards models reveal significant associations between low ncRNA expression and poor prognosis. Only patients with available data for all variables shown were included in the analyses (HR = Hazard Ratio, CI = Confidence Interval).

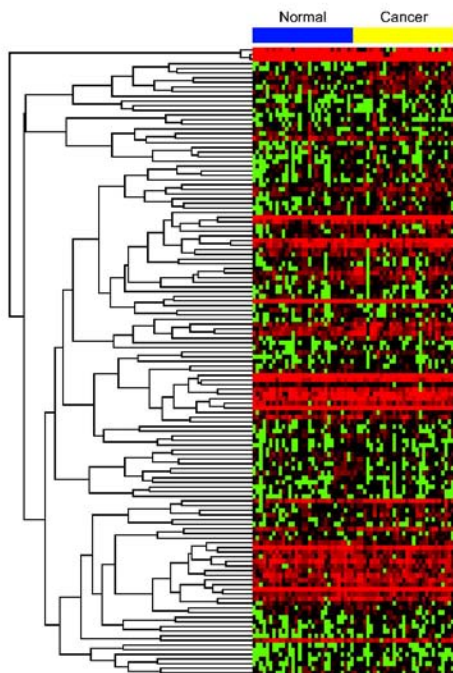
Patient Mortality (n = 75)		
(a)	HR (95% CI)	P-Value
<i>lnc-LCE5A-1</i> Expression (Low vs High)	2.46 (1.26-4.80)	0.008
Age at Initial Diagnosis	1.07 (1.03-1.11)	0.001
Gender (Male vs Female)	1.08 (0.49-2.39)	0.841
Tumor Grade (G3-G4 vs G1-G2)	1.36 (0.71-2.60)	0.359
Clinical Stage (Reference: Stage I)		
Stage II	0.93 (0.12-6.98)	0.947
Stage III	0.40 (0.05-3.54)	0.412
Stage IVA	0.32 (0.03-2.90)	0.308
Stage IVB	1.99 (0.14-28.79)	0.614
Pathologic Stage (Reference: Stage I)		
Stage II	1.02 (0.13-7.80)	0.986
Stage III	5.64 (0.61-51.99)	0.127
Stage IVA	8.60 (0.94-78.68)	0.057
Stage IVB	11.48 (0.80-165.35)	0.073
Patient Mortality (n = 75)		
(b)	HR (95% CI)	P-Value
<i>lnc-KCTD6-3</i> Expression (Low vs. High)	1.93 (1.06-3.54)	0.033
Age at Initial Diagnosis	1.06 (1.02-1.11)	0.002
Gender (Male vs. Female)	1.35 (0.65-2.81)	0.419
Tumor Grade (G3-G4 vs. G1-G2)	1.37 (0.71-2.63)	0.348
Clinical Stage (Reference: Stage I)		
Stage II	1.50 (0.14-16.01)	0.735
Stage III	0.69 (0.06-8.33)	0.767
Stage IVA	0.59 (0.05-7.41)	0.681
Stage IVB	2.84 (0.14-56.16)	0.494

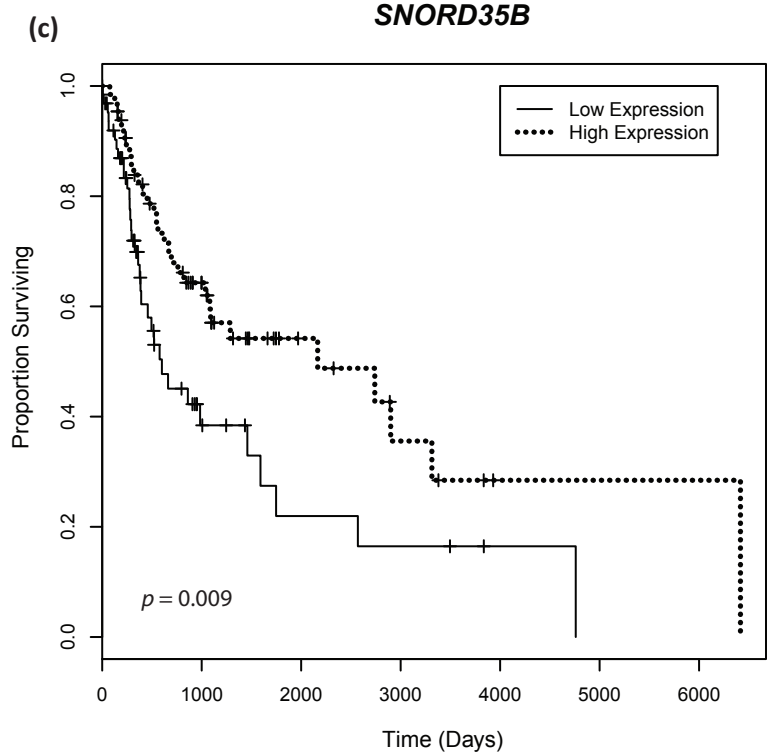
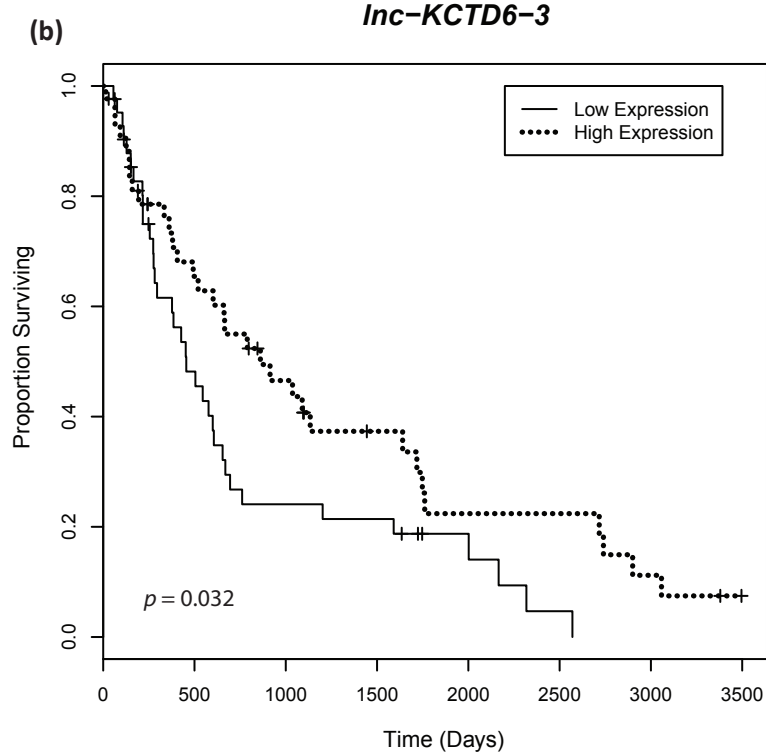
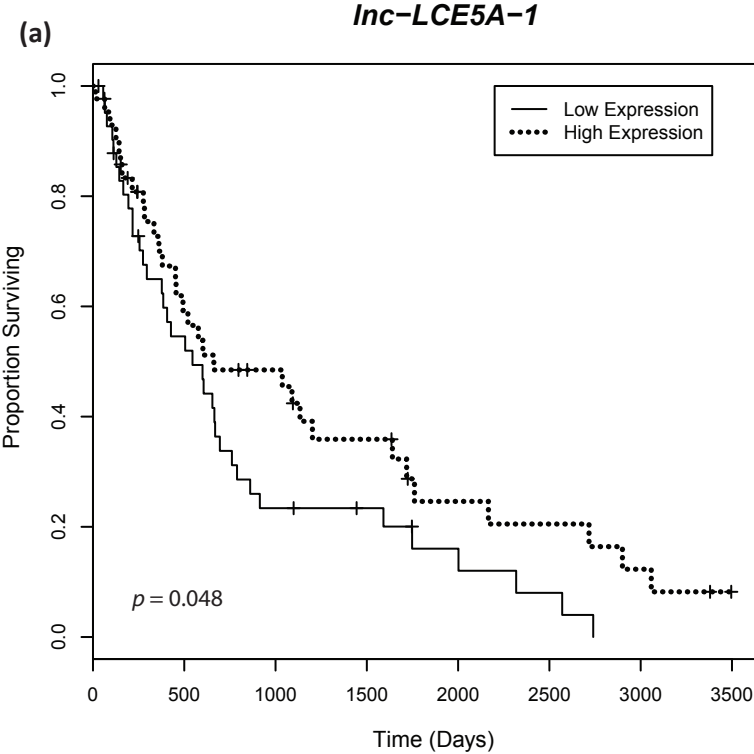
Pathologic Stage (Reference: Stage I)		
Stage II	0.66 (0.06-7.33)	0.734
Stage III	3.36 (0.26-43.19)	0.349
Stage IVA	4.46 (0.35-57.42)	0.252
Stage IVB	5.35 (0.26-111.18)	0.279

(c)	Patient Mortality (<i>n</i> = 109)	
	HR (95% CI)	<i>P</i> -Value
snoRNA <i>SNORD35B</i> Expression (Low vs. High)	2.93 (1.56-5.52)	0.0008
Age at Initial Diagnosis	1.0 (1.02-1.08)	0.002
Gender (Male vs. Female)	1.27 (0.59-2.74)	0.543
Tumor Grade (G3-G4 vs. G1-G2)	1.32 (0.79-2.21)	0.293
Clinical Stage (Reference: Stage I)		
Stage II	2.22 (0.44-11.29)	0.337
Stage III	1.33 (0.25-7.04)	0.736
Stage IVA	0.99 (0.14-5.67)	0.994
Stage IVB	4.44 (0.47-41.74)	0.192
Pathologic Stage (Reference: Stage I)		
Stage II	1.14 (0.21-6.19)	0.870
Stage III	1.57 (0.29-8.54)	0.600
Stage IVA	4.19 (0.80-21.99)	0.091
Stage IVB	5.30 (0.62-45.16)	0.127

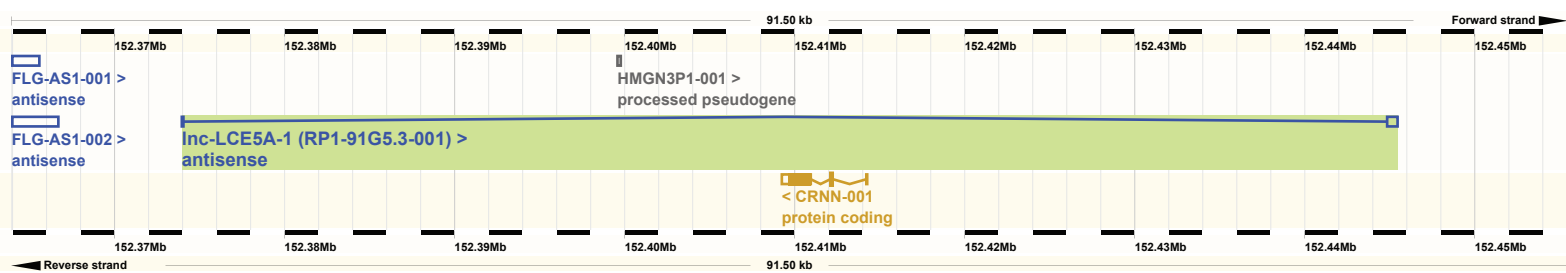




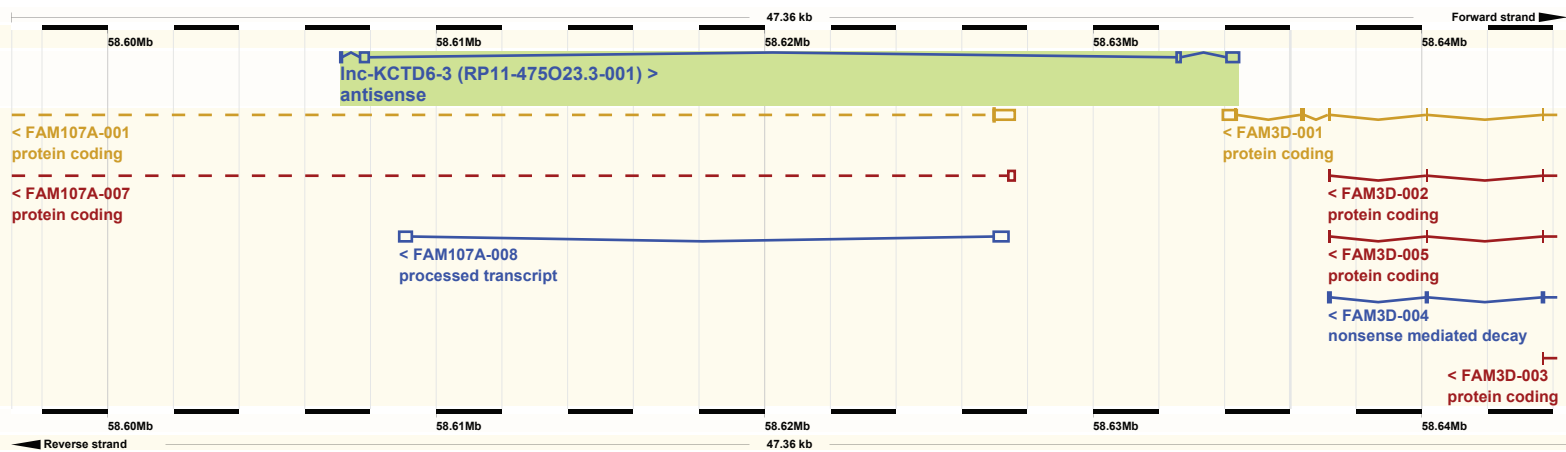




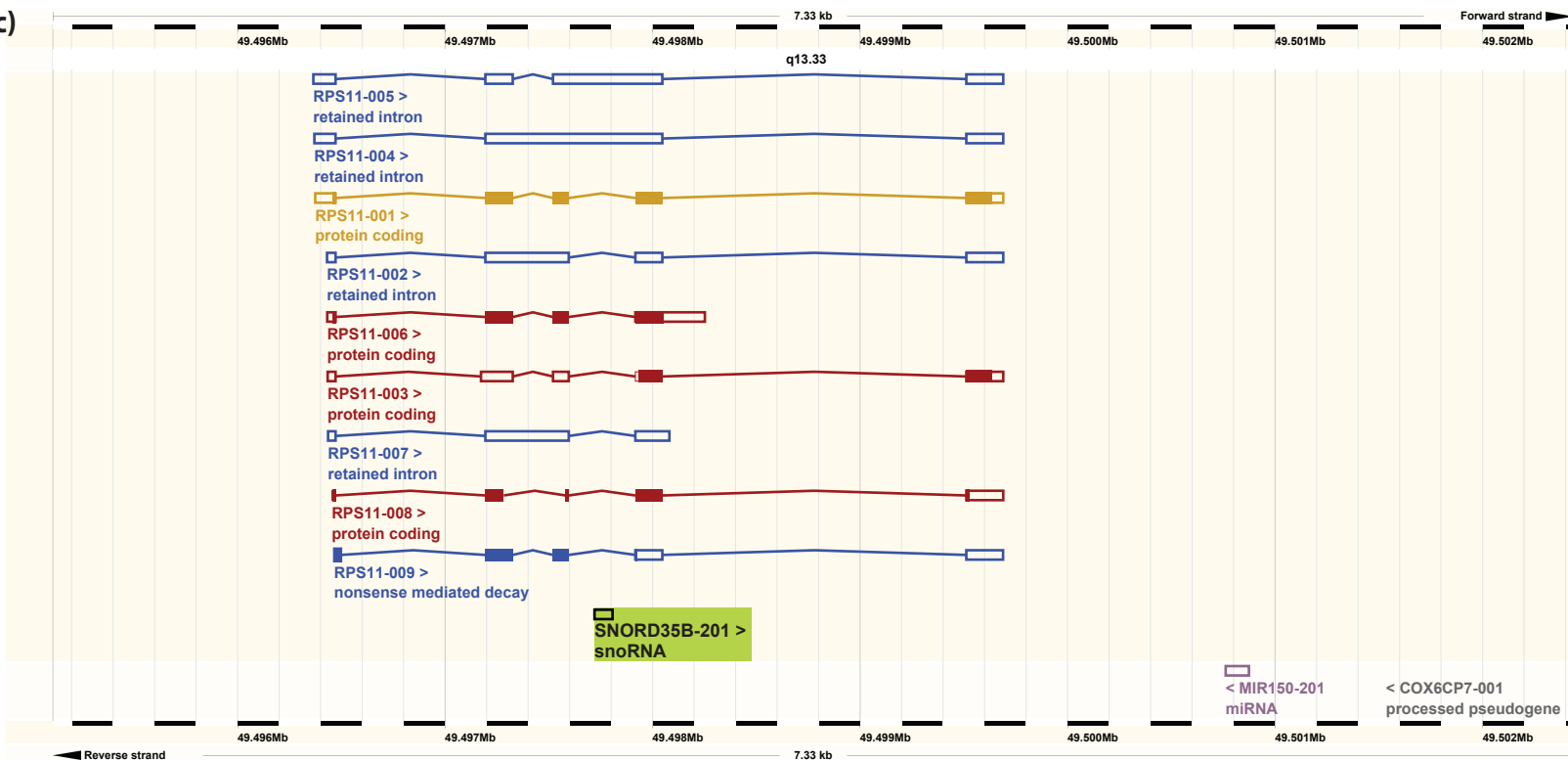
(a)

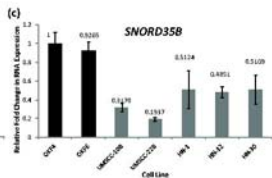
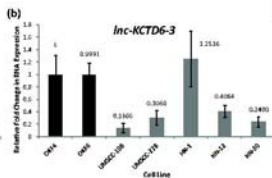
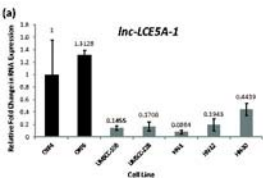


(b)

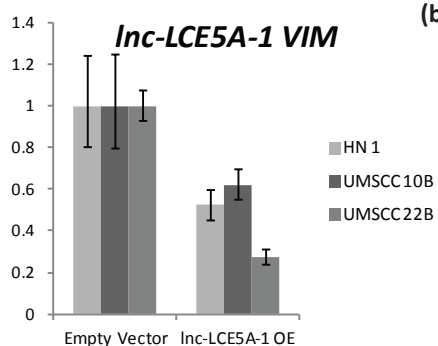
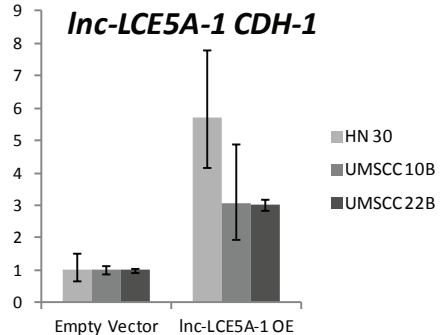


(c)

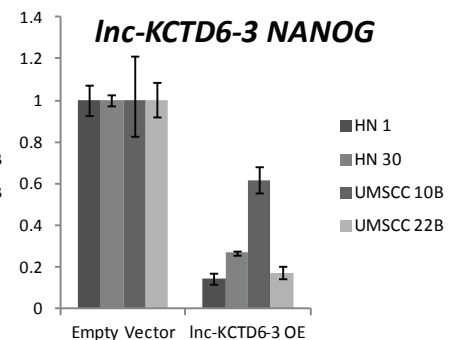
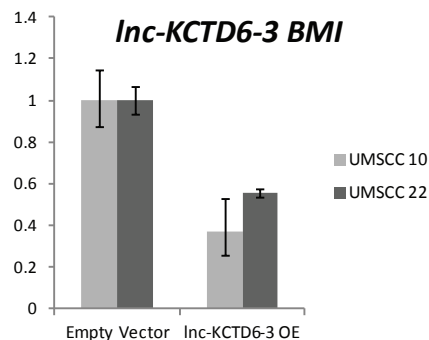
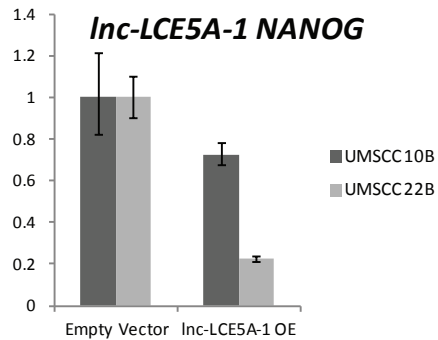
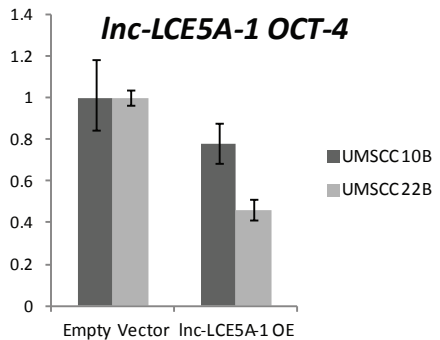
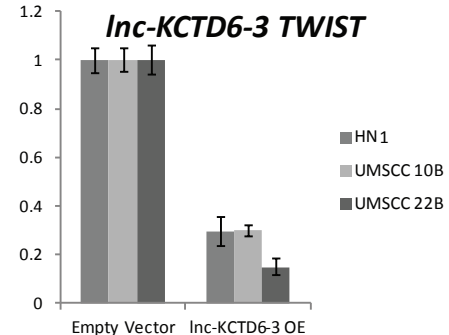
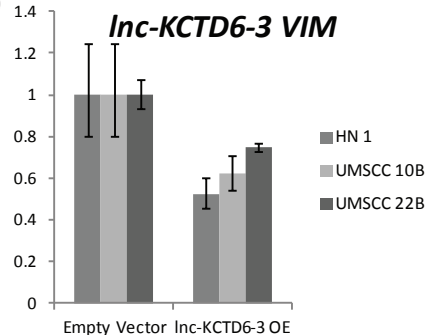




(a)



(b)

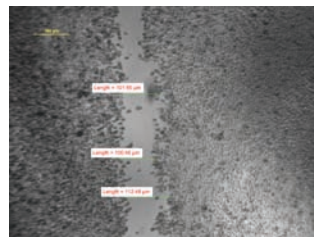
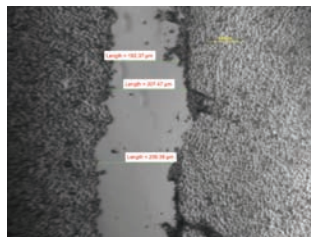
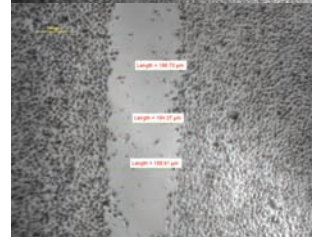
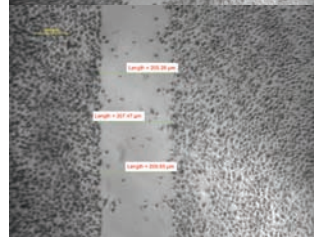
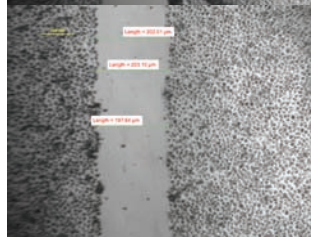
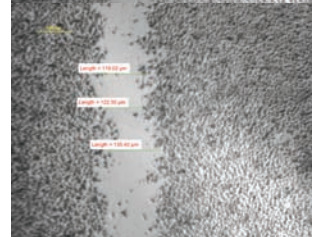
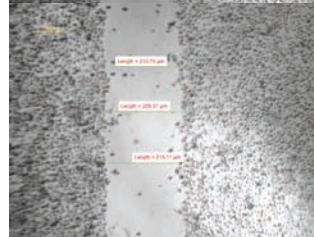
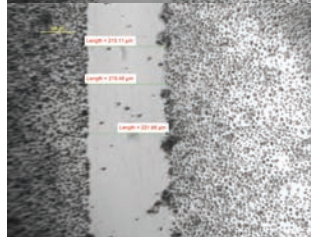
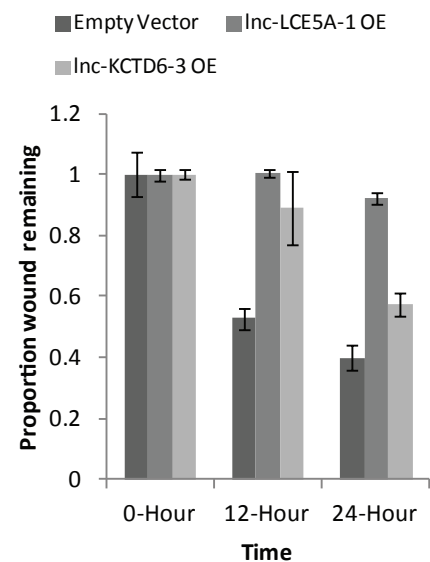
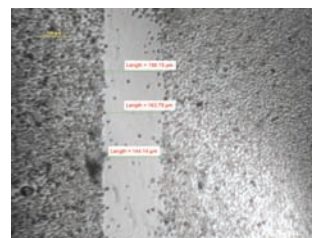
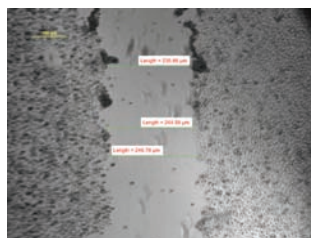
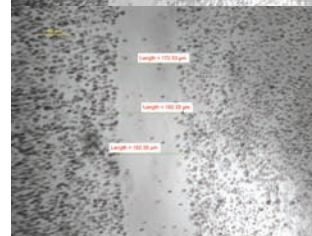
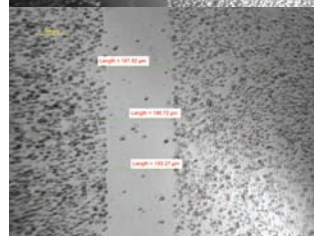
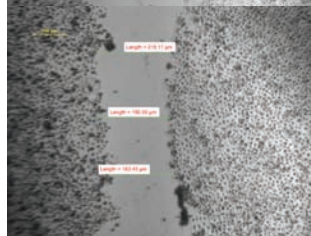
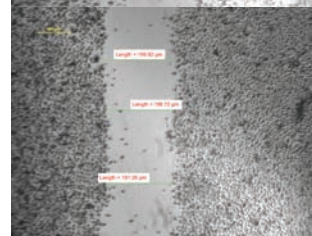
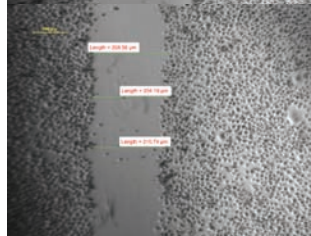


(a)

0 Hours

12 Hours

24 Hours

UMSCC-10B
Empty
VectorUMSCC-10B
Inc-LCE5A-1
OEUMSCC-10B
Inc-KCTD6-3
OE**UMSCC-10B****(b)**UMSCC-22B
Empty
VectorUMSCC-22B
Inc-LCE5A-1
OEUMSCC-22B
Inc-KCTD6-3
OE**UMSCC-22B**